



# The Search for a Title

A Profound Subtitle

Dr. John Smith



Copyright © 2013 John Smith

PUBLISHED BY PUBLISHER

BOOK-WEBSITE.COM

Licensed under the Creative Commons Attribution-NonCommercial 3.0 Unported License (the “License”). You may not use this file except in compliance with the License. You may obtain a copy of the License at <http://creativecommons.org/licenses/by-nc/3.0>. Unless required by applicable law or agreed to in writing, software distributed under the License is distributed on an “AS IS” BASIS, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. See the License for the specific language governing permissions and limitations under the License.

*First printing, March 2013*

# Contents

I	Part One	
<b>1</b>	.....	<b>7</b>
1.1	Lecture 1, August 24	7
1.2	Lecture 2, August 26	8
1.3	Lecture 3, August 28	10
1.4	Lecture 4, August 31	12
1.5	Lecture 5, September 2	14
1.6	Lecture 6, September 4	16
1.7	Lecture 7, September 9	17
1.8	Lecture 8, September 11	19
<b>2</b>	.....	<b>21</b>
2.1	Lecture 9, September 14	21
2.2	Lecture 10, September 16	25
2.3	Lecture 11, September 18	27
2.4	Lecture 12, September 23	28
<b>3</b>	.....	<b>31</b>
3.1	Lecture 12, September 23 (continued)	31
3.2	Lecture 13, September 25	32
3.3	Lecture 14, September 28	33
3.4	Lecture 15, September 30	34

3.5	Lecture 16, October 2	36
3.6	Lecture 17, October 5	37
3.7	Lecture 18, October 7	38
3.8	Lecture 19, October 9	40
	<b>Index</b> .....	<b>43</b>



# Part One

<b>1</b>	.....	<b>7</b>
1.1	Lecture 1, August 24	
1.2	Lecture 2, August 26	
1.3	Lecture 3, August 28	
1.4	Lecture 4, August 31	
1.5	Lecture 5, September 2	
1.6	Lecture 6, September 4	
1.7	Lecture 7, September 9	
1.8	Lecture 8, September 11	
<b>2</b>	.....	<b>21</b>
2.1	Lecture 9, September 14	
2.2	Lecture 10, September 16	
2.3	Lecture 11, September 18	
2.4	Lecture 12, September 23	
<b>3</b>	.....	<b>31</b>
3.1	Lecture 12, September 23 (continued)	
3.2	Lecture 13, September 25	
3.3	Lecture 14, September 28	
3.4	Lecture 15, September 30	
3.5	Lecture 16, October 2	
3.6	Lecture 17, October 5	
3.7	Lecture 18, October 7	
3.8	Lecture 19, October 9	
<b>Index</b>	.....	<b>43</b>



# 1.

## 1.1 Lecture 1, August 24

**Definition 1.1.1 — Field.** A class  $\mathcal{F}$  of subsets of a non-empty set  $\Omega$  is called a *field* if it satisfies

1.  $\emptyset \in \mathcal{F}$  and  $\Omega \in \mathcal{F}$
2. If  $A \in \mathcal{F}$ , then  $A^c \in \mathcal{F}$  (closed under complementation)
3. If  $A \in \mathcal{F}$  and  $B \in \mathcal{F}$ , then  $A \cup B \in \mathcal{F}$  (closed under finite unions)

*Note:* The term *algebra* is often used to denote a field.

**Definition 1.1.2 —  $\sigma$ -Field.** A field is a  $\sigma$ -field or  $\sigma$ -algebra if it is closed under countable unions. That is, if  $A_n \in \mathcal{F}$  is a  $\sigma$ -field, then  $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$ .

**R** A field is a *set of sets*. Suppose we are flipping a coin twice, so the sample space is  $\Omega = \{HH, HT, TH, TT\}$ . We observe an event  $A \subset \Omega$ , say  $A = \{TT\}$ . If we consider the set of sets  $\mathcal{G} = \{\emptyset, \Omega, A, A^c\}$ , then we *cannot* say  $TT \in \mathcal{G}$ . Rather, we should write  $\{TT\} \in \mathcal{G}$  or  $A \in \mathcal{G}$ .

The following are some examples of fields.

- $\{\emptyset, \Omega\}$
- $\{\emptyset, \Omega, A, A^c\}$

The above are very easy to verify using the three properties of Definition 1.1.1. The next three take a little more thought.

■ **Example 1.1 — Finite-cofinite field.** Define the set  $\mathcal{F} = \{A : \text{either } A \text{ is finite or } A^c \text{ is finite}\}$ . The term “finite-cofinite” comes from this definition. Every element is either finite, or its complement is. Suppose our sample space is the natural numbers,  $\Omega = \{1, 2, 3, \dots\}$ . Clearly  $\emptyset \in \mathcal{F}$  and  $\Omega \in \mathcal{F}$ , satisfying property 1.

For property 2, suppose  $A \in \mathcal{F}$ . Again, this means that either  $A$  is finite or  $A^c$  is. If  $A$  is finite, then  $A^c \in \mathcal{F}$  because  $(A^c)^c = A$  is finite. If  $A^c$  is finite, then  $A^c \in \mathcal{F}$  by our construction of  $\mathcal{F}$ . Either way, we see that  $A^c \in \mathcal{F}$  and therefore property 2 holds.

For property 3, there are four cases in which  $A \in \mathcal{F}$  and  $B \in \mathcal{F}$ :  $A$  and  $B$  are finite,  $A^c$  and  $B$  are

finite,  $A$  and  $B^c$  are finite, or  $A^c$  and  $B^c$  are finite. Obviously if  $A$  and  $B$  are finite then  $A \cup B$  is also finite and  $A \cup B \in \mathcal{F}$ . If  $A^c$  and  $B$  are finite, then  $(A \cup B)^c = A^c \cap B^c \subset A^c$ , so  $(A \cup B)^c$  is finite and  $A \cup B \in \mathcal{F}$ . This argument shows that  $A \cup B \in \mathcal{F}$  in the third and fourth cases as well. Therefore property 3 holds, and we can conclude that  $\mathcal{F}$  is a field.

We might also ask whether  $\mathcal{F}$  is a  $\sigma$ -field. Take  $A_1 = \{2\}$ ,  $A_2 = \{4\}$ ,  $\dots$ ,  $A_n = \{2n\}$ . Then  $\bigcup_{n=1}^{\infty} A_n = \{2, 4, 6, 8, \dots\}$ , the set of all even integers. This is an infinite set, as is its complement, the set of all odd integers. Thus  $\bigcup_{n=1}^{\infty} A_n \notin \mathcal{F}$ , meaning that  $\mathcal{F}$  is a field but *not* a  $\sigma$ -field. ■

■ **Example 1.2 — Countable-cocountable field.** Define the set  $\mathcal{F} = \{A : \text{either } A \text{ is countable or } A^c \text{ is countable}\}$ . Suppose the sample space is some infinite set. Again, property 1 is satisfied because  $\emptyset \in \mathcal{F}$  and  $\Omega \in \mathcal{F}$ . The arguments for why properties 2 and 3 are satisfied follow the same reasoning in Example 1.1.

Is  $\mathcal{F}$  a  $\sigma$ -field? If  $\Omega$  is a countable set, then  $\mathcal{F} = \mathcal{P}(\Omega)$ , the *power set* of  $\Omega$  (simply the set of all possible subsets of  $\Omega$ ).  $\mathcal{F}$  is a  $\sigma$ -field, because any  $\bigcup_{n=1}^{\infty} A_n$  is a subset of  $\Omega$ , and therefore is contained in  $\mathcal{F}$ . If  $\Omega$  is uncountable, then  $\mathcal{F}$  is still a  $\sigma$ -field. ■

The previous two examples show two sets that appear similar and are both fields, but the one in Example 1.2 is a  $\sigma$ -field whereas the one in Example 1.1 is not.

■ **Example 1.3** Define  $\mathcal{I} = \{(a, b] : 0 \leq a \leq b \leq 1\}$ . Then if we take  $\mathcal{B}_0 = \{\emptyset, \text{all finite disjoint unions of sets from } \mathcal{I}\}$ , it turns out that  $\mathcal{B}_0$  is a field on  $(0, 1]$ . Again, it is easy to verify that properties 1 and 2 hold. Next, take  $A = \bigcup_{i=1}^k I_i$  and  $B = \bigcup_{j=1}^m I_j$ . Then  $A^c = \bigcap_{i=1}^k I_i^c$ , which is still a finite union of disjoint intervals. So  $A^c \in \mathcal{B}_0$ , and  $\mathcal{B}_0$  is indeed a field.

To determine whether it is a  $\sigma$ -field, take  $A_n = (0, 1 - \frac{1}{n}]$ . Then  $\bigcup_{n=1}^{\infty} A_n = (0, 1)$ . This open interval cannot be written as a finite disjoint union of intervals, so  $\mathcal{B}_0$  is *not* a  $\sigma$ -field. ■

In the following definitions, let  $\mathcal{A}$  be a class of subsets of a non-empty set  $\Omega$ .

**Definition 1.1.3** The field *generated* by  $\mathcal{A}$  is the smallest field containing  $\mathcal{A}$ :

$$f(\mathcal{A}) = \bigcap_{\text{field } \mathcal{G} \supset \mathcal{A}} \mathcal{G}.$$

The  $\sigma$ -field *generated* by  $\mathcal{A}$  is the smallest  $\sigma$ -field containing  $\mathcal{A}$ :

$$\sigma(\mathcal{A}) = \bigcap_{\sigma\text{-field } \mathcal{G} \supset \mathcal{A}} \mathcal{G}.$$

A very useful  $\sigma$ -field is the *Borel  $\sigma$ -field*.

■ **Definition 1.1.4**  $\mathcal{B} = \sigma(\mathcal{I}) = \sigma(\mathcal{B}_0) = \sigma(\mathcal{I}_0) = \sigma(\text{open sets}) = \sigma(\text{open intervals})$  is the *Borel  $\sigma$ -field* on  $(0, 1]$ , where  $\mathcal{I}_0 = \{(a, b] \in \mathcal{I} : a, b \text{ rationals}\}$ . Sets in  $\mathcal{B}$  are called *Borel sets*.

## 1.2 Lecture 2, August 26

We will now show that  $\sigma(\mathcal{I}_0) = \sigma(\mathcal{I})$ . It is clear from the definitions of the sets that  $\mathcal{I}_0 \subset \mathcal{I}$ , so trivially  $\sigma(\mathcal{I}_0) \subset \sigma(\mathcal{I})$ . Now suppose  $(a, b] \in \mathcal{I}$ . We can always find  $a_n \in (a, b]$  such that  $a_n$  is rational and  $a_n \downarrow a$ , meaning that the sequence of  $a_n$ 's decreases and approaches  $a$ . Then taking the infinite union  $\bigcup_{n=1}^{\infty} (a_n, b] = (a, b]$ . Similarly, we can find a rational  $b_n$  such that  $b_n \downarrow b$ . Then  $\bigcap_{n=1}^{\infty} (a_n, b_n] = (a, b]$ , and each  $(a_n, b_n] \in \mathcal{I}_0$ . Thus  $\mathcal{I} \subset \sigma(\mathcal{I}_0) \subset \sigma(\mathcal{I})$ , meaning that  $\sigma(\mathcal{I}) \subset \sigma(\mathcal{I}_0) \subset \sigma(\mathcal{I})$ . We can conclude from this that  $\sigma(\mathcal{I}) = \sigma(\mathcal{I}_0)$ .



A  $\sigma$ -field is *countably generated*, or *separable*, if it is generated by some countable class of sets.



**Theorem 1.2.1** For a nonempty class  $\mathcal{A}$ , the field  $f(\mathcal{A})$  generated by  $\mathcal{A}$  is minimal (if  $\mathcal{H}$  is a field and  $\mathcal{A} \subset \mathcal{H}$ , then  $f(\mathcal{A}) \subset \mathcal{H}$ ) and has the form

$$\mathcal{G} = f(\mathcal{A}) = \left\{ \bigcup_{i=1}^m \bigcap_{j=1}^{n_i} A_{ij} : A_{ij} \text{ or } A_{ij}^c \in \mathcal{A}, \bigcap_{j=1}^{n_i} A_{ij} \text{ are disjoint} \right\}.$$

In short, the sets in  $f(\mathcal{A})$  can be explicitly presented, which is not generally true of the sets in  $\sigma(\mathcal{A})$ .

*Proof.* Clearly  $\mathcal{A} \subset \mathcal{G}$ , and any field containing  $\mathcal{A}$  will also contain  $\mathcal{G}$ . So the “minimality” idea will hold, and all that remains is to show that  $\mathcal{G}$  is indeed a field.

Take  $C = \bigcup_{i=1}^m \bigcap_{j=1}^{n_i} A_{ij}$  and  $D = \bigcup_{s=1}^k \bigcap_{l=1}^{r_s} B_{sl}$  such that  $C, D \in \mathcal{G}$ . That is, all  $\bigcap_{j=1}^{n_i} A_{ij}$  and  $\bigcap_{l=1}^{r_s} B_{sl}$  are disjoint, either  $A_{ij}$  or  $A_{ij}^c$  are in  $\mathcal{A}$ , and either  $B_{sl}$  or  $B_{sl}^c$  are in  $\mathcal{A}$ . Then  $C \cup D = \bigcup_{i=1}^m \bigcup_{s=1}^k \left( \bigcap_{j=1}^{n_i} A_{ij} \right) \cap \left( \bigcap_{l=1}^{r_s} B_{sl} \right)$ . Everything to the right of the first union is disjoint, so  $C \cup D \in \mathcal{G}$ . Furthermore,  $C^c = \bigcap_{i=1}^m \bigcup_{j=1}^{n_i} A_{ij}^c = \bigcup_{j=1}^{n_i} \left( A_{ij}^c \cap \bigcap_{k=1}^{i-1} A_{ik} \right)$ , and again everything in the parentheses is disjoint, so  $C^c \in \mathcal{G}$ . Therefore  $\mathcal{G}$  is a field. ■

As an example of this fact, consider again the class  $\mathcal{I} = \{(a, b] : 0 \leq a \leq b \leq 1\}$ , where  $\Omega = (0, 1]$ . Here, using Theorem 1.2.1,  $f(\mathcal{I}) = \{\emptyset, \text{finite disjoint unions of intervals}\}$ . There is no such nice description of the Borel  $\sigma$ -field  $\mathcal{B}$ , or of  $\sigma$ -fields in general.

Some elementary properties of fields:

1. If  $\mathcal{A}$  consists of singleton sets, then  $f(\mathcal{A})$  is the finite-cofinite field  $f(\mathcal{A}) = \{A : \text{either } A \text{ is finite or } A^c \text{ is finite}\}$ .
2.  $f(\mathcal{A}) \subset \sigma(\mathcal{A})$ .
3. If  $\mathcal{A}$  is finite, then  $f(\mathcal{A}) = \sigma(\mathcal{A})$ .
4. If  $\mathcal{A}$  is countable, then  $f(\mathcal{A})$  is countable.
5. If  $\mathcal{F}_1$  and  $\mathcal{F}_2$  are fields, then  $f(\mathcal{F}_1 \cup \mathcal{F}_2) = \mathcal{G}$ , where

$$\mathcal{G} = \left\{ \bigcup_{i=1}^m (A_i \cap B_i) : A_i \in \mathcal{F}_1, B_i \in \mathcal{F}_2, A_i \cap B_i \text{ are disjoint} \right\}.$$

This can be seen by noting that  $\mathcal{G}$  is closed under intersections, and  $A^c \cap B^c = A^c \cup (A \cap B^c)$ . The important fact here is that the union of two fields is not necessarily a field itself. For example, let  $\Omega = \{1, 2, 3, 4\}$ ,  $A = \{1\}$ ,  $B = \{2\}$ ,  $\mathcal{F}_1 = \{\emptyset, \Omega, A, A^c\}$ , and  $\mathcal{F}_2 = \{\emptyset, \Omega, B, B^c\}$ . Then  $\mathcal{F}_1 \cup \mathcal{F}_2 = \{\emptyset, \Omega, A, B, A^c, B^c\}$ , which is not a field because  $A \cup B$  is not there.

6. If  $\mathcal{F}_n \subset \mathcal{F}_{n+1}$  are fields, then  $\bigcup_n \mathcal{F}_n$  is a field. However, it is not necessarily a  $\sigma$ -field, even if the  $\mathcal{F}_n$  are all  $\sigma$ -fields. For example, let  $\Omega = \{1, 2, 3, \dots\}$  and  $\mathcal{F}_n = \sigma(\{\{k\} : 1 \leq k \leq n\})$ . Then  $\bigcup_n \mathcal{F}_n$  is the finite-cofinite field, which we have already seen is not a  $\sigma$ -field.
7.  $\sigma(\mathcal{A})$  is countably generated (separable) if  $\mathcal{A}$  is countable.
8. The Borel  $\sigma$ -field  $\mathcal{B} = \sigma(\{(a, b] : 0 \leq a \leq b \leq 1 \text{ rationals}\})$  is countably generated.
9.  $\mathcal{F} = \{A \subset (0, 1] : A \text{ or } A^c \text{ is countable}\}$  is *not* countably generated. To show this, suppose  $\mathcal{F} = \sigma(\{A_1, A_2, \dots\}) = \sigma(\{B_1, B_2, \dots\})$ , where  $B_i = A_i$  if  $A_i$  is countable, or  $B_i = A_i^c$  if  $A_i$  is not countable. In short, we are assuming that  $\mathcal{F}$  is separable, and we will work towards a contradiction. We have constructed every  $B_i$  to be countable, and so we can define  $\Omega_0 = \bigcup_{i=1}^\infty B_i$  that is countable as well. Now  $\mathcal{F} = \sigma(\mathcal{A}_0)$ , where  $\mathcal{A}_0 = \{\{x\} : x \in \Omega_0\}$ . Since  $\mathcal{A}_0 \subset \mathcal{G} = \{B, B \cup \Omega_0^c : B \subset \Omega_0\} \subset \mathcal{F}$ , and  $\mathcal{G}$  is a  $\sigma$ -field, it follows that  $\mathcal{G} = \mathcal{F}$ . But if  $y \in \Omega_0^c$ , then  $\{y\} \notin \mathcal{G}$ . This is a contradiction, and so  $\mathcal{F}$  is not separable.

10. If  $\mathcal{F}_1 \subset \mathcal{F}_2$  and  $\mathcal{F}_2$  is countably generated, then  $\mathcal{F}_1$  need not be countably generated. This is counterintuitive, but it is readily apparent from the previous examples; suppose  $\mathcal{F}_1$  is the countable-cocountable  $\sigma$ -field on  $(0, 1]$  and  $\mathcal{F}_2 = \mathcal{B}$ , the Borel  $\sigma$ -field.

### 1.3 Lecture 3, August 28

**Definition 1.3.1 — Probability.** Let  $\mathcal{F}$  be a field on a nonempty set  $\Omega$ . A set function  $P$  on  $\mathcal{F}$  is called a *probability* if it satisfies

- (I)  $0 \leq P(A) \leq 1$  for all  $A \in \mathcal{F}$
- (II)  $P(\emptyset) = 0$  and  $P(\Omega) = 1$
- (III) *Countable additivity* — If  $A_n$  are disjoint sets in  $\mathcal{F}$ , and  $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$ , then  $P(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n)$
- (IV) *Finite additivity* — For  $A_i \in \mathcal{F}$  disjoint,  $P(\bigcup_{i=1}^n A_i) = \sum_{i=1}^n P(A_i)$
- (V) *Countable subadditivity* — If  $B_n, \bigcup_{n=1}^{\infty} B_n \in \mathcal{F}$  then  $P(\bigcup_{n=1}^{\infty} B_n) \leq \sum_{n=1}^{\infty} P(B_n)$ .

The inequality in (V) is known as *Boole's Inequality*.

Under conditions (I) and (II), condition (III) implies (IV) and (V), and conversely (IV) and (V) together imply (III). The latter will now be proven.

*Proof.* Suppose  $A_n \in \mathcal{F}$  are disjoint,  $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$ , and conditions (IV) and (V) hold. Then

$$\sum_{i=1}^n P(A_i) = P\left(\bigcup_{i=1}^n A_i\right) \leq P\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} P(A_i).$$

Letting  $n \rightarrow \infty$ , we have

$$\sum_{i=1}^{\infty} P(A_i) \leq P\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} P(A_i)$$

and therefore  $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$  as desired. ■

Note that (IV) alone does not guarantee (III). Take a nonempty sample space  $\Omega = \{1, 2, \dots\}$  and  $\mathcal{F}$  to be the finite-cofinite field. Furthermore, define  $P$  to be a function on  $\mathcal{F}$  where

$$P(A) = \begin{cases} 0 & \text{if } A \text{ is finite} \\ 1 & \text{if } A^c \text{ is finite.} \end{cases}$$

It is clear that properties (I) and (II) hold in this case. (IV) also holds: Take  $A, B$  disjoint (so  $A \cap B = \emptyset$ ). If both  $A$  and  $B$  are finite, then the union is also finite, so  $P(A \cup B) = 0 = P(A) + P(B)$ . If one is cofinite, then  $A \cup B$  is cofinite, so  $P(A \cup B) = 1 = P(A) + P(B)$ . If both are cofinite, then it turns out that  $A$  and  $B$  could not have been disjoint in the first place. This is because  $(A \cap B)^c = A^c \cup B^c$  is finite, so  $A \cap B$  must be cofinite and therefore  $A \cap B \neq \emptyset$ . So we don't need to consider this case at all.

On the other hand, define  $A_n = \{n\}$ . The  $A_n$  are disjoint, and  $\bigcup_{n=1}^{\infty} A_n = \Omega \in \mathcal{F}$ . However, each  $P(A_n) = 0$ , so  $\sum_{n=1}^{\infty} P(A_i) = 0$ , yet  $P(\bigcup_{n=1}^{\infty} A_n) = P(\Omega) = 1$ . Countable additivity fails.



If we have a general, non-disjoint group of sets  $B_1, B_2, \dots \in \mathcal{F}$ , we can construct a disjoint group  $A_1, A_2, \dots$  such that the union of all sets in each group are the same. This process of

*disjointification* often proves useful. Define

$$\begin{aligned} A_1 &= B_1 \\ A_2 &= B_2 \cap B_1^c \\ A_3 &= B_3 \cap B_2^c \cap B_1^c \\ &\vdots \\ A_n &= B_n \cap \left( \bigcup_{j < n} B_j \right)^c. \end{aligned}$$

Then  $\bigcup_{i=1}^n A_i = \bigcup_{i=1}^n B_i$ , and the  $A_i$  are disjoint.

We can use disjointification to prove some useful statements. For example, we can show that if  $C \subset D$ , then  $P(C) \leq P(D)$ :

$$\begin{aligned} D &= C \cup (D \cap C^c) && \text{(by disjointification)} \\ P(D) &= P(C) + P(D \cap C^c) \\ &\geq P(C). && \text{(probability is non-negative)} \end{aligned}$$

**Definition 1.3.2 — Continuity from below/above.** If  $A_n \in \mathcal{F}$ ,  $A \in \mathcal{F}$ , and  $A_n \uparrow A$  (meaning that  $A_1 \subset A_2 \subset \dots \subset A_n$  and  $\bigcup_i A_i = A$ ), then  $P(A_n) \uparrow P(A)$ . This is called *continuity from below*.

Similarly, if  $B_n \in \mathcal{F}$ ,  $B \in \mathcal{F}$ , and  $B_n \downarrow B$  (meaning that  $B_1 \supset B_2 \supset \dots \supset B_n$  and  $\bigcap_i B_i = B$ ), then  $P(B_n) \downarrow P(B)$ . This is called *continuity from above*.

Recall the function  $P$  defined earlier:

$$P(A) = \begin{cases} 0 & \text{if } A \text{ is finite} \\ 1 & \text{if } A^c \text{ is finite.} \end{cases}$$

This is *not* a probability on  $\Omega = \{1, 2, \dots\}$ . Continuity from below is not satisfied. To see this, define  $A_n = \{1, 2, \dots, n\}$ .  $A_n$  is finite, so  $P(A_n) = 0$ . Then  $A_n \uparrow \Omega = \{1, 2, 3, \dots\}$ , but  $P(\Omega) = 1$ .

**Theorem 1.3.1** Countable additivity implies  $P(C_n) \downarrow 0$  whenever  $C_n \in \mathcal{F}$  and  $C_n \downarrow \emptyset$ .

*Proof.* Suppose we have a sequence of sets  $C_n$  such that  $C_n \downarrow \emptyset$ . That is,  $C_1 \supset C_2 \supset \dots \supset C_n$ . Construct a sequence of sets  $A_n$  in the following way:

$$\begin{aligned} A_1 &= C_1 - C_2 \\ A_2 &= C_2 - C_3 \\ &\vdots \\ A_n &= C_n - A_{n+1}. \end{aligned}$$

Then the  $A_n$  are disjoint, and  $C_n = \bigcup_{i=n}^{\infty} A_i$ . Note that this means that  $C_1 = \bigcup_{i=1}^{\infty} A_i$ . Now  $P(C_1) = P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i) \leq 1 < \infty$ , by countable additivity. Thus  $P(C_n) = \sum_{i=n}^{\infty} P(A_i) \rightarrow 0$  because it is the tail of a converging series. ■

We will now prove that Theorem 1.3.1 and finite additivity together imply countable additivity.

*Proof.* If we define  $D_n = \bigcup_{i=1}^n A_i$  and  $D = \bigcup_{i=1}^{\infty} A_i$ , then  $D - D_n \downarrow \emptyset$  and  $P(D - D_n) \downarrow 0$ . Now  $P(D) = P(D - D_n) + P(D_n)$ , and  $P(D_n) = \sum_{i=1}^n P(A_i) \rightarrow \sum_{i=1}^{\infty} P(A_i)$ . ■

Furthermore, Theorem 1.3.1 implies continuity from above and continuity from below.

**Definition 1.3.3 — Probability space.** If  $\mathcal{F}$  is a  $\sigma$ -field on  $\Omega$  and  $P$  is a probability measure on  $\mathcal{F}$ , the triple  $(\Omega, \mathcal{F}, P)$  is called a *probability measure space*, or simply a *probability space*.

The following examples illustrate the concept of a probability space.

■ **Example 1.4** Let  $\mathcal{F}$  be the finite-cofinite field on some nonempty set  $\Omega$ , and let  $P$  be defined such that  $P(A) = 0$  or  $1$  depending on whether  $A$  is finite or not. If  $\Omega$  is a countable set, then  $P$  is not a probability (as we saw earlier). However, if  $\Omega$  is uncountable, then  $P$  is a valid probability, and  $(\Omega, \mathcal{F}, P)$  is a probability space. ■

■ **Example 1.5** Let  $\mathcal{F}$  be the countable-cocountable  $\sigma$ -field on an uncountable set  $\Omega$ . Define  $P$  such that  $P(A) = 0$  or  $1$  depending on whether  $A$  is countable or not. Then  $P$  is a probability measure, and  $(\Omega, \mathcal{F}, P)$  is a probability space. ■

■ **Example 1.6** Let  $\Omega = (0, 1]$ , and define a function  $\lambda(a, b) = b - a$ , where  $0 \leq a \leq b \leq 1$ . This is our standard concept of *length* for an interval. Furthermore, the length of the union of disjoint intervals is the sum of the lengths of each individual interval:

$$\lambda \left( \bigcup_{i=1}^k (c_i, d_i] \right) = \sum_{i=1}^k (d_i - c_i), \quad (c_i, d_i] \text{ are disjoint.}$$

Then  $\lambda$  is a (countably additive) probability measure on  $\mathcal{B}_0 = \mathcal{I}$ . ■

## 1.4 Lecture 4, August 31

How can we extend this  $\lambda$  to  $\sigma(\mathcal{I}) = \sigma(\mathcal{B}_0)$ , the Borel  $\sigma$ -field? As previously mentioned, there is not a nice concise way to describe  $\sigma$ -fields. We could define  $\mathcal{B}_\sigma = \{\bigcup_{i=1}^\infty A_i : A_i \in \mathcal{B}_0\}$ , but this is not closed under complementation and so is not a  $\sigma$ -field. Then we might try  $\mathcal{B}_{\sigma\delta} = \{\bigcap_{i=1}^\infty B_i : B_i \in \mathcal{B}_\sigma\}$ , but this still won't work. Even if we tried an infinite sequence  $\mathcal{B}_{\sigma\delta\sigma\delta\dots}$  of these unions and intersections, we still would not be left with a  $\sigma$ -field. The bottom line is that we will need another way to describe them.

**Definition 1.4.1 — Outer measure.** Let  $\Omega$  be a non-empty set.  $P^*$  on  $2^\Omega$  (also known as the power set  $\mathcal{P}(\Omega)$ ) is an *outer measure* if it satisfies the following properties.

- (I)  $P^*(\emptyset) = 0$
- (II)  $P^*(A) \geq 0$
- (III)  $P^*(A) \leq P^*(B)$  whenever  $A \subseteq B$
- (IV)  $P^*(\bigcup_i A_i) \leq \sum_i P^*(A_i)$  — countable sub-additivity

**Notation 1.1.** For convenience, we will sometimes omit the intersection symbol  $\cap$ . For example,  $AE = A \cap E$ .

A good measure should satisfy  $P^*(A) + P^*(A^c) = 1$  on  $\sigma(\mathcal{F}_0)$ . This intuitively makes sense when one thinks about the usual concept of probability. However, it turns out that this property is somewhat difficult to handle in practice. It is easier to work with something like

$$\mathcal{M}' = \{A \subset \Omega : P^*(AE) + P^*(A^cE) = P^*(E) \text{ for all } E \subset \Omega\}.$$

Inequalities are even more useful, so we relax this to

$$\mathcal{M} = \{A \subset \Omega : P^*(AE) + P^*(A^cE) \leq P^*(E) \text{ for all } E \subset \Omega\}.$$

We will now present and prove some facts about this particular collection based on an outer measure.

**Theorem 1.4.1** Define  $\mathcal{M} = \mathcal{M}(P^*) = \{A \subset \Omega : P^*(AE) + P^*(A^cE) \leq P^*(E) \text{ for all } E \subset \Omega\}$ , where  $P^*$  is an outer measure on  $\mathcal{F}_0$ . Then

- (a)  $\mathcal{M}$  is a field,
- (b) If  $A_i \in \mathcal{M}$  are disjoint, then  $P^*(E \cap (\bigcup_i A_i)) = \sum_i P^*(EA_i)$ ,
- (c)  $\mathcal{M}$  is a  $\sigma$ -field and  $P^*$  is countably additive on  $\mathcal{M}$ .
- (d)  $\mathcal{F}_0 \subset \mathcal{M}$
- (e)  $P^*$  agrees with  $P$  everywhere on  $\mathcal{F}_0$ .

*Proof.* The proofs for (a), (b), and (c) depend only on properties (I)-(IV) in Definition 2.4.1.

First we will show (a). Recall that to prove  $\mathcal{M}$  is a field, we need to verify that  $\emptyset, \Omega \in \mathcal{M}$ , that  $\mathcal{M}$  is closed under complementation, and that it is closed under finite unions. Inspection of the definition of  $\mathcal{M}$  will reveal that it is symmetric about  $A$ ; we can switch  $A$  with  $A^c$  with no change. Therefore  $\mathcal{M}$  is trivially closed under complementation. Also,  $P^*(\emptyset \cap E) + P^*(\Omega \cap E) = P^*(E)$ , so  $\emptyset, \Omega \in \mathcal{M}$ .

Now we only need to show closure under finite unions. For any  $A, B \in \mathcal{M}$ ,  $(AB)^c = AB^c \cup A^cB$  (this is easy to verify using a Venn diagram).

$$\begin{aligned}
 P^*(ABE) + P^*((AB)^cE) &\leq P^*(ABE) + P^*(AB^cE) + P^*(A^cB^cE) + P^*(A^cBE) && \text{(by property (IV))} \\
 &\leq P^*(ABE) + P^*(B^cE) + P^*(A^cBE) && \text{(because } A \in \mathcal{M}) \\
 &\leq P^*(B^cE) + P^*(BE) && \text{(because } A \in \mathcal{M}) \\
 &\leq P^*(E). && \text{(because } B \in \mathcal{M})
 \end{aligned}$$

Therefore  $AB \in \mathcal{M}$ . We have shown that  $\mathcal{M}$  is closed under finite intersection. If  $A \cap B \in \mathcal{M}$ , then so is  $(A \cap B)^c = A^c \cup B^c$ , and so is  $A \cup B$ . Thus  $\mathcal{M}$  is closed under finite union as well. We can conclude that  $\mathcal{M}$  is a field.

Next, we will prove (b). Suppose  $A_1, A_2 \in \mathcal{M}$  and  $A_1$  and  $A_2$  are disjoint. Then

$$\begin{aligned}
 P^*(E(A_1 \cup A_2)) &= P^*(E(A_1 \cup A_2)A_1) + P^*(E(A_1 \cup A_2)A_1^c) \\
 &= P^*(EA_1) + P^*(EA_2). && \text{(because } A_2 \subset A_1^c)
 \end{aligned}$$

By induction, for disjoint  $A_i \in \mathcal{M}$  we have

$$P^*\left(E \cap \left(\bigcup_{i=1}^n A_i\right)\right) = \sum_{i=1}^n P^*(EA_i).$$

Because  $P^*(A) \geq 0$  for any  $A$ , It follows that  $P^*(E \cap (\bigcup_{i=1}^\infty A_i)) \geq P^*(E \cap (\bigcup_{i=1}^n A_i))$ . Taking the limit leads us to conclude that

$$P^*\left(E \cap \left(\bigcup_{i=1}^\infty A_i\right)\right) \geq P^*\left(E \cap \left(\bigcup_{i=1}^n A_i\right)\right) = \sum_{i=1}^n P^*(EA_i) \rightarrow \sum_{i=1}^\infty P^*(EA_i).$$

But property (IV), countable subadditivity, tells us that  $P^*(E \cap (\bigcup_{i=1}^\infty A_i)) \leq \sum_{i=1}^\infty P^*(EA_i)$ . Therefore

$$P^*\left(E \cap \left(\bigcup_{i=1}^\infty A_i\right)\right) = \sum_{i=1}^\infty P^*(EA_i)$$

as desired.

Finally we show (c). We already know  $\mathcal{M}$  is a field, so to show it is a  $\sigma$ -field all we need is closure under countable unions. Let  $B_1, B_2, \dots \in \mathcal{M}$ , and  $B = \bigcup_{i=1}^{\infty} B_i$ . Disjointify the  $B_i$ s, so define  $A_1 = B_1, \dots, A_n = B_n \cap (\bigcup_{i < n} B_i)^c$ . Then all  $A_i \in \mathcal{M}$ , the  $A_i$  are disjoint, and the unions are the same:  $\bigcup_{i=1}^n A_i = \bigcup_{i=1}^n B_i \uparrow B$ .

$$\begin{aligned} P^*(E) &= P^*\left(E \cap \left(\bigcup_{i=1}^n A_i\right)\right) + P^*\left(E \cap \left(\bigcup_{i=1}^n A_i\right)^c\right) \\ &\geq \sum_{i=1}^n P^*(EA_i) + P^*(EB^c) \rightarrow \sum_{i=1}^{\infty} P^*(EA_i) + P^*(EB^c) && \text{(taking limit as } n \rightarrow \infty) \\ &\geq P^*(EB) + P^*(EB^c). && \text{(by property (IV))} \end{aligned}$$

So  $B = \bigcup_{i=1}^{\infty} B_i \in \mathcal{M}$ , and hence  $\mathcal{M}$  is a  $\sigma$ -field. We can easily see that  $P^*$  is countably additive on  $\mathcal{M}$  by taking  $E = \Omega$  in (b). ■

Again, to this point we have used only properties (I)-(IV) in Definition 2.4.1.

## 1.5 Lecture 5, September 2

Now, assume that  $P$  is a probability measure on a field  $\mathcal{F}_0$  of subsets of a nonempty set  $\Omega$ . Define

$$P^*(E) = \inf \left\{ \sum_i P(A_i) : E \subset \bigcup_i A_i, \text{ and } A_i \in \mathcal{F}_0 \right\}.$$

This  $P^*$  satisfies all four properties of Definition 2.4.1.

*Proof.* Property (I) is clearly satisfied by letting all the  $A_i = \emptyset$ . Also,  $P$  is nonnegative on  $\mathcal{F}_0$ , so  $P^*$  must also be nonnegative, and thus property (II) is satisfied. This non-negativity is why property (III) is satisfied as well.

Property (IV) is a bit harder to show. Let  $A_n \subset \Omega$  and fix some  $\varepsilon > 0$ . Then

$$P^*(A_n) = \inf \left\{ \sum_k P(C_{nk}) : A_n \subset \bigcup_k C_{nk}, C_{nk} \in \mathcal{F}_0 \right\}.$$

Now there is some collection  $B_{nk} \in \mathcal{F}_0$  such that  $A_n \subset \bigcup_k B_{nk}$  and  $\sum_k P(B_{nk}) < P^*(A_n) + \varepsilon 2^{-n}$ . Since  $\bigcup_n A_n \subset \bigcup_n \bigcup_k B_{nk}$ , we have

$$P^*\left(\bigcup_n A_n\right) \leq \sum_n \sum_k P(B_{nk}) \leq \sum_n P^*(A_n) + \varepsilon.$$

$\varepsilon$  can be taken to be arbitrarily small, and thus we have countable sub-additivity.  $P^*$  is therefore an outer measure. ■

Next we will establish the remaining two facts from Theorem 1.4.1.

- $\mathcal{F}_0 \subset \mathcal{M}$
- $P^*$  agrees with  $P$  everywhere on  $\mathcal{F}_0$ .

■ **Example 1.7 — Exercise 3.3(a).** Consider the following scenario: Let  $\Omega = \{1, 2, 3\}$ . The power set of  $\Omega$  is  $2^\Omega = \{\emptyset, \Omega, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}\}$ . Now let  $P_1(\{2, 3\}) = 1, P_1(\{1\}) = 0$ , and  $P_2(\{2, 3\}) = 0, P_2(\{1\}) = 1$  be two probability measures on the field  $\mathcal{F}_0 = \{\emptyset, \Omega, \{1\}, \{2, 3\}\}$ .

First we will establish  $\mathcal{F}_0 \subset \mathcal{M}(P_1^*)$ . If  $A = \{2\}$ , then  $P_1^*(A\Omega) = P_1^*(A) = P_1^*(\{2, 3\}) = 1$  and  $P_1^*(A^c\Omega) = P_1^*(A^c) = P_1^*(\{1, 3\}) = 1$ . But  $P_1^*(\Omega) = 1$ , and  $1 + 1 > 1$ . This means that  $\{2\} \notin \mathcal{M}(P_1^*)$

and  $\{1, 3\} \notin \mathcal{M}(P_1^*)$ . Similarly, we can check that  $\{3\}, \{1, 2\} \notin \mathcal{M}(P_1^*)$ . The only sets left from  $2^\Omega$  are the ones that make up  $\mathcal{F}_0$ , so  $\mathcal{F}_0 = \mathcal{M}(P_1^*)$ .

Now we focus on  $P_2^*$ . For a particular  $A$  and any  $E$ , property (III) of Definition 2.4.1 guarantees that  $P_2^*(AE) \leq P_2^*(A)$  because  $A \cap E \subseteq A$ . Suppose  $1 \notin A$ . If this is the case, then  $P_2^*(A) = 0$  (since the only two possibilities for  $A$  in  $\mathcal{F}$  are  $\emptyset$  and  $\{2, 3\}$ , and  $P_2^*$  is zero on each of these). This establishes that  $P_2^*(AE) = 0$ . It is also true that  $P_2^*(A^cE) \leq P_2^*(E)$ , also by property (III) of Definition 2.4.1. Putting these together,

$$P_2^*(AE) + P_2^*(A^cE) \leq 0 + P_2^*(E)$$

which indicates that  $A \in \mathcal{M}(P_2^*)$  and, as a direct result,  $A^c \in \mathcal{M}(P_2^*)$ . This covers every possible set  $A$ , and therefore  $\mathcal{M}(P_2^*) = 2^\Omega$ . ■

We will now prove that  $\mathcal{F}_0 \subset \mathcal{M}$ .

*Proof.* Let  $A \in \mathcal{F}_0$ , and  $E \subset \Omega$ . Let  $A_i \in \mathcal{F}_0$  such that  $E \subset \bigcup_{i=1}^\infty A_i$ . Note that  $AE \subset \bigcup_{i=1}^\infty AA_i$  and  $A^cE \subset \bigcup_{i=1}^\infty A^cA_i$ . So

$$P^*(EA) + P^*(EA^c) \leq \sum_{i=1}^\infty P(AA_i) + \sum_{i=1}^\infty P(A^cA_i) = \sum_{i=1}^\infty P(A_i).$$

So  $P^*(EA) + P^*(EA^c) \leq P^*(E)$ , and thus  $A \in \mathcal{M}$ . ■

Only the finite additivity of  $P$  was used above. Next we will show that  $P^* = P$  on  $\mathcal{F}_0$ .

*Proof.* We have already seen that  $P^*(A) \leq P(A)$  if  $A \in \mathcal{F}_0$ . If we have some  $A_i \in \mathcal{F}_0$  such that  $\bigcup_{i=1}^\infty A_i \supset A$ , then

$$P(A) \leq \sum_{i=1}^\infty P(AA_i) \leq \sum_{i=1}^\infty P(A_i)$$

by the countable sub-additivity of  $P$  on  $\mathcal{F}_0$ . Notice that we cannot say directly that  $P(A) \leq \sum_{i=1}^\infty P(A_i)$  because there is no guarantee that  $\bigcup_{i=1}^\infty A_i \in \mathcal{F}_0$ . The intermediate step is therefore necessary.

The above gives us that  $P(A) \leq P^*(A)$ . Therefore  $P^*$  restricted to  $\mathcal{F}_0$  agrees with  $P$ , as desired. ■

This leads us to an important result.

**Theorem 1.5.1 — Extension Theorem.** Every probability measure on a field  $\mathcal{F}_0$  has an extension to  $\sigma(\mathcal{F}_0)$ .

*Proof.*  $\mathcal{M}$  is a  $\sigma$ -field containing  $\mathcal{F}_0$ , so  $\sigma(\mathcal{F}_0) \subset \mathcal{M}$ . Furthermore,  $P^*$  is a probability measure on  $\mathcal{M}$ , and hence  $P^*$  restricted to  $\sigma(\mathcal{F}_0)$  is a probability measure as well. Finally,  $P^*(A) = P(A)$  for all  $A \in \mathcal{F}_0$ . ■

When we restrict  $P^*$  to  $\sigma(\mathcal{F}_0)$ , then  $P^*$  is defined on all subsets of  $\Omega$ . Define

$$\begin{aligned} P_0(A) &= P^*(A) & \text{if } A \in \mathcal{M} \\ P_1(A) &= P^*(A) & \text{if } A \in \sigma(\mathcal{F}_0) \\ P_2(A) &= P^*(A) & \text{if } A \in \mathcal{F}_0. \end{aligned}$$

$P_0$  is not a probability measure on  $\sigma(\mathcal{F}_0)$  or  $\mathcal{F}_0$ , for example. The values are the same (and are equal to the values of  $P^*$ ), but the *domains*, the collections of sets on which they are defined, differ.

Is it possible to have two probability measures, say  $Q_1$  and  $Q_2$ , on  $\sigma(\mathcal{F}_0)$  such that  $Q_1 = Q_2$  on  $\mathcal{F}_0$ ? We will now work towards a uniqueness theorem. First some definitions.

**Definition 1.5.1 —  $\pi$ -system.** A class  $\mathcal{P}$  is called a  $\pi$ -system if  $A, B \in \mathcal{P}$  implies  $AB \in \mathcal{P}$  (closed under intersection).

**Definition 1.5.2 —  $\lambda$ -system.** A class  $\mathcal{L}$  is called a  $\lambda$ -system if it satisfies the following conditions:

- ( $\lambda 1$ )  $\Omega \in \mathcal{L}$
- ( $\lambda 2$ )  $A \in \mathcal{L}$  implies  $A^c \in \mathcal{L}$
- ( $\lambda 3$ ) If  $A_1, A_2, \dots \in \mathcal{L}$  and are disjoint, then  $\bigcup_i A_i \in \mathcal{L}$ . (Closed under countable disjoint unions)

A  $\lambda$ -system is “almost” a  $\sigma$ -field. The only difference is the additional restriction of disjointness in property ( $\lambda 3$ ). Why is this distinction made? Suppose we have  $P_1$  and  $P_2$  such that for all  $A_i$ ,  $P_1(A_i) = P_2(A_i)$ . We can’t really say that

$$P_1 \left( \bigcup_{i=1}^{\infty} A_i \right) = P_2 \left( \bigcup_{i=1}^{\infty} A_i \right).$$

We need disjointness to prove uniqueness.

■ **Example 1.8 —  $\lambda$ -system but not a  $\sigma$ -field.** Suppose  $\Omega = \{1, 2, 3, 4\}$ . Let

$$\mathcal{L} = \{\emptyset, \Omega, \{1, 2\}, \{2, 3\}, \{3, 4\}, \{1, 4\}\}.$$

Notice that  $\mathcal{L}$  is of the form  $\mathcal{L} = \{\emptyset, \Omega, A, B, A^c, B^c\}$ , where  $\emptyset \neq A \neq B \neq \Omega$ ,  $A \neq B^c$ , and  $A \cap B \neq \emptyset$ .

■

**Theorem 1.5.2** A class  $\mathcal{F}$  that is both a  $\pi$ -system and a  $\lambda$ -system, is a  $\sigma$ -field.

*Proof.* We only need to show that  $\mathcal{F}$  is closed under countable unions. Let  $B_i \in \mathcal{F}$ . Break the  $B_i$  into disjoint sets in the usual way:

$$A_1 = B_1, \dots, A_n = B_n \cap \left( \bigcup_{i < n} B_i \right)^c.$$

Then  $A_i$  are disjoint and  $\bigcup_i A_i = \bigcup_i B_i$ . Since  $B_i \in \mathcal{F}$ , we have  $B_i^c \in \mathcal{F}$  by ( $\lambda 2$ ).

Because  $\mathcal{F}$  is a  $\pi$ -system, we have  $A_n = B_n \cap (B_1^c \cdots B_{n-1}^c) \in \mathcal{F}$ . So by ( $\lambda 3$ ),  $\bigcup_i A_i \in \mathcal{F}$ . Hence  $\mathcal{F}$  is a  $\sigma$ -field. ■

## 1.6 Lecture 6, September 4

**Theorem 1.6.1** If  $\mathcal{P}$  is a  $\pi$ -system,  $\mathcal{L}$  is a  $\lambda$ -system, and  $\mathcal{P} \subset \mathcal{L}$ , then  $\sigma(\mathcal{P}) \subset \mathcal{L}$ .

*Proof.* Let  $\mathcal{L}_0$  be the smallest  $\lambda$ -system containing  $\mathcal{P}$ . As  $\mathcal{L}_0 \subset \mathcal{L}$ , by Theorem 1.5.1 it is enough to show that  $\mathcal{L}_0$  is a  $\pi$ -system. For any fixed  $A \subset \Omega$  (where  $A$  is not necessarily in  $\mathcal{L}_0$ ), define

$$\mathcal{L}_A = \{B \in \mathcal{L}_0 : AB \in \mathcal{L}_0\}.$$

Note that the intersection of  $\lambda$ -systems is itself a  $\lambda$ -system. In the context of collections of sets, “intersection” means sets that are common to both.

**Part 1.** First we will show that if  $A \in \mathcal{L}_0$ , then  $\mathcal{L}_A$  is a  $\lambda$ -system.

1. If  $A \in \mathcal{L}_0$ , then  $A \cap \Omega \in \mathcal{L}_0$ . So  $\Omega \in \mathcal{L}_A$ , verifying property ( $\lambda 1$ ).



2. If  $A \in \mathcal{L}_0$ , then  $A^c \in \mathcal{L}_0$ . Furthermore, if  $B \in \mathcal{L}_A$ , then  $B \in \mathcal{L}_0$ . Now consider the sets  $AB$  and  $A^c$ . These are disjoint, and each is in  $\mathcal{L}_0$ . As a result,  $(AB \cup A^c)^c \in \mathcal{L}_0$ , and  $(AB \cup A^c)^c = A \cap (AB)^c = B^c A \in \mathcal{L}_0$ . We know that  $B^c \in \mathcal{L}_0$ , and so this implies that  $B^c \in \mathcal{L}_A$ . Therefore  $\mathcal{L}_A$  is closed under complementation, verifying property  $(\lambda 2)$ .
  3. If  $A \in \mathcal{L}_0$  and  $B_n \in \mathcal{L}_0$  are disjoint, then  $AB_n \in \mathcal{L}_0$  and are disjoint. So  $\bigcup_n B_n \in \mathcal{L}_0$  and  $A \cap \bigcup_n B_n \in \mathcal{L}_0$ . This implies that  $\bigcup_n B_n \in \mathcal{L}_A$ , verifying property  $(\lambda 3)$ .
- Thus if  $A \in \mathcal{L}_0$ , then  $\mathcal{L}_A$  is a  $\lambda$ -system.

**Part 2.** We will now show that  $\mathcal{L}_0$  is a  $\pi$ -system. There are three possibilities here:

1.  $A \in \mathcal{P}, B \in \mathcal{P}$
2.  $A \in \mathcal{P}, B \in \mathcal{L}_0$
3.  $A \in \mathcal{L}_0, B \in \mathcal{L}_0$ .

We must show that in each case,  $A \cap B \in \mathcal{L}_0$ .

First, fix  $A \in \mathcal{P}$ . Then  $B \in \mathcal{P}$  implies  $AB \in \mathcal{P} \subset \mathcal{L}_0$ . So  $B \in \mathcal{L}_A$  and hence  $\mathcal{P} \subset \mathcal{L}_A$ . Now, we showed in Part 1 above that  $\mathcal{L}_A$  is a  $\lambda$ -system since  $A \in \mathcal{L}_0$ . To this point we have showed that  $A \in \mathcal{P}$  implies  $\mathcal{L}_0 \subset \mathcal{L}_A$ . Since we already knew that  $\mathcal{L}_A \subset \mathcal{L}_0$ , this actually means  $A \in \mathcal{P} \implies \mathcal{L}_A = \mathcal{L}_0$ .

From the definition of  $\mathcal{L}_A$ , if  $A \in \mathcal{P}$  and  $B \in \mathcal{L}_0$ , then  $AB \in \mathcal{L}_0$ . This is exactly what we needed for case 2.

Now fix  $B \in \mathcal{L}_0$ . If  $A \in \mathcal{P}$ , then  $AB \in \mathcal{L}_0$ . This is simply a restatement of the above conclusion. This implies that  $A \in \mathcal{L}_B = \{C \in \mathcal{L}_0 : CB \in \mathcal{L}_0\}$ . So  $\mathcal{P} \subset \mathcal{L}_B$ , which immediately implies  $\mathcal{L}_0 \subset \mathcal{L}_B$  whenever  $B \in \mathcal{L}_0$ . Hence  $\mathcal{L}_0 = \mathcal{L}_B$  for all  $B \in \mathcal{L}_0$ .

The conclusion is if  $A \in \mathcal{L}_0$  and  $B \in \mathcal{L}_0$ , then  $AB \in \mathcal{L}_0$ , which is case 3 above.

So  $\mathcal{L}_0$  is a  $\pi$ -system and a  $\lambda$ -system, and thus by Theorem 1.5.2 it is a  $\sigma$ -field. Also it contains  $\mathcal{P}$ , so  $\mathcal{P} \subset \mathcal{L}_0 \subset \mathcal{L}$ . ■

**Theorem 1.6.2 — Uniqueness Theorem.** Suppose  $P_1$  and  $P_2$  are probability measures on  $\sigma(\mathcal{P})$ , where  $\mathcal{P}$  is a  $\pi$ -system. If  $P_1$  and  $P_2$  agree on  $\mathcal{P}$ , then they agree on  $\sigma(\mathcal{P})$ .

*Proof.* Let  $\mathcal{L} = \{A \in \sigma(\mathcal{P}) : P_1(A) = P_2(A)\}$ . This  $\mathcal{L}$  is a  $\lambda$ -system (conditions are straightforward to verify), and it contains the  $\pi$ -system  $\mathcal{P}$ . By Theorem 1.6.1,  $\sigma(\mathcal{P}) \subset \mathcal{L}$  as required. ■

**Definition 1.6.1 — Support.**  $S \in \mathcal{F}$  is called a *support* of  $P$  if  $P(S) = 1$ .

**Definition 1.6.2 — Null-set.** A set with probability measure zero is called a *null-set*.

**Definition 1.6.3 — Completeness.** A probability measure space  $(\Omega, \mathcal{F}_0, P_0)$  is *complete* if

$$A \subset B, B \in \mathcal{F}_0, \text{ and } P_0(B) = 0 \implies A \in \mathcal{F}_0.$$

That is, if there is a null-set in  $\mathcal{F}_0$ , all its subsets are also in  $\mathcal{F}_0$ .

The Borel  $\sigma$ -field  $\mathcal{B}$  is *not* complete. There exist uncountable subsets which are still measure zero. Cantor sets, on the other hand, are complete.

## 1.7 Lecture 7, September 9

It is always possible to complete a probability space.

**Theorem 1.7.1** For every probability space, there exists a unique minimal complete extension.

*Proof.* For any probability space  $(\Omega, \mathcal{F}, P)$ , define the outer measure  $P^*$  using  $\mathcal{F}_0$  such that  $\mathcal{F} = \sigma(\mathcal{F})$ . Then  $P^*$  restricted to  $\mathcal{M}(P^*)$  is a probability measure: If  $P^*(B) = 0$  and  $A \subset B$ , then  $P^*(A) = 0$  and  $P^*(A \cap E) + P^*(A^c \cap E) \leq P^*(B) + P^*(E) = P^*(E)$ , and so  $A \in \mathcal{M}(P^*)$ .

Therefore  $(\Omega, \mathcal{M}(P^*), P^*)$  is a complete probability space.  $\blacksquare$


The Borel  $\sigma$ -field on  $(0, 1]$ ,  $\mathcal{B}$ , together with the Lebesgue measure  $\lambda$ , can be completed this way. The sets in  $\mathcal{M}(\lambda^*)$  are called *Lebesgue sets*.  $\lambda^*$  is still called *Lebesgue measure*, and is often denoted simply by  $\lambda$ . On  $\Omega = (0, 1]$ ,

$$\mathcal{I} \subset \mathcal{B}_0 \subset \mathcal{B} \subset \mathcal{M}(\lambda^*).$$

In terms of the classification of the above, this is  $(\pi\text{-system}) \subset (\text{field}) \subset (\sigma\text{-field}) \subset (\text{complete Lebesgue } \sigma\text{-field})$ .

For any probability space  $(\Omega, \mathcal{F}, P)$ , define

$$\mathcal{F}^+ = \{A : A \triangle B \subset C \text{ for some } B, C \in \mathcal{F} \text{ satisfying } P(C) = 0\}.$$

 We use  $\triangle$  as the *symmetric difference* operator.

$$A \triangle B = (A \cap B^c) \cup (A^c \cap B).$$

Notice that  $A^c \triangle B^c = A \triangle B$ .

Notice that  $A \in \mathcal{F}^+$  implies  $A^c \in \mathcal{F}^+$ , since  $A^c \triangle B^c = A \triangle B \subset C$ . If  $A \in \mathcal{F}$ , then taking  $B = A$  and  $C = \emptyset$  shows that  $\mathcal{F} \subset \mathcal{F}^+$ . Finally,  $\mathcal{F}^+$  is closed under countable unions: For  $A_i \in \mathcal{F}^+$ , we can get  $B_i, C_i \in \mathcal{F}$  such that  $A_i \triangle B_i \subset C_i$ .

$$\begin{aligned} \left( \bigcup_{i=1}^{\infty} A_i \right) \cap \left( \bigcup_{j=1}^{\infty} B_j \right)^c &\subset \bigcup_{i=1}^{\infty} (A_i \cap B_i^c) \\ \left( \bigcup_{i=1}^{\infty} B_i \right) \cap \left( \bigcup_{j=1}^{\infty} A_j \right)^c &\subset \bigcup_{i=1}^{\infty} (B_i \cap A_i^c) \\ \left( \bigcup_{i=1}^{\infty} A_i \right) \triangle \left( \bigcup_{j=1}^{\infty} B_j \right) &\subset \bigcup_{i=1}^{\infty} (A_i \triangle B_i) \subset \bigcup_{i=1}^{\infty} C_i. \end{aligned}$$

So  $\mathcal{F}^+$  is a  $\sigma$ -field on  $\Omega$  and  $\mathcal{F} \subset \mathcal{F}^+$ .

We want to extend  $P$  from  $\mathcal{F}$  onto  $\mathcal{F}^+$ . For  $A \in \mathcal{F}^+$ , there exist some  $B, C \in \mathcal{F}$  satisfying  $P(C) = 0$  and  $A \triangle B \subset C$ . This is directly from the definition of  $\mathcal{F}^+$ . If we have another two sets  $D$  and  $E$  from  $\mathcal{F}$  that satisfy  $A \triangle D \subset E$  and  $P(E) = 0$ , then  $B \triangle D \subset C \cup E$ , and hence  $P(B) = P(D)$ . Define  $P^+(A) = P(B)$ . This  $P^+$  is an unambiguous probability measure on  $\mathcal{F}^+$ .

This completes the derivation of  $(\Omega, \mathcal{F}^+, P^+)$ , the minimal complete extension of  $(\Omega, \mathcal{F}, P)$ .

We now turn our attention to  $\lambda$ , the Lebesgue measure on the Borel  $\sigma$ -field  $\mathcal{B}$ . Define the  $\oplus$  operator as follows:

$$x \oplus y = \begin{cases} x + y & \text{if } x + y \in (0, 1] \\ x + y - 1 & \text{otherwise.} \end{cases}$$

The  $\oplus$  operator is also defined with respect to a set and a scalar.  $A \oplus x = \{a \oplus x : a \in A\}$ . Now consider the set

$$\mathcal{L} = \{A \in \mathcal{B} : A \oplus x \in \mathcal{B} \text{ and } \lambda(A \oplus x) = \lambda(A)\}.$$

Since  $(A \oplus x)^c = A^c \oplus x$ ,  $\mathcal{L}$  is a  $\lambda$ -system containing  $\mathcal{I}$ . By Theorem 1.6.1,  $\mathcal{L} = \mathcal{B}$ . Shifting all sets by  $x$  did not their measure. In this sense,  $\lambda$  is *translation invariant* on  $\mathcal{B}$ . In fact, we can go even further.

**Theorem 1.7.2** Lebesgue measure  $\lambda^*$  is translation invariant on all subsets of  $\Omega$ .

*Proof.* Let  $\mathcal{B}_0$  be the Borel field generated by  $\mathcal{I}$ . As we have just seen, for  $A \in \mathcal{B}_0$ ,  $A \oplus x \in \mathcal{B}_0$  and  $\lambda(A \oplus x) = \lambda(A)$ . If  $B \subset \bigcup_{i=1}^{\infty} A_i$  for  $A_i \in \mathcal{B}_0$ , then  $B \oplus x \subset \bigcup_{i=1}^{\infty} (A_i \oplus x)$ . Hence

$$\lambda^*(B \oplus x) \leq \sum_{i=1}^{\infty} \lambda(A_i \oplus x) = \sum_{i=1}^{\infty} \lambda(A_i),$$

which implies that  $\lambda^*(B \oplus x) \leq \lambda^*(B)$ . As  $B = (B \oplus x) \oplus (1 - x)$ , it follows that  $\lambda^*(B) \leq \lambda^*(B \oplus x)$ . So in fact  $\lambda^*(B) = \lambda^*(B \oplus x)$ . Thus  $\lambda^*$  is translation invariant on all subsets of  $\Omega$ . ■

Just because  $\lambda^*$  is translation invariant on all subsets of  $\Omega$  does not mean that it is a probability measure on all subsets of  $\Omega$ . It is only a probability measure on sets belonging to a certain class, appropriately named *Lebesgue-measurable sets*.

**Definition 1.7.1 — Lebesgue-measurable.** The sets in  $\mathcal{M} = \mathcal{M}(\lambda^*)$  are called *Lebesgue-measurable sets*.  $\lambda^*$  (called *Lebesgue measure*) is a probability measure on  $\mathcal{M}$ .

It can be proven that  $\lambda^*$  is translation-invariant on  $\mathcal{M}$  as follows. If  $A \in \mathcal{M}$ , then  $\lambda^*(AE) + \lambda^*(A^cE) \leq \lambda^*(E)$  for all  $E \subset \Omega$  because  $\lambda^*$  is an outer measure. For any  $B \subset \Omega$ ,

$$(B \oplus x) \cap E = (B \oplus x) \cap ((E \oplus (1 - x)) \oplus x) = (B \cap (E \oplus (1 - x))) \oplus x.$$

It follows that

$$\lambda^*((B \oplus x) \cap E) = \lambda^*(B \cap (E \oplus (1 - x))) \oplus x = \lambda^*(B \cap (E \oplus (1 - x))).$$

So

$$\lambda^*((A \oplus x) \cap E) + \lambda^*((A \oplus x)^c \cap E) \leq \lambda^*(E \oplus (1 - x)) = \lambda^*(E).$$

Hence  $A \oplus x \in \mathcal{M}$ . This establishes that the Lebesgue measure on  $\mathcal{M}$  is *translation invariant*.

Define the relation  $\sim$  such that  $x \sim y$  if  $x \oplus r = y$  for some rational  $r \in (0, 1]$ . This relation partitions  $\Omega$  into equivalence classes  $\{A_\theta : \theta \in \Theta\}$ . Construct a set  $H$  consisting of exactly one point from each class  $A_\theta$ . Further define  $Q$  to be the set of rationals in  $\Omega$  and  $H_r = H \oplus r$ .

Then  $\bigcup_{r \in Q} H_r = \Omega$ , and  $H_r \cap H_s = \emptyset$  for  $r \neq s \in Q$ . That is, the  $H$ 's are disjoint. To see this, consider a particular  $x \in \Omega$ . Because  $\bigcup_{\theta} A_\theta = \Omega$ , there is some  $\theta$  such that  $x \in A_\theta$ . Choose the  $y$  from that same  $A_\theta$  where  $y \in H$ . Now suppose  $x \in H_r \cap H_s$ . Then  $x = y + r$ ,  $y \in H$ , and  $x = y' + s$ ,  $y' \in H$ . This is a contradiction, since we built  $H$  by taking exactly one element from each  $A_\theta$ . Thus  $H_r \cap H_s = \emptyset$ .

Now if  $P$  is a translation invariant probability measure on all subsets of  $\Omega$ , then

$$1 = P\left(\bigcup_{r \in Q} H_r\right) = \sum_{r \in Q} P(H_r) = \infty \cdot P(H_{(1/2)})$$

since this is a countably infinite sum. Clearly we have reached a contradiction. Therefore, *there is no translation-invariant probability measure on  $2^{(0,1]}$  (all subsets of  $(0, 1]$ )*. This also implies that  $\mathcal{M}(\lambda^*) \neq 2^{(0,1]}$ , that is, there exists some set that is not Lebesgue-measurable.

**Definition 1.8.1 — Monotone class.** A class  $\mathcal{M}$  of subsets of  $\Omega$  is called a *monotone class* if it is closed under countable monotone limits. That is,

- (i)  $A_1, A_2, \dots \in \mathcal{M}$  and  $A_n \uparrow A$  imply  $A \in \mathcal{M}$  and
- (ii)  $A_1, A_2, \dots \in \mathcal{M}$  and  $A_n \downarrow A$  imply  $A \in \mathcal{M}$ .

Notice that a class containing only one set is automatically monotone. For example,  $\mathcal{M} = \{\emptyset\}$  satisfies the conditions in Definition 1.8.1.

**Theorem 1.8.1** If  $\mathcal{F}$  is a field and  $\mathcal{M}$  is a monotone class, then  $\mathcal{F} \subset \mathcal{M}$  implies  $\sigma(\mathcal{F}) \subset \mathcal{M}$ .

*Proof.* First we will show that a class  $\mathcal{M}$  that is both monotone and a field is a  $\sigma$ -field. Take  $c_1, c_2, \dots \in \mathcal{M}$ . Then  $A_n = \bigcup_{i=1}^n c_i \in \mathcal{M}$  (fields contain finite unions). But  $A_n \uparrow \bigcup_{i=1}^{\infty} c_i \in \mathcal{M}$  because  $\mathcal{M}$  is closed under countable monotone limits. Thus  $\mathcal{M}$  is a  $\sigma$ -field.

Now let  $\mathcal{M}_0$  be the smallest monotone class containing  $\mathcal{F}$ . Then  $\mathcal{F} \subset \mathcal{M}_0 \subset \mathcal{M}$ . Thus it is enough to show that  $\mathcal{M}_0$  is a field.

Define

$$\mathcal{M}_A = \{B \in \mathcal{M}_0 : A \cap B, A \cap B^c \text{ and } A^c \cap B \in \mathcal{M}_0\}.$$

Obviously  $\mathcal{M}_A \subset \mathcal{M}_0$ . Fix  $A \in \mathcal{F}$ . We want to show that  $\mathcal{M}_A$  is a monotone class. To do that, let  $B_n \in \mathcal{M}_A$  such that  $B_n \uparrow B$ . From the definition of  $\mathcal{M}_A$  we know that

$$\left. \begin{array}{l} A \cap B_n \\ A \cap B_n^c \\ A^c \cap B_n \\ B_n \end{array} \right\} \in \mathcal{M}_0.$$

Taking increasing limits,  $B_n \uparrow B \in \mathcal{M}_0$ ,  $A \cap B_n \uparrow A \cap B \in \mathcal{M}_0$ ,  $A^c \cap B_n \uparrow A^c \cap B \in \mathcal{M}_0$ , and  $A \cap B_n^c \downarrow A \cap B^c \in \mathcal{M}_0$ . All this means that  $B \in \mathcal{M}_A$ . So  $\mathcal{M}_A$  is indeed monotone. In fact, this is true for any  $A$ , not just those in  $\mathcal{F}$ . (The above argument can be modified slightly to show that if  $C_n \in \mathcal{M}_A$  and  $C_n \downarrow C$ , then  $C \in \mathcal{M}_A$ .)

The next step is to show that if  $A \in \mathcal{F}$ , then  $\mathcal{F} \subset \mathcal{M}_A$ . Take  $A, B \in \mathcal{F}$ . Then by definition of  $\mathcal{M}_A$ ,  $B \in \mathcal{M}_A$ . Hence  $\mathcal{M}_0 \subset \mathcal{M}_A$  because  $\mathcal{M}_0$  is minimal. However, we already saw that  $\mathcal{M}_A \subset \mathcal{M}_0$ . Therefore  $\mathcal{M}_0 = \mathcal{M}_A$ .

So, for  $B \in \mathcal{M}_0$  and  $A \in \mathcal{F}$ , we know that  $A \cap B$ ,  $A \cap B^c$ , and  $A^c \cap B \in \mathcal{M}_0$ . Again,  $\mathcal{F} \subset \mathcal{M}_B$  and by minimality  $\mathcal{M}_0 = \mathcal{M}_B$ .

Finally, if  $A, B \in \mathcal{M}_0 = \mathcal{M}_A$ , then  $A \cap B$ ,  $A \cap B^c$ , and  $A^c \cap B \in \mathcal{M}_0$ . So  $\mathcal{M}_0$  is a field. ■

■ **Example 1.9** The above theorem requires that  $\mathcal{M}$  contain a field  $\mathcal{F}$ . It does not hold if this requirement is relaxed to, say,  $\mathcal{M}$  containing a  $\pi$ -system. As a counterexample, consider  $\mathcal{P} = \mathcal{M} = \{\emptyset\}$ , where  $\Omega \neq \emptyset$ . Clearly  $\mathcal{P}$  is a  $\pi$ -system, and  $\mathcal{M}$  contains a single element so it is a monotone class. However,  $\sigma(\mathcal{P}) = \{\emptyset, \Omega\} \not\subset \mathcal{M}$ . ■

## 2.

### 2.1 Lecture 9, September 14

**Definition 2.1.1 — Lim inf and lim sup.**

$$\liminf_n A_n = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k$$
$$\limsup_n A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k.$$

Suppose we have a probability space  $(\Omega, \mathcal{F}, P)$  such that  $A_n \in \mathcal{F}$ . Then  $\liminf_n A_n, \limsup_n A_n \in \mathcal{F}$ . So it makes sense to think about  $P(\liminf_n A_n)$  and  $P(\limsup_n A_n)$ . In fact, we will often use the following theorem about these values.

**Theorem 2.1.1** For any  $A_n$ ,

$$P\left(\liminf_n A_n\right) \leq \liminf_n P(A_n) \leq \limsup_n P(A_n) \leq P\left(\limsup_n A_n\right).$$

Note that  $A_n \subset \bigcup_{k \geq n} A_k$ . So  $P(A_n) \leq P\left(\bigcup_{k \geq n} A_k\right) \downarrow P(\limsup_n A_n)$ .

It is possible for the inequalities in Theorem 2.1.1 to all be strict, as in the following example.

■ **Example 2.1** Define a sequence of sets  $A_n$  such that  $A_{2n+1} = (0, 1/3]$  and  $A_{2n} = (1/3, 1]$ . In this

situation we have

$$\begin{aligned} P\left(\liminf_n A_n\right) &= 0, \\ \liminf_n P(A_n) &= \frac{1}{3}, \\ \limsup_n P(A_n) &= \frac{2}{3}, \text{ and} \\ P\left(\limsup_n A_n\right) &= 1. \end{aligned}$$

Clearly

$$P\left(\liminf_n A_n\right) < \liminf_n P(A_n) < \limsup_n P(A_n) < P\left(\limsup_n A_n\right).$$

■

**Definition 2.1.2 — Independent events.** Two events  $A$  and  $B$  are *independent* if  $P(AB) = P(A)P(B)$ .

Three events  $A$ ,  $B$ , and  $C$  are independent if all three are pairwise independent with each other, and  $P(ABC) = P(A)P(B)P(C)$ .

In general,  $A_1, \dots, A_n$  are independent if

$$P(A_{k_1} \cap \dots \cap A_{k_j}) = P(A_{k_1}) \cdots P(A_{k_j})$$

for  $2 \leq j \leq n$  and  $1 \leq k_1 < \dots < k_j \leq n$ .

■ **Example 2.2** Suppose  $\Omega = \{1, 2, 3, 4, 5, 6, 7, 8\}$ , and  $P(\{i\}) = \frac{1}{8}$ . Define the events

$$A = \{1, 2, 3, 4\}, \quad B = \{1, 2, 5, 6\}, \quad C = \{1, 2, 7, 8\}.$$

It is easy to check that any two of  $A$ ,  $B$ , and  $C$  are independent. However,  $P(ABC) = \frac{1}{4}$  but  $P(A)P(B)P(C) = \frac{1}{8}$ .

Now suppose we define

$$A = \{1, 2, 3, 4\}, \quad D = \{4, 5, 6, 7\}, \quad E = \{4, 5, 6, 7\}.$$

Here  $P(ADE) = P(A)P(D)P(E) = \frac{1}{8}$ , but no two of  $A$ ,  $D$ , and  $E$  are independent. In each of these scenarios, the three events are not independent. ■

Notice that technically, independence is not a property of the events themselves, but rather of the probability measure  $P$ .

If we have two events  $A, B$  that are independent, what can we say about  $A^c$  and  $B$ ?

$$\begin{aligned} P(A^c \cap B) &= P(B) - P(A \cap B) \\ &= P(B) - P(A)P(B) \\ &= P(B)[1 - P(A)] \\ &= P(B)P(A^c). \end{aligned}$$

$A^c$  and  $B$  are also independent.

**Definition 2.1.3 — Independent classes.** The classes  $\mathcal{F}_1, \dots, \mathcal{F}_n$  are independent if for every possible choice of  $A_i \in \mathcal{F}_i$ ,  $A_1, \dots, A_n$  are independent.

An infinite collection of classes  $\{\mathcal{F}_\theta : \theta \in \Theta\}$  is independent if each finite subcollection  $\mathcal{F}_{\theta_1}, \dots, \mathcal{F}_{\theta_n}$  of classes is independent.

**Theorem 2.1.2**

- If  $\mathcal{F}_1, \dots, \mathcal{F}_n$  are independent and each  $\mathcal{F}_i$  is a  $\pi$ -system, then  $\sigma(\mathcal{F}_1), \dots, \sigma(\mathcal{F}_n)$  are independent.
- If the classes  $\{\mathcal{F}_\theta : \theta \in \Theta\}$  are independent and each  $\mathcal{F}_\theta$  is a  $\pi$ -system, then  $\{\mathcal{F}_\theta : \theta \in \Theta\}$  are independent.

*Proof.* Fix  $A_2 \in \mathcal{F}_2, \dots, A_n \in \mathcal{F}_n$ . Let

$$\mathcal{G} = \{B \in \sigma(\mathcal{F}_1) : P(BB_2 \cdots B_n) = P(B)P(B_2) \cdots P(B_n), B_i = A_i \text{ or } \Omega\}.$$

Clearly  $\mathcal{F}_1 \in \mathcal{G}$ . Now check that  $\mathcal{G}$  is a  $\lambda$ -system. Recall that this means we need to check

- $\Omega \in \mathcal{G}$
- $C \in \mathcal{G}$  implies  $C^c \in \mathcal{G}$
- For disjoint  $C_i \in \mathcal{G}$ ,  $\bigcup_i C_i \in \mathcal{G}$ .

The first is trivial. For the second, fix some  $C \in \mathcal{G}$ . That means that  $P(C \cap B_2 \cap \cdots \cap B_n) = P(C)P(B_2) \cdots P(B_n)$ . We are trying to determine if  $P(C^c \cap B_2 \cap \cdots \cap B_n) = P(C^c)P(B_2) \cdots P(B_n)$ . Taking the probability of the union of the two left hand sides,

$$\begin{aligned} P((C \cap B_2 \cap \cdots \cap B_n) \cup (C^c \cap B_2 \cap \cdots \cap B_n)) &= P(\Omega \cap B_2 \cap \cdots \cap B_n) \\ &= P(B_2 \cap \cdots \cap B_n). \end{aligned}$$

Adding the two right hand sides gives

$$\begin{aligned} P(C)P(B_2) \cdots P(B_n) + P(C^c)P(B_2) \cdots P(B_n) &= P(\Omega)P(B_2) \cdots P(B_n) \\ &= P(B_2) \cdots P(B_n). \end{aligned}$$

Since the  $B$ s are independent it is indeed true that  $P(B_2 \cap \cdots \cap B_n) = P(B_2) \cdots P(B_n)$ . So  $C^c \in \mathcal{G}$ .

For the third property,

$$\begin{aligned} P\left(\bigcup_i C_i \cap B_2 \cap \cdots \cap B_n\right) &= \sum_{i=1}^{\infty} P(C_i \cap B_2 \cap \cdots \cap B_n) && (C_i \text{ disjoint}) \\ &= \left(\sum_{i=1}^{\infty} P(C_i)\right) P(B_2) \cdots P(B_n) \\ &= P\left(\bigcup_i C_i\right) P(B_2) \cdots P(B_n). && (C_i \text{ disjoint}) \end{aligned}$$

So  $\bigcup_i C_i \in \mathcal{G}$ , and thus  $\mathcal{G}$  is a  $\lambda$ -system, containing the  $\pi$ -system  $\mathcal{F}_1$ . Therefore  $\mathcal{G} \supset \sigma(\mathcal{F}_1)$ .

Now we can fix  $A_1 \in \sigma(\mathcal{F}_1), A_3 \in \mathcal{F}_3, \dots, A_n \in \mathcal{F}_n$ . Define

$$\mathcal{G}_2 = \{B \in \sigma(\mathcal{F}_2) : P(BB_2 \cdots B_n) = P(B)P(B_2) \cdots P(B_n); B_i = A_i \text{ or } \Omega\}.$$

We can proceed in the exact same way to show that  $\mathcal{G}_2 \supset \sigma(\mathcal{F}_2)$ , etc. This method suffices to prove that  $\sigma(\mathcal{F}_1), \sigma(\mathcal{F}_2), \dots, \sigma(\mathcal{F}_n)$  are independent. ■



**Corollary 2.1.3** Suppose that the array

$$\begin{array}{ccc} A_{11} & A_{12} & \cdots \\ A_{21} & A_{22} & \cdots \\ \vdots & \vdots & \ddots \end{array}$$

of events is independent. If  $\mathcal{F}_i$  is the  $\sigma$ -field generated by the  $i$ th row, then  $\mathcal{F}_1, \mathcal{F}_2, \dots$  are independent.

*Proof.* If  $\mathcal{A}_i$  is the class of all finite intersections of elements of the  $i$ th row, then each  $\mathcal{A}_i$  is a  $\pi$ -system, and by Theorem 2.1.2,  $\mathcal{F}_i = \sigma(\mathcal{A}_i)$ . ■

Why is this useful? Say we want to generate a sequence of normal random variables. It turns out that this can be done with just a coin. When we flip a coin  $n$  times we are generating the events  $Z_1, \dots, Z_n$  where  $P(Z_i = 1) = P(Z_i = 0) = \frac{1}{2}$ . Define

$$Z_0 = \sum_{i=1}^{\infty} \frac{Z_i}{2^i}.$$

Notice that  $Z_0 \in [0, 1]$  and  $P(Z_0 = x) = x$  for  $0 < x < 1$ . That is,  $Z_0$  is uniform on  $(0, 1)$ . We can then generate infinitely many of these by arranging them into an array like the above, and they will all be independent.

**Theorem 2.1.4 — Borel-Cantelli Lemmas.** Let  $(\Omega, \mathcal{F}, P)$  be a probability space, and let  $A_i \in \mathcal{F}$ .

1. If  $\sum_n P(A_n)$  converges, then  $P(\limsup_n A_n) = 0$ .
2. If  $\{A_n\}$  are independent and  $\sum_n P(A_n)$  diverges, then  $P(\limsup_n A_n) = 1$ .

*Proof.* Proof of 1:

$$P\left(\limsup_n A_n\right) \leq P\left(\bigcup_{k=m}^{\infty} A_k\right) \leq \sum_{k=m}^{\infty} P(A_k) \xrightarrow{m \rightarrow \infty} 0.$$

The above uses the countable subadditivity property of  $P$ .

Proof of 2: consider the complement.  $(\limsup_n A_n)^c = (\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k)^c = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k^c$ .

$$\begin{aligned} P\left(\bigcap_{k=m}^{\infty} A_k^c\right) &\leq P\left(\bigcap_{k=m}^r A_k^c\right) && \text{(nonnegativity)} \\ &= \prod_{k=m}^r (1 - P(A_k)) && \text{(independence)} \\ &\leq e^{-\sum_{k=m}^r P(A_k)}. && \text{(because } 1 - x \leq e^{-x} \text{ } \forall x) \end{aligned}$$

As  $\sum_n P(A_n)$  is a divergent series,  $e^{-\sum_{k=m}^r P(A_k)} \rightarrow 0$  as  $r \rightarrow \infty$ . So  $P(\bigcup_{m=1}^{\infty} \bigcap_{k=m}^{\infty} A_k^c) = 0$ , and thus the complement  $P(\limsup_n A_n) = 1$ . ■

The Borel-Cantelli Lemmas tell us that if we have infinitely many independent events,  $P(\limsup_n A_n) = 0$  or  $1$  always, regardless of the choice of probability measure  $P$ .

Is independence necessary for the lemmas to hold? The answer is yes, as the following example illustrates.



■ **Example 2.3** Consider the unit interval Lebesgue measure. Define  $A = (0, 1/3]$  and  $B = (2/3, 1]$ . Let  $A_{2n} = A, A_{2n+1} = B$ . So  $A_k$  is an infinite sequence of sets  $A, B, A, B, \dots$ .

Now each  $P(A_n) = 1/3$ , so  $\sum P(A_n) = \infty$ . However,  $\limsup_n A_n = A \cup B$ , and so  $P(\limsup_n A_n) = 2/3$ , not 0 or 1 as the Borel-Cantelli Lemmas guarantee for independent events. ■

## 2.2 Lecture 10, September 16

There is a version of Borel-Cantelli that does work for dependent events, however.

**Theorem 2.2.1 — Borel-Cantelli Lemma, dependent events.** Let

$$\theta_n = \frac{\sum_{j=1}^m \sum_{k=1}^n P(A_j A_k)}{[\sum_{k=1}^n P(A_k)]^2}.$$

If  $\liminf_n \theta_n \leq 1$  and  $\sum_n P(A_n) = \infty$ , then  $P(\limsup_n A_n) = 1$ .

*Proof.* Let  $N_n = \sum_{i=1}^n I_{A_i}$  and  $m_n = \sum_{k \leq n} P(A_k)$ . If  $m_n > x$ , then

$$\begin{aligned} P(N_n \leq x) &= P(|N_n - m_n| \geq m_n - x) \\ &\leq \text{Var}(N_n) / (m_n - x)^2 && \text{(Chebyshev)} \\ &= (\theta_n - 1) m_n^2 / (m_n - x)^2. && (N_n \text{ is binomial}) \end{aligned}$$

So  $\liminf_n P(N_n \leq x) = 0$ . Since  $P(\sup_k N_k < x) \leq P(N_n \leq x)$ ,

$$P(\sup_k N_k < \infty) = P\left(\bigcup_{m=1}^{\infty} (\sup_k N_k < m)\right) = 0.$$

Hence  $P(\limsup_n A_n) = P(\sup_n N_n = \infty) = 1$ . ■

**Corollary 2.2.2** If  $\{A_n\}$  are pairwise independent, and  $\sum_n P(A_n) = \infty$ , then  $P(\limsup_n A_n) = 1$ .

We will now introduce the concept of a *tail  $\sigma$ -field*.

**Definition 2.2.1 — Tail  $\sigma$ -field.**  $\mathcal{T} = \bigcap_{n=1}^{\infty} \sigma(A_n, A_{n+1}, \dots)$  is called the *tail  $\sigma$ -field* associated with  $A_1, A_2, \dots$ .

To understand this conceptually, consider the sequence

$$\sigma(A_1, A_2, \dots) \supset \sigma(A_2, A_3, \dots) \supset \sigma(A_3, A_4, \dots).$$

These are decreasing  $\sigma$ -fields, and  $\mathcal{T}$  is the intersection of all of these. It is usually very small. For example, if  $A_1 = A_2 = \dots = A$ , then  $\mathcal{T} = \sigma(\{A\}) = (\emptyset, \Omega, A, A^c)$ .

**Theorem 2.2.3 — Zero-one law.** If  $A_1, A_2, \dots$  are independent, then  $P(A)$  is either 0 or 1 for each  $A \in \mathcal{T}$ .

*Proof.* Clearly  $\sigma(A_1, \dots, A_n)$  and  $\sigma(A_{n+1}, \dots)$  are independent for all  $n$ .  $A \in \mathcal{T}$  implies that  $A$  and  $A_1, \dots, A_n$  are independent, so  $A$  and  $A_1, A_2, \dots$  are independent. Thus  $A$  and  $\mathcal{T}$  are independent, and hence  $A$  and  $A^c$  are independent. So

$$0 = P(A \cap A^c) = P(A)P(A^c) \implies P(A) = 0 \text{ or } 1.$$

■

Suppose  $A = \limsup_n A_n$ . Is  $A \in \mathcal{T}$ ?

$$\begin{aligned} A &= \bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} A_n = \{\omega : \omega \in A_k \text{ for infinitely many } k = 1, 2, \dots\} \\ &= \{\omega : \omega \in A_k \text{ for infinitely many } k = 2, 3, \dots\} \\ &= \{\omega : \omega \in A_k \text{ for infinitely many } k = r, r+1, \dots \text{ for any } r\} \\ &\in \sigma(A_r, A_{r+1}, \dots) \end{aligned}$$

and so  $A \in \mathcal{T}$ . Also,

$$A = \bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} A_n = \bigcap_{m=2}^{\infty} \bigcup_{n=m}^{\infty} A_n = \bigcap_{m=r}^{\infty} \bigcup_{n=m}^{\infty} A_n \in \sigma(A_r, A_{r+1}, \dots).$$

So  $\liminf_n$  and  $\limsup_n$  are both in the tail  $\sigma$ -field.

Note that Borel-Cantelli can actually help tell you *when*  $P(A)$  is 0 or 1. The zero-one law just tells you that it must be one or the other.

**Theorem 2.2.4 — Bernstein's Theorem.** If  $f$  is continuous, then  $B_n(x) \rightarrow f(x)$  uniformly on  $[0, 1]$ , where

$$B_n(x) = \sum_{k=0}^n f(k/n) \binom{n}{k} x^k (1-x)^{n-k}$$

are Bernstein polynomials.

*Proof.* Let  $M = \sup_x |f(x)|$ , and let  $\delta(\varepsilon) = \sup\{|f(x) - f(y)| : |x - y| \leq \varepsilon\}$ . This value is known as the *modulus of continuity* of  $f$ . It will be shown that

$$\sup_x |f(x) - B_n(x)| \leq \delta(\varepsilon) + \frac{2M}{n\varepsilon^2}.$$

Because  $f$  is uniformly continuous,  $\lim_{\varepsilon \rightarrow 0} \delta(\varepsilon) = 0$ , and so this inequality (provided we pick a sufficiently small  $\varepsilon$ , say  $\varepsilon = n^{-1/3}$ ) will give the result in the theorem.

Fix  $n \geq 1$  and  $x \in [0, 1]$  for the moment. Let  $X_1, \dots, X_n$  be independent random variables (on some probability space) such that  $P(X_i = 1) = x$  and  $P(X_i = 0) = 1 - x$ . Define  $S = X_1 + \dots + X_n$ . Since  $S$  is just the sum of  $n$  Bernoulli random variables,  $P(S = k) = \binom{n}{k} x^k (1-x)^{n-k}$ , and by the formula for calculating expected values of functions of random variables,  $E[f(S/n)] = B_n(x)$ . Additionally,  $E[S/n] = x$ , so by the law of large numbers, there should be a high probability that  $S/n$  is near  $x$ , and hence (because  $f$  is continuous) that  $f(S/n) = f(x)$ . Therefore,  $E[f(S/n)]$  should be near  $f(x)$ . This is the probabilistic idea behind the definition of the Bernstein polynomial  $B_n(x)$ .

For some  $\varepsilon > 0$ , consider the set  $\{|S/n - x| < \varepsilon\}$ . This is the set of values of  $x$  that are “near” (within  $\varepsilon$  of)  $S/n$ . On this set,  $|f(S/n) - f(x)|$  is bounded by  $\delta(\varepsilon)$ . On  $\{|S/n - x| \geq \varepsilon\}$ , the complementary set,  $|f(S/n) - f(x)|$  is bounded by  $2M$ . Now

$$\begin{aligned} |B_n(x) - f(x)| &= |E[f(S/n) - f(x)]| \\ &\leq E[|f(S/n) - f(x)|] \\ &\leq \delta(\varepsilon)P[|S/n - x| < \varepsilon] + 2MP[|S/n - x| \geq \varepsilon] \\ &\leq \delta(\varepsilon) + 2M\text{Var}[S]/n^2\varepsilon^2. \end{aligned} \quad (\text{Chebyshev's Inequality})$$

Since  $\text{Var}[S] = nx(1-x) \leq n$ , we indeed find that

$$\sup_x |f(x) - B_n(x)| \leq \delta(\varepsilon) + \frac{2M}{n\varepsilon^2}.$$

■

### 2.3 Lecture 11, September 18

We can't begin to define constructs like probability density functions without first developing the concept of an *infinite measure*.

**Definition 2.3.1 — General measure.** A set function  $\mu$  on a field  $\mathcal{F}$  on  $\Omega$  is a *measure* if it satisfies

- (I)  $0 \leq \mu(A) \leq \infty$  for  $A \in \mathcal{F}$
- (II)  $\mu(\emptyset) = 0$
- (III) If  $\{A_n\}$  is a sequence of disjoint sets in  $\mathcal{F}$  and if  $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$ , then

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n).$$

Furthermore,  $\mu$  is a *finite measure* if  $\mu(\Omega) < \infty$ . On the other hand,  $\mu$  is an *infinite measure* if  $\mu(\Omega) = \infty$ .

**Definition 2.3.2 —  $\sigma$ -finite measure.**  $\mu$  is a  *$\sigma$ -finite measure* if  $\Omega = A_1 \cup A_2 \cup \dots$  for some finite or countable sequence of sets in  $\mathcal{F}$  satisfying  $\mu(A_k) < \infty$ .

A finite measure is by definition  $\sigma$ -finite, but a  $\sigma$ -finite measure may be finite or infinite. If  $\mathcal{A}$  is a subclass of  $\mathcal{F}$ , then  $\mu$  is  *$\sigma$ -finite on  $\mathcal{A}$*  if  $\Omega = \bigcup_k A_k$  for some finite or infinite sequence of sets in  $\mathcal{A}$  satisfying  $\mu(A_k) < \infty$ . These sets need not necessarily be disjoint. It is important to understand that  $\sigma$ -finiteness is a joint property of the space  $\Omega$ , the measure  $\mu$ , and the class  $\mathcal{A}$ .

Results about probability measures can be extended to finite measures simply by redefining

$$P(A) = \frac{\mu(A)}{\mu(\Omega)}.$$

We can extend to  $\sigma$ -finite measures as well, since they are actually made up of finite intervals:  $\Omega = \bigcup_k A_k$ ,  $A_k \in \mathcal{F}$ , and  $\mu(A_i) < \infty$ . For example, the real line can be written

$$\mathbb{R} = \bigcup_{n=1}^{\infty} (-n, n) = \bigcup_{m=-\infty}^{\infty} (m, m+1].$$

■ **Example 2.4 — Counting measure.** Suppose  $\Omega = \{1, 2, 3, \dots\}$  and  $\mathcal{F}$  consists of all subsets of  $\Omega$ . Define  $\mu(A)$  to be the number of points in  $A$  if  $A$  is a finite set, and  $\infty$  otherwise. This  $\mu$  is called the *counting measure*, for obvious reasons.

We can write  $\Omega = \bigcup_{n=1}^{\infty} \{n\}$ , and  $\mu(\{n\}) = 1 < \infty$  for all  $n$ . So  $\mu$  is  $\sigma$ -finite.

Now suppose instead that  $\Omega = \mathbb{R}$ . In this case,  $\mu(\text{any interval}) = \infty$ . To write  $\mathbb{R} = \Omega = \bigcup_i A_i$ , at least one of the  $A_i$  must be infinite. So  $\mu$  is not  $\sigma$ -finite. ■

■ **Example 2.5** The restriction of a  $\sigma$ -finite measure to a smaller  $\sigma$ -field may not necessarily be  $\sigma$ -finite itself. The counting measure is not  $\sigma$ -finite on  $\{\emptyset, \Omega\}$  if  $\Omega$  is a countably infinite set. ■

**Theorem 2.3.1** Let  $\mu$  be a measure on a field  $\mathcal{F}$  and  $A_n, A \in \mathcal{F}$ .

- (I) *Continuity from below:* if  $A_n \uparrow A$ , then  $\mu(A_n) \uparrow \mu(A)$ .
- (II) *Continuity from above:* if  $A_n \downarrow A$ , and  $\mu(A_k) < \infty$  for some  $k$ , then  $\mu(A_n) \downarrow \mu(A)$ .
- (III) *Countable subadditivity:*

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} \mu(A_n).$$

(IV) If  $\mu$  is  $\sigma$ -finite on  $\mathcal{F}$ , then  $\mathcal{F}$  cannot contain an uncountable, disjoint collection of sets of positive  $\mu$ -measure.

Notice the extra condition in (II). A simple example demonstrating why this is necessary is as follows. Suppose  $\Omega = \{1, 2, \dots\}$  and  $\mathcal{F}$  consists of all subsets of  $\Omega$ . Define  $\mu$  to be the counting measure as before. Now let  $A_n = \{n, n+1, \dots\}$ . Then  $A_n \downarrow \emptyset$ , and  $\mu(\emptyset) = 0$ , yet  $\mu(A_n) = \infty$  for all  $n$ . So clearly  $\mu(A_n)$  does not converge to  $\mu(\emptyset)$ .

## 2.4 Lecture 12, September 23

**Theorem 2.4.1** Suppose that  $\mu_1$  and  $\mu_2$  are measures on  $\sigma(\mathcal{P})$ , where  $\mathcal{P}$  is a  $\pi$ -system, and suppose they are  $\sigma$ -finite on  $\mathcal{P}$ . If  $\mu_1$  and  $\mu_2$  agree on  $\mathcal{P}$ , then they agree on  $\sigma(\mathcal{P})$ .

*Proof.* Suppose  $A \in \mathcal{P}$ , and  $0 < \mu_1(A) = \mu_2(A) < \infty$ . Define the class

$$\mathcal{G} = \{B \in \sigma(\mathcal{P}) : \mu_1(A \cap B) = \mu_2(A \cap B)\}.$$

Now  $\mathcal{P} \subset \mathcal{G}$ . And  $B \in \mathcal{P}$  implies  $A \cap B \in \mathcal{P}$ , since  $A \in \mathcal{P}$  and  $\mathcal{P}$  is a  $\pi$ -system. Also,

$$\begin{aligned} B \in \mathcal{G} &\implies \mu_1(A \cap B^c) = \mu_1(A) - \mu_1(A \cap B) \\ &= \mu_2(A) - \mu_2(A \cap B) = \mu_2(A \cap B^c). \end{aligned}$$

So  $\mathcal{G}$  is closed under complementation.

Now take  $B_1, B_2, \dots$  disjoint.

$$\begin{aligned} \mu_1\left(A \cap \bigcup_{i=1}^{\infty} B_i\right) &= \sum_i \mu_1(A \cap B_i) \\ &= \sum_i \mu_2(A \cap B_i) \\ &= \mu_2\left(A \cap \bigcup_{i=1}^{\infty} B_i\right), \end{aligned}$$

so  $\mathcal{G}$  is closed under countable union. Thus  $\mathcal{G}$  is a  $\lambda$ -system, and it contains a  $\pi$ -system  $\mathcal{P}$ . Thus  $\mathcal{G} \supset \sigma(\mathcal{P})$ . Clearly  $\mathcal{G} \subset \sigma(\mathcal{P})$ . Therefore  $\mathcal{G} = \sigma(\mathcal{P})$ . And so  $\mu_1$  and  $\mu_2$  agree on  $\sigma(\mathcal{P})$ . ■

**Theorem 2.4.2** Suppose  $\mu_1$  and  $\mu_2$  are finite measures on  $\sigma(\mathcal{P})$ , where  $\mathcal{P}$  is a  $\pi$ -system and  $\Omega$  is a finite or countable union of sets in  $\mathcal{P}$ . If  $\mu_1$  and  $\mu_2$  agree on  $\mathcal{P}$ , then they agree on  $\sigma(\mathcal{P})$ .

*Proof.* Suppose  $\mu_1$  and  $\mu_2$  are finite measures on  $\sigma(\mathcal{P})$ . We want to find sets  $A_1, A_2, \dots \in \mathcal{P}$  such that  $\bigcup_{i=1}^{\infty} A_i = \Omega$  where  $\mu_1(A_i) < \infty$ . It is important to note that we can't simply disjointify the  $A_i$  here because there is no guarantee that  $A_1 - A_2$ , say, will be in  $\mathcal{P}$ . Instead, let  $C \in \sigma(\mathcal{P})$ . Then,

using the inclusion-exclusion principle,

$$\begin{aligned}
 \mu_1 \left( C \cap \bigcup_{i=1}^k A_i \right) &= \mu_1 \left( \bigcup_{i=1}^k (C \cap A_i) \right) \\
 &= \sum_{i=1}^k \mu_1(C \cap A_i) - \sum_{i < j} \mu_1(C \cap A_i \cap A_j) + \cdots \\
 &\quad + (-1)^{k+1} \mu_1(C \cap A_1 \cap A_2 \cap \cdots \cap A_k) \\
 &= \sum_{i=1}^k \mu_2(C \cap A_i) - \sum_{i < j} \mu_2(C \cap A_i \cap A_j) + \cdots \\
 &\quad + (-1)^{k+1} \mu_2(C \cap A_1 \cap A_2 \cap \cdots \cap A_k) \\
 &= \mu_2 \left( C \cap \bigcup_{i=1}^k A_i \right).
 \end{aligned}$$

So  $\mu_1(C) = \mu_2(C)$ . ■

■ **Example 2.6 — Counterexample to Theorem 2.4.2.** Suppose  $\mathcal{P} = \{\emptyset\}$ . Then  $\mathcal{P}$  is a  $\pi$ -system and  $\sigma(\mathcal{P}) = \{\Omega, \emptyset\}$ . Any two finite measures agree on  $\mathcal{P}$ , but they do not necessarily agree on  $\sigma(\mathcal{P})$ . Theorem 2.4.2 does not apply in this case because  $\Omega$  is not a countable union of sets in  $\mathcal{P}$ . ■

■ **Example 2.7 — Another counterexample.** Consider the real number line  $\mathbb{R}$  and the Borel  $\sigma$ -field  $\mathcal{B}$ . Define  $\mathcal{P}_0 = \{(-\infty, x] : x \in \mathbb{R}\}$ . This is a  $\pi$ -system: it should be clear that the intersection of two such sets will also be in  $\mathcal{P}_0$ .

If we take two sets  $(-\infty, a]$  and  $(-\infty, b] \in \mathcal{P}_0$ , notice that  $(-\infty, a]^c \cap (\infty, b] = (a, b]$ . Define  $\mathcal{P} = \{(a, b] : a \leq b < \infty\}$ . Then  $\sigma(\mathcal{P}_0) = \sigma(\mathcal{P}) = \mathcal{B}$ .

Now take  $\mu_1(a, b] = b - a$  and  $\mu_2(a, b] = 2(b - a)$ . Then  $\mu_1 = \lambda$ , our standard concept of Lebesgue measure, and  $\mu_2 = 2\lambda$ . These are distinct, yet  $\mu_1 = \mu_2$  on  $\mathcal{P}_0$ . Theorem 2.4.2 does not hold in this situation because  $\mu_1$  and  $\mu_2$  are infinite measures. ■

**Definition 2.4.1 — Outer measure.** An *outer measure* is a set function  $\mu^*$  on all subsets of  $\Omega$  satisfying:

- (i)  $\mu^*(\emptyset) = 0$
- (ii)  $0 \leq \mu^*(A) \leq \infty$  for all  $A \subset \Omega$
- (iii)  $\mu^*(A) \leq \mu^*(B)$  whenever  $A \subset B$
- (iv)  $\mu^*(\bigcup_i A_i) \leq \sum_i \mu^*(A_i)$  (Countable subadditivity)

**Theorem 2.4.3** Recall the class definition from earlier.

$$\mathcal{M}(\mu^*) = \{A \subset \Omega : \mu^*(AE) + \mu^*(A^c E) \leq \mu^*(E) \text{ for all } E \subset \Omega\}.$$

If  $\mu^*$  is an outer measure, then  $\mathcal{M}(\mu^*)$  is a  $\sigma$ -field, and  $\mu^*$  restricted to  $\mathcal{M}(\mu^*)$  is a measure.

**Theorem 2.4.4 — Extension theorem, general measures.** A measure  $\mu$  on a field  $\mathcal{F}_0$  has an extension to the generating  $\sigma$ -field.

*Proof.* The proof follows in the same manner as the one for Theorem 1.5.1. ■



# 3.

## 3.1 Lecture 12, September 23 (continued)

**Definition 3.1.1 — Measurable function.** The function  $T : (\Omega, \mathcal{F}) \rightarrow (\Omega', \mathcal{F}')$  is *measurable*  $\mathcal{F}/\mathcal{F}'$  if  $T^{-1}(A') \in \mathcal{F}$  for all  $A' \in \mathcal{F}'$ . Sometimes this is simply written “ $T$  is measurable  $\mathcal{F}$ .”

We can use this concept to define a measure on another space. If  $\mu$  is a measure on  $(\Omega, \mathcal{F})$ , then  $\mu(T^{-1}(A'))$  for  $A' \in \mathcal{F}'$  is a measure on  $(\Omega', \mathcal{F}')$ . The following property of measurable functions is useful:

$$T^{-1}\left(\bigcup_{i=1}^{\infty} A'_i\right) = \bigcup_{i=1}^{\infty} T^{-1}(A'_i).$$

To see this, take any point  $x \in T^{-1}(\bigcup_{i=1}^{\infty} A'_i)$ . Then  $T(x) \in \bigcup_{i=1}^{\infty} A'_i$ , which implies  $T(x) \in A'_j$  for some  $j$ , which implies  $x \in T^{-1}(A'_j)$ , which finally implies  $x \in \bigcup_{i=1}^{\infty} T^{-1}(A'_i)$ .

This is most useful for translating a measure to a measure on a real line, working on it there, and then translating it back.

**Theorem 3.1.1** The following two statements are true.

- (a)  $T^{-1}(A') \in \mathcal{F}$  for all  $A' \in \mathcal{A}'$  implies that  $T$  is  $\mathcal{F}/\sigma(\mathcal{A}')$  measurable.
- (b) If  $T : (\Omega, \mathcal{F}) \rightarrow (\Omega', \mathcal{F}')$  and  $T' : (\Omega', \mathcal{F}') \rightarrow (\Omega'', \mathcal{F}'')$  are measurable, then  $T'T : (\Omega, \mathcal{F}) \rightarrow (\Omega'', \mathcal{F}'')$  is measurable.

*Proof.* For (a): Consider the class

$$\mathcal{G} = \{A' \in \mathcal{F}' : T^{-1}(A') \in \mathcal{F}\}.$$

Then because  $T^{-1}(A') \in \mathcal{F}$  for all  $A' \in \mathcal{A}'$ ,  $\mathcal{G} \supset \mathcal{A}'$ . It is straightforward to show, as we have done many times before, that  $\mathcal{G}$  is a  $\sigma$ -field on  $\Omega'$ . Therefore  $T^{-1}(A') \in \mathcal{F}$  for all  $A' \in \sigma(\mathcal{A}')$ , which means that  $T$  is  $\mathcal{F}/\sigma(\mathcal{A}')$  measurable.

For (b):  $(T'T)^{-1}(A'') = T^{-1}(T'^{-1}(A''))$ .  $A'' \in \mathcal{F}''$  implies  $T'^{-1}(A'') \in \mathcal{F}'$ . So  $T^{-1}(T'^{-1}(A'')) \in \mathcal{F}$ . ■

### 3.2 Lecture 13, September 25

Recall that the  $\sigma$ -field  $\mathcal{R}^k$  generated by all  $k$ -dimensional rectangles in  $\mathbb{R}^k$  is called the *Borel  $\sigma$ -field*.

$$\begin{aligned}\mathcal{R}^k &= \{y_1, \dots, y_k : a_i \leq y_i \leq b_i, i = 1, \dots, k\} \\ &= [a_1, b_1] \times [a_2, b_2] \times \dots \times [a_k, b_k].\end{aligned}$$

**Definition 3.2.1 — Random vector.** A random vector is a measurable function  $f : (\Omega, \mathcal{F}) \rightarrow (\mathbb{R}^k, \mathcal{R}^k)$ . If  $k = 1$ , then  $f$  is called a *random variable*.

A measurable function  $f : (\mathbb{R}^i, \mathcal{R}^i) \rightarrow (\mathbb{R}^k, \mathcal{R}^k)$  is called a *Borel function*.

Note that  $f(\omega) = (f_1(\omega), f_2(\omega), \dots, f_k(\omega))$  is a random vector if and only if each  $f_i : \Omega \rightarrow \mathbb{R}$  is a random variable.

Look at the class  $\mathcal{A} = \{(-\infty, x] : x \in \mathbb{R}\}$ . Then  $\sigma(\mathcal{A}) = \mathcal{R}$ , the Borel  $\sigma$ -field. This is because  $\mathcal{R} = \sigma(\mathcal{I})$ , where  $\mathcal{I} = \{(a, b] : a \leq b\}$ , and  $(-\infty, x] = \bigcup_{-n < x} (-n, x]$ . Again, observe that  $(-\infty, b] \cap (-\infty, a]^c = (a, b]$ . So the two characterizations are identical. This is why CDFs only use sets of the type in the class  $\mathcal{A}$ : they give all the information you need.

#### Theorem 3.2.1

- (a) Continuous functions on  $\mathbb{R}^j$  into  $\mathbb{R}^k$  are measurable.
- (b) If  $f_i : \Omega \rightarrow \mathbb{R}$  are random variables, and  $g : \mathbb{R}^k \rightarrow \mathbb{R}$  is a Borel function, then  $g(f_1, \dots, f_k)$  is measurable  $\mathcal{F}$ .

*Proof.* Part (a) is obvious. If  $f : \mathbb{R}^j \rightarrow \mathbb{R}^k$  is continuous, then write  $f$  as

$$f(x_1, \dots, x_j) = (f_1(x_1, \dots, x_j), \dots, f_k(x_1, \dots, x_j)).$$

Then each  $f_i : \mathbb{R}^j \rightarrow \mathbb{R}$  is continuous.

For part (b), if  $f : (\Omega, \mathcal{F}) \xrightarrow{\text{meas}} (\mathbb{R}^k, \mathcal{R}^k)$ , then  $(g \circ f)(\omega) = g(f(\omega))$  is measurable if  $g$  is. ■

From this theorem it immediately follows that  $\sum_{i=1}^k f_i$  and  $\prod_{i=1}^k f_i$  are measurable if the  $f_i$  are.

We now shift our focus to the limits of sequences of measurable functions. The “extended real line,” denoted  $\overline{\mathbb{R}}$ , is simply  $\mathbb{R}$  with the symbols  $-\infty$  and  $\infty$  appended. Consider a function  $f : (\Omega, \mathcal{F}) \rightarrow (\overline{\mathbb{R}}, \overline{\mathcal{R}})$ . Here  $\overline{\mathbb{R}} = (-\infty, \infty]$  and  $\overline{\mathcal{R}} = \sigma(\mathcal{R}, \{-\infty\}, \{\infty\}) = \sigma(\{(a, b] : a \leq b\}, \{-\infty\}, \{\infty\})$ . If  $f$  is measurable, and the sets  $\{\omega : f(\omega) = \infty\}$  and  $\{\omega : f(\omega) = -\infty\}$  are both in  $\mathcal{F}$ , then we say that  $f$  is *extended measurable*.

**Theorem 3.2.2** Suppose  $f_1, f_2, \dots$  are real functions measurable  $\mathcal{F}$  (extended measurable or not).

- (a) The functions  $\sup_n f_n$ ,  $\inf_n f_n$ ,  $\limsup_n f_n$ , and  $\liminf_n f_n$  are all measurable  $\mathcal{F}$ .
- (b) If  $\lim_n f_n$  exists everywhere, then it is measurable  $\mathcal{F}$ .
- (c)  $L = \{\omega : \lim_n f_n(\omega) \text{ exists}\}$  is measurable.
- (d) If  $f$  is measurable  $\mathcal{F}$ , then  $G = \{\omega : f_n(\omega) \rightarrow f(\omega)\} \in \mathcal{F}$ .

*Proof.* Part (a): Notice that  $\{\sup_n f_n \leq x\} = \bigcap_n \{f_n \leq x\}$  lies in  $\mathcal{F}$ , even if  $x = -\infty$  or  $\infty$ . Thus  $\sup_n f_n$  is measurable. We can conclude that  $\inf_n f_n$  is measurable by the same argument. Hence  $\limsup_n f_n = \inf_n \sup_{k \geq n} f_k$  is measurable, and  $\liminf_n f_n = \sup_n \inf_{k \geq n} f_k$  is measurable.

Part (b): If  $\lim_n f_n$  exists, it is equal to both  $\liminf_n f_n$  and  $\limsup_n f_n$ , so of course it is measurable.



Part (c): Define  $g = \liminf_n f_n$  and  $h = \limsup_n f_n$ . The set  $L$  is simply  $L = \{\omega : g(\omega) = h(\omega)\}$ . At a glance this seems to be measurable, but we must be careful with infinity values.

$$\begin{aligned} L &= \{\omega : h(\omega) = \infty, g(\omega) = \infty\} \cup \{\omega : h(\omega) = -\infty, g(\omega) = -\infty\} \\ &\quad \cup \{\omega : -\infty < h(\omega) = g(\omega) < \infty\} \\ &= \left[ \{\omega : h(\omega) = \infty\} \cap \{\omega : g(\omega) = \infty\} \right] \cup \left[ \{\omega : h(\omega) = -\infty\} \cap \{\omega : g(\omega) = -\infty\} \right] \\ &\quad \cup \{\omega : h(\omega) - g(\omega) = 0 \text{ and } -\infty < g(\omega), h(\omega) < \infty\} \end{aligned}$$

The first two terms in this union are in  $\mathcal{F}$ . It is enough to show that the last one is in  $\mathcal{F}$ . This part can be written as

$$\{\omega : h(\omega) - g(\omega) = 0\} \cap \{\omega : |h(\omega)| < \infty, |g(\omega)| < \infty\}.$$

Both of these are indeed in  $\mathcal{F}$ , so finally we can conclude that  $L$  is measurable.

Part (d): The set  $G$  is  $G = \{\omega : \liminf_n f_n(\omega) = \limsup_n f_n(\omega) = f(\omega)\}$ . By a similar argument as above,  $G$  is measurable. ■

### 3.3 Lecture 14, September 28

The next theorem, Egoroff's Theorem, essentially says that if a sequence of real-valued measurable functions converges to another measurable function, we have uniform convergence once we throw out a very small set (of measure less than  $\varepsilon$ ).

**Theorem 3.3.1 — Egoroff's Theorem.** Suppose  $f_n$  and  $f$  are real-valued  $\mathcal{F}$ -measurable functions such that  $f_n(\omega) \rightarrow f(\omega)$  for  $\omega \in A$ , where  $\mu(A) < \infty$ . Then for each  $\varepsilon > 0$ , there exists a measurable subset  $B \subset A$  such that  $\mu(B) < \varepsilon$  and  $f_n(\omega) \rightarrow f(\omega)$  uniformly on  $A - B$ .

*Proof.* Let

$$B_n^{(k)} = \bigcup_{i \geq n} \left\{ \omega \in A : |f(\omega) - f_i(\omega)| > \frac{1}{k} \right\}.$$

Since  $\limsup_n B_n^{(k)} = \emptyset$ ,  $B_n^{(k)} \downarrow \emptyset$  as  $n \rightarrow \infty$ . Choose an  $n_k$  such that  $\mu(B_{n_k}^{(k)}) < \frac{\varepsilon}{2^k}$ , and define  $B = \bigcup_{k=1}^{\infty} B_{n_k}^{(k)}$ . Then

$$A - B = A \bigcap_{k=1}^{\infty} B_{n_k}^{(k)c} = \bigcap_{k=1}^{\infty} \bigcap_{i \geq n_k} \left\{ \omega \in A : |f(\omega) - f_i(\omega)| \leq \frac{1}{k} \right\}.$$

So  $f_n(\omega) \rightarrow f(\omega)$  uniformly on this set. ■

It is important to note that Egoroff's Theorem does not hold without the condition  $\mu(A) < \infty$ , as we will illustrate in the following example.

■ **Example 3.1** Suppose  $\Omega = \{1, 2, \dots\}$ , and let  $\mathcal{F}$  be the class of all subsets of  $\Omega$ . Let  $\mu$  be the counting measure. In this situation, define the functions

$$f_n(\omega) = n \mathbb{I}_{\{n, n+1, \dots\}} = \begin{cases} n & \text{if } \omega \geq n \\ 0 & \text{otherwise.} \end{cases}$$

Here  $f_n(\omega) \rightarrow 0$  as  $n \rightarrow \infty$  for all  $\omega$ , but we don't have uniform convergence, violating the theorem. The reason it fails is that  $\Omega$  has infinite measure. On a finite set, however,  $f_n(\omega)$  does converge to 0 uniformly. ■

We will now discuss the concept of measures on  $\mathbb{R}^k$ . *Lebesgue measure*  $\lambda_k$  on the class  $\mathcal{R}^k$  of  $k$ -dimensional Borel sets is the extension of the regular Lebesgue measure  $\lambda$ .

$$\lambda_k\{\mathbf{x} : a_i < x_i < b_i, i = 1, \dots, k\} = \prod_{i=1}^k (b_i - a_i).$$

Lebesgue measure does not change if the set is shifted by a constant. If  $A \in \mathcal{R}^k$ , then  $A + \mathbf{x} = \{a + \mathbf{x} : a \in A\} \in \mathcal{R}^k$  and  $\lambda_k(A) = \lambda_k(A + \mathbf{x})$ . To prove this, define

$$\mathcal{G} = \{A \in \mathcal{R}^k : A + \mathbf{x} \in \mathcal{R}^k \text{ and } \lambda_k(A) = \lambda_k(A + \mathbf{x})\}.$$

This class  $\mathcal{G}$  contains  $\mathcal{I}_k$ , the class of all rectangles  $[a_1, b_1] \times \dots \times [a_k, b_k]$ . It is straightforward to show that  $\mathcal{G}$  is a  $\lambda$ -system and that  $\mathcal{I}_k$  is a  $\pi$ -system. Thus  $\mathcal{R}^k \supset \mathcal{G} \supset \sigma(\mathcal{I}_k) = \mathcal{R}^k$ , and so  $\mathcal{G} = \mathcal{R}^k$ .

If  $T$  is a non-singular linear transformation on  $\mathbb{R}^k \rightarrow \mathbb{R}^k$ , then  $A \in \mathcal{R}^k$  implies  $T(A) \in \mathcal{R}^k$ , and  $\lambda_k(T(A)) = |\det T| \cdot \lambda_k(A)$ . To prove this, note that by elementary row and column operations,  $T$  can be represented as the product of linear transformations of three special forms:

- (a)  $T(x_1, \dots, x_k) = (x_{\pi 1}, \dots, x_{\pi k})$ ;
- (b)  $T(x_1, \dots, x_k) = (\alpha x_1, \dots, x_k)$ ;
- (c)  $T(x_1, \dots, x_k) = (x_1 + x_2, x_2, \dots, x_k)$ .

Therefore, if we can prove the property holds for each of the above, we have proved that it holds for any  $T$ . Define

$$\mathcal{G} = \{A \in \mathcal{R}^k : T(A) \in \mathcal{R}^k, \lambda_k(T(A)) = |\det T| \lambda_k(A)\}.$$

Also define  $\mathcal{I}_k^0 = \{(a_1, b_1] \times \dots \times (a_k, b_k] : a_i, b_j \text{ are all rationals}\}$ . This is a  $\pi$ -system. By the previous theorem,  $\sigma(\mathcal{I}_k^0) = \mathcal{R}^k$ . Continue in the same way as above, by showing that  $\mathcal{G}$  is a  $\lambda$ -system and concluding that  $\mathcal{G} = \mathcal{R}^k$ .

### 3.4 Lecture 15, September 30

Suppose  $\mu$  is a measure on  $\mathcal{R}$  such that  $\mu(A) < \infty$  for each bounded measurable set  $A$ . Define  $F(x) = \mu(0, x]$  if  $x \geq 0$ , and  $F(x) = -\mu(x, 0]$  if  $x \leq 0$ . This function satisfies  $\mu(a, b] = F(b) - F(a)$ . If  $\mu$  is a *finite measure*, it is common practice to standardize  $F$  by defining it by  $F(x) = \mu(-\infty, x]$ . In this case,  $\lim_{x \rightarrow -\infty} F(x) = 0$  and  $\lim_{x \rightarrow \infty} F(x) = \mu(\mathbb{R})$ .

**Theorem 3.4.1** If  $F$  is a non-decreasing right-continuous real function on the line, then there exists on  $\mathcal{R}$  a unique measure  $\mu$  satisfying  $\mu(a, b] = F(b) - F(a)$ .

■ **Example 3.2 — Problem 12.1.** Suppose  $\mu$  is a measure on  $\mathcal{R}$  that is finite for bounded sets and  $\mu(A + x) = \mu(A)$ . Then  $\mu(A) = c\lambda(A)$  for all  $A \in \mathcal{R}$  and for some  $c \geq 0$ .

Let  $a, b$  be rationals and  $b - a = \frac{r}{m}$ .

$$\begin{aligned} \mu(a, b] &= \mu(0, b - a] = \mu\left(0, \frac{r}{m}\right] = \sum_{i=1}^r \mu\left(\frac{i-1}{m}, \frac{i}{m}\right] = r\mu\left(0, \frac{1}{m}\right] \\ &= \frac{r}{m}\mu(0, 1] = \mu(0, 1]\lambda(a, b]. \end{aligned}$$

Here we have two measures,  $\mu_1 = \mu$  and  $\mu_2 = c\lambda$  where  $c = \mu(0, 1]$ . Define

$$\mathcal{G} = \{A \in \mathcal{R} : \mu_1(A) = \mu_2(A)\} \supset \mathcal{I}_0 = \{(a, b] : a, b \text{ rational}\}.$$

$\mathcal{I}_0$  is a  $\pi$ -system. We can write the real line as  $\mathbb{R} = \bigcup_{n=1}^{\infty} (-n, n]$ , and  $\mu_1(-n, n] = \mu_2(-n, n] < \infty$ ,  $(-n, n] \in \mathcal{I}_0$ . So the property at the beginning of this example characterized Lebesgue measure in a sense. ■

We can think of measures on  $\mathbb{R}^2$  as well, although we need to be a little careful. Suppose  $\mu$  is a finite measure and  $A$  is a rectangle  $A = \{x : a_i < x_i \leq b_i, i = 1, 2\}$ . Then

$$\mu(A) = F(b_1, b_2) - F(b_1, a_2) - F(a_1, b_2) + F(a_1, a_2).$$

This clearly requires that the right hand side is non-negative for all  $a_i, b_i$ . We need a bit more than monotonicity to satisfy this.

**Definition 3.4.1 — Distribution function.** If  $\mu$  is a probability measure, then  $F$  as described above is called a (*cumulative*) *distribution function*. In this case

$$\lim_{x \rightarrow -\infty} F(x) = 0 \text{ and } \lim_{x \rightarrow \infty} F(x) = 1.$$

**Theorem 3.4.2** If  $F$  is a non-decreasing, right-continuous function satisfying Definition 3.4.1, then there exists on some probability space a random variable (r.v.)  $X$  for which  $F(x) = P(X \leq x)$ .

*Proof.* This follows from Theorem 3.4.1, since there exists a measure  $\mu$  on the real line such that  $\mu(\mathbb{R}) = 1$ . Take  $(\Omega, \mathcal{F}, P) = (\mathbb{R}, \mathcal{R}, \mu)$ , and  $X(\omega) = \omega$ . ■

*Alternative proof.* Define  $\varphi(u) = \inf\{x : u \leq F(x)\}$  for  $0 < u < 1$ . The *quantile function*  $\varphi$  is the inverse  $F^{-1}$ .

By definition of  $\varphi$ , if  $F(x) \geq u$ , then  $x \geq \varphi(u)$ . On the other hand, as  $F$  is non-decreasing,  $x \geq \varphi(u) \implies F(x) \geq F(\varphi(u))$ . Let  $A_u = \{y : F(y) \geq u\}$ . Then there is a sequence  $\{y_k\} \subseteq A_u$  such that  $\lim_k y_k = \inf A_u = \varphi(u)$ . Because  $y_k \in A_u, y_k \geq \varphi(u)$ . By the right continuity of  $F$ , we have  $\lim_k F(y_k) = F(\varphi(u))$ . But we also know that  $F(y_k) \geq u$  for each  $k$ . So  $F(x) \geq F(\varphi(u)) \geq u$ . Hence  $u \leq F(x)$  if and only if  $\varphi(u) \leq x$ .

Thus  $\{x : u \leq F(x)\}$  is the interval  $[\varphi(u), \infty)$ . If we define  $X(\omega) = \varphi(\omega)$  for  $\omega \in (0, 1)$ , then  $X$  is a r.v. and  $P(X \leq x) = F(x)$ . ■

Some properties of the quantile function follow. Let  $\varphi(u-)$  denote the left limit of  $\varphi$  at  $u$ . For  $0 < u < 1$ ,

1.  $F(x) < u$  if and only if  $x < \varphi(u)$ . This follows from the fact (derived in the alternative proof above) that  $u \leq F(x)$  if and only if  $\varphi(u) \leq x$ .
2.  $F(\varphi(u)-) \leq u \leq F(\varphi(u))$ . The fact that  $F(\varphi(u)) \leq u$  was established previously. Also, by property 1, whenever  $x < \varphi(u)$ , we have  $F(x) < u$ . So  $F(\varphi(u)-) \leq u$ .
3. If  $F$  is continuous at  $\varphi(u)$ , then  $F(\varphi(u)) = u$ . This follows from property 2.

Also, if  $X$  has continuous distribution  $F$ , then  $F(X)$  has a uniform distribution on the unit interval. Use properties 1 and 3 to see this.

We now shift focus to integration. Specifically, we will explore the difference between the Riemann integral and the integral with respect to a measure.

**Definition 3.4.2** Let  $f$  be a non-negative measurable function on  $(\Omega, \mathcal{F}, \mu)$ . The *integral* of  $f$  is defined as

$$\int f d\mu = \sup \sum_i \left[ \inf_{\omega \in A_i} f(\omega) \right] \mu(A_i),$$

where the supremum is taken over all finite decompositions  $\{A_i\}$  of  $\Omega$ .

If  $\Omega = \mathbb{R}$  and  $\mu = \lambda$  is Lebesgue measure, then  $\int f d\lambda$  is written  $\int f(x) dx$  and is called the *Lebesgue integral*. The Riemann integral, when it exists, coincides with the Lebesgue integral. If  $\mathbb{Q}$  is the set of rational numbers and  $f = \mathbb{I}_{\mathbb{Q}}$ , then  $f$  is not Riemann integrable, but its Lebesgue integral exists and  $\int f d\lambda = \lambda(\mathbb{Q}) = 0$ .

### 3.5 Lecture 16, October 2

**Theorem 3.5.1** If  $f$  is real and measurable  $\mathcal{F}$ , then there exists a sequence  $\{f_n\}$  of simple functions, each measurable  $\mathcal{F}$ , such that

$$\begin{aligned} 0 \leq f_n(\omega) \uparrow f(\omega) & \text{ if } f(\omega) \geq 0, \text{ and} \\ 0 \geq f_n(\omega) \downarrow f(\omega) & \text{ if } f(\omega) \leq 0. \end{aligned}$$

*Proof.* Use the function

$$f_n(\omega) = \begin{cases} -n & \text{if } -\infty \leq f(\omega) \leq -n, \\ -\frac{k-1}{2^n} & \text{if } -\frac{k}{2^n} \leq f(\omega) \leq -\frac{k-1}{2^n}, 1 \leq k \leq n2^n, \\ \frac{k-1}{2^n} & \text{if } \frac{k-1}{2^n} \leq f(\omega) \leq \frac{k}{2^n}, 1 \leq k \leq n2^n, \\ n & \text{if } n \leq f(\omega) \leq \infty. \end{cases}$$

The difference between  $f_n$  and  $f$  is always less than  $\frac{1}{2^n}$ , so as  $n \rightarrow \infty$ ,  $f_n$  approaches  $f$ . If  $f(\omega) \geq 0$ , then  $f_n$  is always smaller than  $f$ , so  $f_n \uparrow f$ . If  $f(\omega) \leq 0$ , then  $f_n$  is always larger than  $f$ , so  $f_n \downarrow f$ . ■

Next we will look at some properties of integrals.

**Theorem 3.5.2** Suppose  $f, g, f_n$  are all non-negative measurable functions and  $a, b \geq 0$ .

- i) If  $f = \sum_i x_i \mathbb{I}_{A_i}$ , where the  $A_i$  are disjoint, then  $\int f d\mu = \sum_i x_i \mu(A_i)$ .
- ii)  $0 \leq f \leq g \implies 0 \leq \int f d\mu \leq \int g d\mu$ .
- iii)  $0 \leq f_n \uparrow f \implies 0 \leq \int f_n d\mu \uparrow \int f d\mu$ .
- iv)  $\int (af + bg) d\mu = a \int f d\mu + b \int g d\mu$ .

**R** The integral is uniquely defined by i) and iii). Suppose  $\Lambda(\sum_i x_i \mathbb{I}_{A_i} = \sum_i x_i \mu(A_i))$  and  $\Lambda$  also satisfies iii). We just shows that for  $f \geq 0$  there exists a sequence  $\{f_n\}$  of simple functions such that  $0 \leq f_n \uparrow f$ . So by iii),

$$\Lambda(f) = \lim_n \Lambda(f_n) = \lim_n \int f_n d\mu = \int f d\mu.$$

*Proof.* i) Let  $\{B_j\}$  be a finite measurable partition of  $\Omega$ . If  $B_j \cap A_i \neq \emptyset$ , then  $x_i \geq b_j = \inf\{f(\omega) : \omega \in B_j\}$ . So

$$\sum_j b_j \mu(B_j) = \sum_{i,j} b_j \mu(B_j \cap A_i) \leq \sum_{i,j} x_i \mu(B_j \cap A_i) = \sum_i x_i \mu(A_i).$$

Equality holds if the partition  $\{B_j\}$  coincides with  $\{A_i\}$ .

iii) It is enough to show that  $\int f d\mu = \lim_n \int f_n d\mu$  or that

$$S = \sum_{i=1}^m v_i \mu(A_i) \leq \lim_n \int f_n d\mu$$

if  $A_1, \dots, A_m$  is a measurable partition of  $\Omega$  and  $v_i = \inf\{f(\omega) : \omega \in A_i\}$ .

We only consider  $i$  with  $v_i \mu(A_i) > 0$ , since  $S = 0$  if  $v_i \mu(A_i) = 0$  for all  $i$ . Choose  $a_i$  satisfying  $0 < a_i < v_i$ . Clearly  $A_{i,n} = \{\omega \in A_i : f_n(\omega) > a_i\} \subset A_i$  and  $A_{i,n} \uparrow A_i$  as  $f_n \uparrow f$ .  $f_n \geq \sum_i a_i \mathbb{I}_{A_{i,n}}$  implies  $\int f_n d\mu \geq \sum_i a_i \mu(A_{i,n}) \rightarrow \sum_i a_i \mu(A_i)$ . Now let  $a_i \uparrow v_i$  to get the desired result.

iv)

$$\begin{aligned}
\int (af_n + bg_n)d\mu &= \int (\sum_i aa_i \mathbb{I}_{A_i} + \sum_i bb_i \mathbb{I}_{B_i})d\mu \\
&= \sum_i aa_i \mu(A_i) + \sum_i bb_i \mu(B_i) \\
&= a \sum_i a_i \mu(A_i) + b \sum_i b_i \mu(B_i) \\
&= a \int f_n d\mu + b \int g_n d\mu.
\end{aligned}$$

Now  $a \int f_n d\mu \rightarrow a \int f d\mu$  and  $b \int g_n d\mu \rightarrow b \int g d\mu$ , and  $\int (af_n + bg_n)d\mu \rightarrow \int (af + bg)d\mu$ . iv) follows. ■

We can use the above to prove many other interesting properties.

**Notation 3.1.** Almost everywhere, abbreviated *a.e.*, means “outside a null set.”

**Theorem 3.5.3** Suppose  $f, g \geq 0$ . Let  $A = \{\omega : f(\omega) > 0\}$ .

- i)  $f = 0$  a.e.  $\implies \int f d\mu = 0$ .
- ii)  $\mu(A) > 0 \implies \int f d\mu > 0$ .
- iii)  $\int f d\mu < \infty \implies f < \infty$  a.e.
- iv)  $f \leq g$  a.e.  $\implies \int f d\mu \leq \int g d\mu$ .
- v)  $f = g$  a.e.  $\implies \int f d\mu = \int g d\mu$ .

*Proof.* i) Since  $\mu(A) = 0$ ,  $f = f \mathbb{I}_A \leq \infty \cdot \mathbb{I}_A$ . Then  $0 \leq \int f d\mu = \int f \mathbb{I}_A d\mu \leq \infty \cdot \mu(A) = 0$ .

ii) Let  $A_n = \{\omega : f(\omega) > \frac{1}{n}\}$ . Then  $A_n \uparrow A$  and  $f \geq f \mathbb{I}_{A_n} \geq \frac{1}{n} \mathbb{I}_{A_n}$ . So  $\int f d\mu \geq \int f \mathbb{I}_{A_n} d\mu \geq \int \frac{1}{n} \mathbb{I}_{A_n} d\mu \geq \frac{1}{n} \mu(A_n) > 0$  for all large  $n$ .

iii) Let  $B = \{\omega : f(\omega) = \infty\}$ . Then  $f \geq f \mathbb{I}_B = \infty \cdot \mathbb{I}_B$ . So  $\infty > \int f d\mu \geq \infty \cdot \mu(B) \implies \mu(B) = 0$ .

iv) Let  $G = \{\omega : f(\omega) = \infty\}$ . By i),  $\mu(G^c) = 0$  implies  $\int f \mathbb{I}_{G^c} d\mu = \int g \mathbb{I}_{G^c} d\mu = 0$ . So  $\int f d\mu = \int f \mathbb{I}_G d\mu \leq \int g \mathbb{I}_G d\mu = \int g d\mu$ .

v) Trivially follows from iv). ■

### 3.6 Lecture 17, October 5

If  $f$  is measurable, then both  $f^+ = \max(0, f)$  and  $f^- = \max(0, -f)$  are measurable, and  $f = f^+ - f^-$ ,  $|f| = f^+ + f^-$ .

If  $\int f^+ d\mu < \infty$  or  $\int f^- d\mu < \infty$ , then  $\int f(\omega) d\mu(\omega) = \int f(\omega) \mu(d\omega) = \int f d\mu = \int f^+ d\mu - \int f^- d\mu$ . This is called the *definite integral* of  $f$ .

**Definition 3.6.1 — Integrable function.**  $f$  is *integrable* if  $\int |f| d\mu < \infty$ . In this case,  $\int f^+ d\mu + \int f^- d\mu = \int |f| d\mu < \infty$ . So  $f$  is integrable if and only if both  $\int f^+ d\mu$  and  $\int f^- d\mu$  are finite.

If  $f$  has a definite integral and  $f = g$  a.e., then  $\int f d\mu = \int g d\mu$ . This is easily verified:  $f^+ = g^+$  a.e. and  $f^- = g^-$  a.e., so  $\int f^\pm d\mu = \int g^\pm d\mu$ .

**Theorem 3.6.1** Suppose  $f, g$  are integrable functions.

- i)  $f \leq g$  a.e.  $\implies \int f d\mu \leq \int g d\mu$ .
- ii) If  $a$  and  $b$  finite real numbers, then  $(af + bg)$  is integrable and  $\int (af + bg) d\mu = a \int f d\mu + b \int g d\mu$ .
- iii)  $|\int f d\mu| \leq \int |f| d\mu$  and  $|\int f d\mu - \int g d\mu| \leq \int |f - g| d\mu$ .

*Proof.* i) Suppose  $f \leq g$  a.e. This implies that  $f^+ \leq g^+$  a.e. and  $g^- \leq f^-$  a.e. As  $\int f^+ d\mu \leq \int g^+ d\mu < \infty$ , and  $\int g^- d\mu \leq \int f^- d\mu < \infty$ , we have  $\int f d\mu \leq \int g d\mu$ .

ii)  $af + bg$  is integrable, since  $|af + bg| \leq |a||f| + |b||g|$ . For  $a > 0$ ,  $\int af d\mu = \int (af)^+ d\mu - \int (af)^- d\mu = a \int f^+ d\mu - a \int f^- d\mu = a \int f d\mu$ . Similarly for  $a < 0$ . So it is enough to prove the result for one value of  $a$ . For simplicity, take  $a = b = 1$ . Note that  $(f + g)^+ - (f + g)^- = f + g = f^+ + g^+ - f^- - g^-$  implies  $(f + g)^+ + f^- + g^- = (f + g)^- + f^+ + g^+$ . Since these functions are all non-negative,  $\int (f + g)^+ d\mu + \int f^- d\mu + \int g^- d\mu = \int (f + g)^- d\mu + \int f^+ d\mu + \int g^+ d\mu$ . A rearrangement of this completes the proof.

iii) By i) and ii),  $-|f| \leq f \leq |f|$  implies  $-\int |f| d\mu \leq \int f d\mu \leq \int |f| d\mu$ . ■

■ **Example 3.3** Suppose  $\mathcal{F}$  consists of all subsets of  $\Omega = \{1, 2, \dots\}$  and  $\mu$  is the counting measure. A sequence  $\{x_n\}$  of real numbers can be considered a measurable function  $f$ ,  $f(n) = x_n$ . So  $f$  is integrable if and only if the sequence  $\{x_n\}$  is absolutely convergent. The sequence given by  $x_m = (-1)^m \frac{1}{m}$ , for example, is not integrable. ■

■ **Example 3.4** Take  $\Omega$ ,  $\mathcal{F}$ , and  $\mu$  as in the previous example. Let  $f \equiv 0$ ,  $f_n = \mathbb{I}_{\{n, n+1, \dots\}}$ . Then  $\lim_n f_n(m) \rightarrow f(m)$  for all  $m$ , but  $\int f d\mu = 0$  and  $\int f_n d\mu = \infty$  for all  $n$ . ■

■ **Example 3.5** Suppose  $f$  is a Borel function, bounded on bounded intervals. Both  $f_n = f \mathbb{I}_{(-n, n)}$ ,  $g_n = f \mathbb{I}_{(-n, n+1)}$  converge pointwise to  $f$ . Even if the limits  $\lim_n \int f_n d\lambda$  and  $\lim_n \int g_n d\lambda$  exist, they may not be equal. For example,  $f(x) = x$ . ■

■ **Example 3.6** The function  $f = \sum_{k=1}^{\infty} (-1)^k \frac{1}{k} \mathbb{I}_{(k, k+1)}$  has no integral, even though the limit  $\lim_n \int f \mathbb{I}_{(-n, n)} d\lambda$  exists. ■

■ **Example 3.7** Let  $f_n = n^2 \mathbb{I}_{(0, 1/n)}$ , and  $f \equiv 0$ . Then  $f_n \rightarrow f$  everywhere,  $\int f d\lambda = 0$  and  $\int f_n d\lambda = n \rightarrow \infty$ . ■

### 3.7 Lecture 18, October 7

Does  $f_n \rightarrow f$  a.e. imply that  $\int f_n d\mu \rightarrow \int f d\mu$ ? Not always; see Examples 3.4 and 3.7 above. However, under certain conditions this is indeed true.

**Theorem 3.7.1 — Monotone Convergence Theorem.** If  $f_n$  and  $f$  satisfy  $0 \leq f_n \uparrow f$  a.e., then  $\int f_n d\mu \uparrow \int f d\mu$ .

*Proof.* There exists a measurable set  $A$  such that  $\mu(A^c) = 0$  and  $0 \leq f_n \mathbb{I}_A \uparrow f \mathbb{I}_A$ . So  $\int f_n d\mu = \int f_n \mathbb{I}_A d\mu \rightarrow \int f \mathbb{I}_A d\mu = \int f d\mu$ . ■

■ **Example 3.8** For each  $m$ ,  $0 \leq x_{nm} \uparrow x_m$  as  $n \rightarrow \infty$  implies  $\lim_n \sum_m x_{nm} = \sum_m x_m$  (see examples 16.1 and 16.3 in Billingsley). This is false for decreasing limits, however. As a counterexample, let  $y_{nm} = \mathbb{I}_{\{m \geq n\}}$  and  $y_m = 0$ . Then  $0 \leq y_{nm} \downarrow y_m$ , but  $\sum_m y_{nm} = \infty \neq 0 = \sum_m y_m$ . ■

■ **Example 3.9** Consider the measure space  $(\Omega, \mathcal{F}, \mu)$ , and let  $\mathcal{F}_0$  be a sub  $\sigma$ -field of  $\mathcal{F}$ . Let the measure  $\mu_0$  on  $\mathcal{F}_0$  be the restriction of  $\mu$ . Finally, let  $f$  be measurable  $\mathcal{F}_0$ . Then  $\int f d\mu = \int f d\mu_0$ . This holds for  $f = \mathbb{I}_A$  for all  $A \in \mathcal{F}_0$ , so therefore it holds for simple functions, increasing limits of non-negative  $f$ , and for all measurable functions  $f$ , where at least one of  $\int f^+ d\mu$  and  $\int f^- d\mu$  is finite. ■

■ **Example 3.10** If  $\mu$  and  $\mu_n$  are measures and  $f \geq 0$ , and if  $\int f d\mu = \sum_n \int f d\mu_n$ , then  $\mu(A) = \sum_n \mu_n(A)$  for  $A \in \mathcal{F}$ . See example 16.5 in Billingsley. ■

**Theorem 3.7.2 — Fatou's Lemma.** For  $f_n \geq 0$ ,

$$\int \liminf_n f_n d\mu \leq \liminf_n \int f_n d\mu.$$

*Proof.* If  $g_n = \inf_{k \geq n} f_k$ , then  $0 \leq g_n \uparrow g = \liminf_n f_n$ . So  $\int f_n d\mu \geq \int g_n d\mu \rightarrow \int g d\mu$ . ■

The next theorem tells us when limits and integrals can be interchanged.

**Theorem 3.7.3 — Lebesgue's Dominated Convergence Theorem.** If  $|f_n| \leq g$  for some integrable  $g$  and if  $f_n \rightarrow f$  a.e., then  $f$  and  $f_n$  are integrable and  $\lim_n \int f_n d\mu = \int \lim_n f_n d\mu = \int f d\mu$ .

*Proof.* Since  $g + f_n$  and  $g - f_n$  are non-negative, Fatou's Lemma (Theorem 3.7.2) tells us that

$$\begin{aligned} \int g d\mu + \int \liminf_n f_n d\mu &= \int \liminf_n (g + f_n) d\mu \\ &\leq \liminf_n \int (g + f_n) d\mu \\ &= \int g d\mu + \liminf_n \int f_n d\mu \end{aligned}$$

and

$$\begin{aligned} \int g d\mu - \int \liminf_n f_n d\mu &= \int \liminf_n (g - f_n) d\mu \\ &\leq \liminf_n \int (g - f_n) d\mu \\ &= \int g d\mu - \liminf_n \int f_n d\mu. \end{aligned}$$

Since  $g$  is integrable,  $\int g d\mu$  is finite and non-negative, and thus we can cancel it. So

$$\int \liminf_n f_n d\mu \leq \liminf_n \int f_n d\mu \leq \limsup_n \int f_n d\mu \leq \int \limsup_n f_n d\mu.$$

■

Below are some properties of integrals.

1. **The Weierstrass M-test for series.** Suppose for each  $m$ ,  $x_{nm} \rightarrow x_m$  as  $n \rightarrow \infty$  and  $|x_{nm}| \leq M_m$ . If  $\sum_m M_m < \infty$ , then  $\sum_m x_{nm} \rightarrow \sum_m x_m$ .
2. **Bounded convergence theorem.** If  $\mu$  is a finite measure and the  $f_n$  are uniformly bounded, then  $f_n \rightarrow f$  a.e. implies that  $\int f_n d\mu \rightarrow \int f d\mu$ .
3. If  $f_n \geq 0$ , then  $\int \sum_n f_n d\mu = \sum_n \int f_n d\mu$ .
4. If  $\sum_n f_n$  converges a.e. and  $|\sum_{k=1}^n f_k| \leq g$  for some integrable  $g$ , then  $\sum_n f_n$  and  $f_n$  are integrable and  $\int \sum_n f_n d\mu = \sum_n \int f_n d\mu$ .
5. If  $\sum_n \int |f_n| d\mu < \infty$ , then  $\sum_n f_n$  converges absolutely a.e. and is integrable, and  $\int \sum_n f_n d\mu = \sum_n \int f_n d\mu$ .
6. The integral of  $f$  over a set is defined as  $\int_A f d\mu = \int \mathbb{I}_A f d\mu$ . If  $A_1, A_2, \dots$  are disjoint, and if  $f$  is either non-negative or integrable, then  $\int_{\cup_n A_n} f d\mu = \sum_n \int_{A_n} f d\mu$ .

### 3.8 Lecture 19, October 9

One way to guarantee the convergence of a sequence of functions is to determine that the integral of the tail of the sequence goes to zero uniformly. How can we show the convergence of integrals?

**Definition 3.8.1** The sequence  $\{f_n\}$  is *uniformly integrable* (u.i.) if

$$\lim_{\alpha \rightarrow \infty} \sup_n \int_{\{|f_n| \geq \alpha\}} |f_n| d\mu = 0.$$

If  $\sup_n \int |f_n|^\beta d\mu < \infty$  for some  $\beta > 1$ , then  $\{f_n\}$  is u.i. Notice that because the integral in Definition 3.8.1 is taken over the set on which  $\{|f_n| \geq \alpha\}$ , the quantity  $\frac{|f_n|}{\alpha}$  is always non-negative. So we can rewrite as

$$\begin{aligned} \int_{\{|f_n| \geq \alpha\}} |f_n| d\mu &\leq |f_n| \left( \frac{|f_n|}{\alpha} \right)^{\beta-1} d\mu \\ &\leq \frac{1}{\alpha^{\beta-1}} \int |f_n|^\beta d\mu \\ &\leq \frac{1}{\alpha^{\beta-1}} \sup_{k \geq 1} \int |f_k|^\beta d\mu \rightarrow 0 \text{ as } \alpha \rightarrow \infty. \end{aligned}$$

If  $\{f_n\}$  and  $\{g_n\}$  are uniformly integrable, then so is  $\{f_n + g_n\}$ . We aim to show that  $\int_{\{|f+g| \geq 2\alpha\}} |f+g| d\mu \rightarrow 0$ . Use the fact that  $|f+g| \leq |f| + |g| \leq 2h$ , where  $h = \max(|f|, |g|)$ . So taking region of integration into account, we have  $2\alpha \leq |f+g| \leq 2h$ , and so  $\alpha \leq h$ . Therefore

$$\begin{aligned} \int_{\{|f+g| \geq 2\alpha\}} |f+g| d\mu &\leq 2 \int_{\{h \geq \alpha\}} h d\mu \\ &\leq 2 \int_{\{|f| \geq \alpha\}} |f| d\mu + 2 \int_{\{|g| \geq \alpha\}} |g| d\mu. \end{aligned}$$

So

$$\sup_n \int_{\{|f_n+g_n| \geq 2\alpha\}} |f_n+g_n| d\mu \leq 2 \sup_n \int_{\{|f_n| \geq \alpha\}} |f_n| d\mu + 2 \sup_n \int_{\{|g_n| \geq \alpha\}} |g_n| d\mu \xrightarrow{\alpha \rightarrow \infty} 0.$$

**Theorem 3.8.1** Suppose  $\mu$  is a finite measure and  $f_n \rightarrow f$  a.e.

- a) If  $\{f_n\}$  is u.i., then  $f$  is integrable and  $\int f_n d\mu \rightarrow \int f d\mu$ .
- b) If  $f, f_n$  are non-negative and integrable, then  $\int f_n d\mu \rightarrow \int f d\mu$  implies  $\{f_n\}$  is u.i.

*Proof.* a) First of all,

$$\begin{aligned} \int |f_n| d\mu &\leq \alpha \mu(\Omega) + \sup_k \int_{\{|f_k| \geq \alpha\}} |f_k| d\mu \\ &\leq \alpha \mu(\Omega) + 1 \end{aligned}$$

since the supremum in the right hand side is less than 1 for large enough  $\alpha$ . This is finite, so we can use Fatou's Lemma to see that

$$\int |f| d\mu \leq \liminf_n \int |f_n| d\mu < \infty$$

and thus  $f$  is integrable. If  $\mu(|f| = \alpha) = 0$ , then  $f_n \mathbb{I}_{\{|f_n| < \alpha\}} \rightarrow f \mathbb{I}_{\{|f| < \alpha\}}$  a.e. and by the bounded convergence theorem,  $\int f_n \mathbb{I}_{\{|f_n| < \alpha\}} d\mu \rightarrow \int f \mathbb{I}_{\{|f| < \alpha\}} d\mu$ . Uniform integrability of  $f_n$  and  $f$  give us that the other parts of the integrals converge. This yields the result.

b) If  $\mu(|f| = \alpha) = 0$ , then  $\int f_n \mathbb{I}_{\{|f_n| \geq \alpha\}} d\mu \rightarrow \int f \mathbb{I}_{\{|f| \geq \alpha\}} d\mu$ . The result follows since  $f$  is integrable. ■



**Theorem 3.8.2** Suppose  $\mu$  is a finite measure,  $f$  and  $f_n$  are integrable, and that  $f_n \rightarrow f$  a.e. The following are equivalent:

- i)  $\{f_n\}$  is u.i.
- ii)  $\int |f_n - f| d\mu \rightarrow 0$ .
- iii)  $\int |f_n| d\mu \rightarrow \int |f| d\mu$ .

*Proof.* i) implies the uniform integrability of  $\{|f_n - f|\}$ . So ii) follows by the previous theorem. ii) implies iii), since  $||f_n| - |f|| \leq |f_n - f|$ . Finally, i) follows from iii) using the theorem above. ■

As a final note on this topic, if  $|f_n| \leq g$  for some integrable  $g$ , then  $\{f_n\}$  is u.i. However, the converse is not true. Consider the measure space  $((0, 1], \mathcal{B}, \lambda)$ .  $\lambda$  is a finite measure. Take the sequence of functions

$$f_n = \frac{n}{\log n} \mathbb{I}_{(0, 1/n)}, \quad n \geq 3.$$

The  $f_n$  are u.i. because the integral is simply  $1/\log n$ , which goes to zero. However, the sequence is not dominated by any integrable function  $g$ .





## Index

Monotone class, 20

Probability, 10