More about events spaces, specifically, σ-algebras and in particular the Borel σ-algebra

References:

Gubner: Note 1 of Chapter 1. Problems 36-39.

R. M. Gray and L. D. Davisson, *An Introduction to Statistical Signal Processing.* This is the only reference that discusses this material at the level of this course. It is available for free download at: www-ee.stanford.edu/~gray/sp.html.

Math 597 or Sat 620 for the full story.

- The Problem: Unfortunately, we can't always take the event space to be the power set $\mathsf{P}(\Omega)$ of the sample space Ω, i.e. set of all subsets of the sample space Ω.

Example 1: There are many random experiments that are well modelled by choosing $\Omega = [0,1]$ and having the probability measure assign probability to subintervals of Ω in the following way

$$P((a,b)) = b-a, \quad \text{when } 0 \le a < b \le 1 \qquad **$$

Though it is much beyond the scope of this course, it can be shown that there is no probability measure on $\mathsf{P}(\Omega)$ that assigns probability to intervals according to **, i.e. there is no function defined on $\mathsf{P}(\Omega)$ that satisfies the axioms and **.

Example 2: There are many random experiments that are well modelled by choosing $\Omega = (-\infty,\infty)$ and having the probability measure assign probability to subintervals of Ω in the following way

$$P((a,b)) = \int_a^b f(x)\,dx \qquad ***$$

where $f(x)$ is some nonnegative function that integrates to one. Again, it can be shown that there is no probability measure on $\mathsf{P}(\Omega)$ that assigns probability to intervals according to ***, i.e. there is no function defined on $\mathsf{P}(\Omega)$ that satisfies the axioms and ***.

The "problem" is that there are some incredibly complex sets in $\mathsf{P}(\Omega)$, making it impossible to extend P from intervals to all sets in $\mathsf{P}(\Omega)$ in a way that satisfies the axioms. This is not at all intuitive. It is not easy to describe the nasty sets that cause problems. This was an unnerving discovery, made around 1900, I think.

In practice we virtually never encounter sets for which the probability axioms fail. That is, we virtually always work with sets that are intervals or are the union of a finite or countably infinite number of intervals, for which ** and *** work fine when determining probabilities. Thus in practice, the existence of this problem is not an issue. Nevertheless, we want to know that all sets in our event space have well defined probabilities that satisfy the axioms, because we want to be convinced that probability theory has a sound basis.

The discussion that follows is a sketch of the way to resolve this issue. Some of the steps are beyond the scope of this course. However, it is good to be aware of the general outline of this sketch.

Note: Example 1 is a special case of Example 2: If

$$f(x) = \begin{cases} 1, & 0 \le x \le 1 \\ 0, & \text{else} \end{cases}$$

then *** reduces to **. We'll focus on Example 1, but the story extends to Example 2 and beyond, e.g. to Ω = cartesian plane.

- The first idea in resolving this problem is to choose the event space to be smaller than the power set. However it should not be too small.

  1. It should include all intervals and unions of intervals, but it should avoid the incredibly complex sets that make difficulties for the power set.
  2. It should be a σ-algebra. That is, the event space E should be "closed" in the following ways:

     a. If $F \in \mathsf{E}$, then $F^c \in \mathsf{E}$

     b. If $F, G \in \mathsf{E}$, then $F \cup G \in \mathsf{E}$

     c. If $F_1, F_2, \ldots \in \mathsf{E}$, then $\cup_{i=1}^{\infty} F_i \in \mathsf{E}$

     Note that the power set is a σ-algebra, but it is too large.

     Note that the set of all intervals is not a σ-algebra.

- What σ-algebra to use? Not just any smaller σ-algebra will permit there to be a valid probability measure that satisfies **.

  Approach: Identify a collection C of elementary events of prime importance. In the examples above, C contains all subintervals of Ω, specifically all subintervals of Ω of the form (a,b), $[a,b]$, $(a,b]$, and $[a,b)$.

  Let $\mathsf{E} = \sigma(C)$, where

  $$\sigma(C) = \text{smallest } \sigma\text{-algebra containing } C$$

  is called the "σ-algebra generated by C". That is $\sigma(C)$ is a σ-algebra, $C \subset \sigma(C)$, and $\sigma(C) \subset \mathsf{F}$ if F is any other σ-algebra containing C.

  How do we know there is a smallest σ-algebra? Here's the reason.

  One can show that $\sigma(X)$ = intersection of all σ-algebras containing C. Here's the idea: There's at least one σ-algebra containing C, namely, the power set.

(In fact there are usually uncountably many σ-algebras containing C.) The intersection of all σ-algebras containing C is not empty because all σ-algebras containing C contain $\{\phi, \Omega\}$. So the intersection at least contains $\{\phi, \Omega\}$. The intersection of any number of σ-algebras is a σ-algebra. (In Homework 2, you will prove a special case, namely, that the intersection of two σ-algebras is a σ-algebra.) Clearly, the intersection of all σ-algebras containing C is a σ-algebra contained in any other σ-algebra containing C. Therefore, it is the smallest σ-algebra containing C.

- In our case, Ω is an interval, C contains the subintervals of Ω, and $\sigma(C)$ is called the "Borel σ-algebra for Ω", denoted B. The sets in B are called "Borel sets".
- Now we need to specify a function on B that satisfies ** and the axioms (i.e. that is a probability measure). Alternatively, it suffices to prove that there exists such a function.

a. We start by using ** to assign probability to intervals:

$$P(\,[a,b]\,) = P(\,(a,b)\,) = P(\,(a,b]\,) = P(\,[a,b)\,) = b-a, \quad 0 \le a \le b \le 1$$

b. Next we assign probabilities to unions of finitely many intervals.
Any such set can be expressed as

$$F = I_1 \cup I_2 \cup \ldots \cup I_N$$

where N is some integer and I_i's are disjoint intervals. And we assign

$$P(F) = P(I_1) + \ldots + P(I_N)$$

c. It can be verified that the collection of sets that are finite unions of intervals and singletons plus the empty set is an "algebra", denoted A. (An algebra is a collection of sets that is closed under complements and finite unions, but not necessarily under countable unions.) Moreover, it can be verified that $\sigma(\mathsf{A}) = \sigma(C) = \mathsf{B}$.

d. It can be straightforwardly verified that the definition of P satisfies the axioms, whenever the sets in question come from A. (Even countable additivity holds, when all the sets and their union comes from A.)

e. The Cartheodory extension theorem, stated below, shows that there is a unique way to define P on B, the entire Borel σ-algebra, in such a way that P satisfies the axioms on B and agrees with the probabilities already defined on intervals and, more generally, on the algebra A.. We say that P has been "extended" from A to B. The probability measure obtained in this way is called the "Borel measure".

f. For technical reasons it is customary to add to $\mathcal{B}$ a collection of trivial sets, namely, all subsets of all sets in $\mathcal{B}$ with Borel measure zero. It can be shown that the P extends uniquely to this enlarged σ-algebra. The resulting extended measure is called the "Lebesgue measure", which you may have heard of. Or perhaps you've heard of Lebesgue integration, which is based on Lebesgue measure.

Cartheodory Extension Theorem: (Ash, RAP, p. 19)

A probability measure P on an algebra $\mathcal{A}$ (by definition it satisfies all the axioms when the sets in question come from $\mathcal{A}$) has a unique extension to a probability measure on $\sigma(\mathcal{A})$. That is, there is a unique probability measure $\overline{P}$ on $\sigma(\mathcal{A})$ such that $P(F) = \overline{P}(F)$ for all $F \in \mathcal{A}$.

The proof is much beyond the scope of this class.

- This has been a brief overview. You are not expected to understand it well. But now that you are aware of the situation, you have the opportunity to pursue this further in the future.

Properties of σ-algebras

- Recall: A <u>σ-algebra</u> is a collection of subsets of Ω such that

  1. $A^c \in E$ whenever $A \in E$
  2. $A \cup B \in E$ whenever $A,B \in E$
  3. $\bigcup_{i=1}^{\infty} A_i \in E$ whenever $A_1, A_2, ... \in E$

  i.e. it is closed under complementation and unioning.

- An <u>algebra</u> is a collection of subsets that satisfy the first two conditions.
- Every σ-algebra is an algebra, but not vice versa.
- When Ω is finite, every algebra is an algebra.
- Example of an algebra that is not a σ-algebra.

  $\Omega = (-\infty,\infty)$. E contains all finite unions of intervals of all kinds, plus ϕ, plus all singletons

- Properties of a σ-algebra that derive from its definition

  4. $\Omega \in E$, since $\Omega = A \cup A^c$
  5. $\phi \in E$, since $\phi = \Omega^c$
  6. $A \cap B \in E$ whenever $A,B \in E$, since $A \cap B = (A^c \cup B^c)^c$
  7. $A - B \in E$ whenever $A,B \in E$, since $A\text{-}B = A \cap B^c$
  8. $\bigcap_{i=1}^{\infty} A_i \in E$, whenever $A_1, A_2, ... \in E$, since $\bigcap_{i=1}^{\infty} A_i = \left(\bigcup_{i=1}^{\infty} A_i^c\right)^c$

A procedure for fiinding the σ-algebra generated by a collection of sets.

Though we have introduced generated σ-algebras in the context of intervals and he Borel σ-algebra, there are occasions where we want to use the generated σ-algebra concept in simpler situations. Here we discuss a procedure for finding the σ-algebra generated by a finite collection of sets. It also works for some countable sets. Working with this procedure generally helps one to gain a better understanding of the concepts of σ-algebra and generated σ-algebra.

Recall:

$\sigma(C)$ = σ-algebra generated by C = smallest σ-algebra containing C

= intersection of all σ-algebras containing C

1. Suppose C is a partition into a finite collection of sets. Then $\sigma(C)$ consists of all sets formed by taking finite unions of members of C, plus the empty set (the empty union).

   Example of using the procedure:

   $\Omega = \{1, 2, 3, 4\}$

   $C = \{ \{1\}, \{2,3\} \}$

   partition = $\{ \{1\}, \{2,3\}, \{4\} \}$

   $\sigma(C) = \{ \phi, \{1\}, \{2,3\}, \{4\}, \{1,2,3\}, \{1,4\}, \{2,3,4\}, \{1,2,3,4\} \}$

   single cells unions of pairs of cells unions of 3 cells

2. If C is a finite collection that is not a partition, then find a partition, such that every member of C is a union of elements of C, and every member of the partition can be formed by union, complement, intersection, subtraction operations on elements of C. (This is the finest partition formable from C. It is also the coarsest partition from which C can be formed.)

   Then $\sigma(C)$ is formed from this partition in the same way as before.

   Example:

   $\Omega = \{1, 2, 3, 4\}$

   $C = \{ \{1,2\}, \{2\} \}$

   partition = $\{ \{1\}, \{2\}, \{3,4\} \}$

   $\sigma(C) = \{ \phi, \{1\}, \{2\}, \{3,4\}, \{1,2\}, \{1,3,4\}, \{2,3,4\}, \{1,2,3,4\} \}$

   single cells unions of pairs of cells unions of 3 cells

3. This method sometimes but not always works when C is a countably infinite collection.

Example when it works:

$\Omega = \{1,2,3, \dots \}$

$C = \{ \{1,2,3,4\}, \{3,4,5,6\}, \{5,6,7,8\}, \dots \}$

partition = $\{ \{1,2\}, \{3,4\}, \{5,6\}, \dots \}$

$\sigma(D)$ consists of all sets that are finite or countable unions of these

Example when it does not work:

$\Omega = [0,1]$

C = all intervals with rational endpoints.

4. This method never works when C contains uncountably many sets.

More about the Borel σ-algebra

We defined the Borel σ-algebra to be σ-algebra generated by the set of all intervals. However, it turns out that there are many generating sets for the Borel σ-algebra. This is nice because it indicates that the Borel σ-algebra is a rather robustly defined quantity. Each of the following is a generating set of the Borel σ-algebra on $\Omega = (-\infty,\infty)$:

The set of all sub intervals of all kinds.

The set of all open intervals, i.e. intervals of the form (a,b)

The set of all closed intervals, i.e. intervals of the form [a,b]

The set of all half open intervals of the form [a,b)

The set of all half open intervals of the form (a,b]

The collection of all finite unions of intervals of any one of the above kinds.

The set of all open subsets.

Generalizations of the Borel σ-algebra and Borel measure:

We have so far focused on the Borel σ-algebra and the Borel measure on Ω = [0,1]. However, as indicated below these concepts extend much beyond this.

1. The Borel σ-algebra $\mathcal{B}$ on the interval $\Omega = (u,v)$, $-\infty \le u < v \le \infty$ is the σ-algebra generated by all intervals of all types in (u,v).

   In this case the Borel measure is specified by

   $$P((a,b)) = \frac{b-a}{v-u}, \quad u \le a < b \le v \qquad ++$$

   For the Borel σ-algebra on (u,v), instead of specifying the probability measure as above, for many random experiments, it may be more appropriate to specify a function nonnegative function f such that $\int_u^v f(x)\,dx = 1$ and to define

   $$P([a,b]) = \int_a^b f(x)\,dx \quad u \le a < b \le v \qquad +++$$

   Via the Cartheodory Extension Theorem, this probability assignment extends uniquely to a probability measure on the Borel σ-algebra on (u,v). This is Borel measure only if $f(x) = 1$, $u < x < v$. In other cases, there is no particular name for the probability measure. The function f is called a probability density, because like any density, it is a quantity that one integrates to compute the total amount.

   This method works when $(u,v) = (-\infty,\infty)$.

   When describing a probability model, one normally has to specify the probability of all sets in the event space. This can be a daunting task when the σ-algebra is large. However, when the event space is a Borel σ-algebra, then we see from the Cartheodory extension theorem that it suffices to specify the probability measure on the algebra $\mathcal{A}$ consisting of all finite unions of intervals and singletons plus the empty set. In fact, usually it suffices just to specify the probability of intervals, provided this is done in a way that the probability of sets in $\mathcal{A}$ satisfy the axioms. This will always be the case if probability of intervals is specified as in ++ or +++ above.

2. The Borel σ-algebra on the Cartesian plane.

   Suppose Ω = some rectangle in the Cartesian plan

   Then the Borel σ-alg $\mathcal{B}$ for Ω is the σ-algebra generated by “subsquares” in Ω. It is also the σ-algebra generated by the “subrectangles”, or by the “subtriangles”, or by the “subcircles”, and so on.

   The Borel measure on $\mathcal{B}$ is the unique extension of the following defined for squares, rectangles and other elementary shapes.

$$P(F) = \frac{\text{area of } F}{\text{area of } \Omega} \qquad ++$$

Alternatively, one can specify a probability measure by specifying a nonnegative function $f(x,y)$ on Ω that integrates to one and defining

$$P(F) = \int_F f(x,y)\, dx\, dy \qquad +++$$

The function f is again called a probability density. The probability measure is not called a Borel measure.

Once again the above two formulas are sufficient to uniquely specify the entire probability measure.

It is not necessary for Ω to be a rectangle. It could be any set with a well defined area, even the entire Cartesian plane. However, ++ applies only when Ω has finite area.

3. The Borel σ-algebra in higher dimensions

The concepts above generalize directly. If Ω is some subset of N-dimensional space with well defined and finite area, then

Borel measure on B is the unique extension of the following defined for squares, rectangles and other elementary shapes.

$$P(F) = \frac{\text{area of } F}{\text{volume of } \Omega} \qquad ++$$

Alternatively, one can specify a probability measure by specifying a nonnegative function $f(x_1,...,x_N)$ on Ω that integrates to one and defining

$$P(F) = \int_F f(x_1,...,x_N)\, dx_1...dx_N \qquad +++$$

The function f is again called a probability density. The probability measure is not called a Borel measure.

Once again the above two formulas are sufficient to uniquely specify the entire probability measure.