MASSACHUSETTS INSTITUTE OF TECHNOLOGY

Fall 2005 6.436J/15.085J

Notes from the 9/12 lecture.

1 Introduction

We first establish that probability measures have a certain continuity property.

We then move to the construction of two basic probability models: (a) A model involving an infinite sequence of independent tosses of a fair coin; (b) A model in which the outcome is "uniformly distributed" over the unit interval $[0, 1]$.

These two models are often encountered in elementary probability and used without further discussion. Rigorously speaking, however, one needs to be sure that these models are well-posed and consistent with the axioms of probability. In other words, one needs to establish that there exist probability spaces that correspond to these models.

The complete proof of the existence of such probability spaces requires quite a bit of technical development (see [W]). In this handout, we go through the steps of this development, omitting most of the proofs.

2 Continuity of probabilities

Consider a probability model in which $\Omega = \Re$. We would like to be able to assert that the probability of the event $[1/n, 1]$ converges to the probability of the event $(0, 1]$, as $n \to \infty$. This is accomplished by the following theorem.

Theorem 1: Let $\mathcal{F}$ be a σ-field of subsets (called "$\mathcal{F}$-measurable sets") of a sample space Ω. Let $\mathbf{P}$ be a function on $\mathcal{F}$ such that $\mathbf{P}(\Omega) = 1$, $\mathbf{P}(A) \geq 0$ for every $A \in \mathcal{F}$, and such that $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B)$ whenever A, B are disjoint elements of $\mathcal{F}$. [This latter property of $\mathbf{P}$ is called "finite additivity."] The following are equivalent:

(a) $\mathbf{P}$ is a probability measure (that is, it also satisfies countable additivity).

(b) If $A_i \in \mathcal{F}$ is an increasing family of sets ($A_i \subseteq A_{i+1}$, for all i), and $A = \cup_{i=1}^{\infty} A_i$, then $\lim_{n\to\infty} \mathbf{P}(A_i) = \mathbf{P}(A)$.

(c) If $A_i \in \mathcal{F}$ is a decreasing family of sets ($A_i \supseteq A_{i+1}$, for all i), and $A = \cap_{i=1}^{\infty} A_i$, then $\lim_{n\to\infty} \mathbf{P}(A_i) = \mathbf{P}(A)$.

(d) If $A_i \in \mathcal{F}$ is a decreasing family of sets ($A_i \supseteq A_{i+1}$, for all i) and $\cap_{i=1}^{\infty} A_i$ is empty, then $\lim_{n\to\infty} \mathbf{P}(A_i) = 0$.

Proof: That (b) follows from (a) is Lemma 5 in [GS]. To establish (c) from (b), just use de Morgan's law. Statement (d) follows from (c) because it is just a special case in which A is empty. It remains to show that (d) implies (a).

Let $B_i \in \mathcal{F}$ be disjoint events. Let $A_n = \cup_{i=n}^{\infty} B_i$. Note that A_n is a decreasing sequence of events and converges to the empty set. [Intuitively, any element of A_1 belongs to some B_n, which means that it does not belong to A_{n+1}. So, every element of A_1 is outside some A_{n+1}, so the intersection of the latter sets is empty.]

Assume that property (d) holds. Using finite additivity for the finitely many $(n+1)$ sets $B_1, B_2, \ldots, B_n, \cup_{i=n+1}^{\infty} B_i$, we have

$$\mathbf{P}\left(\bigcup_{i=1}^{\infty} B_i\right) = \sum_{i=1}^{n} \mathbf{P}(B_i) + \mathbf{P}\left(\bigcup_{i=n+1}^{\infty} B_i\right).$$

This equality holds for any n. Now let $n \to \infty$. The first term on the right-hand side converges to $\sum_{i=1}^{\infty} \mathbf{P}(B_i)$. The second term converges to zero, because $A_n = \cup_{i=n+1}^{\infty} B_i$ converges to the empty set and statement (d) applies. **q.e.d.**

This theorem tells us that when checking for countable additivity, it is enough to check that $\mathbf{P}$ behaves "continuously" along a decreasing sequence of events with empty intersection.

3 The infinite coin toss model

Consider an infinite sequence of fair coin tosses, with each toss being equally likely to result in heads or tails (recorded as 1 and 0, respectively). The sample space for this experiment is the set $\{0,1\}^{\infty}$ of all infinite sequences $\omega = (\omega_1, \omega_2, \ldots)$ of zeroes and ones.

Let $\mathcal{F}_n$ be the set of events whose occurrence can be decided by looking at the results of the first n tosses. For example, the event $\{\omega \mid \omega_1 = 1 \text{ and } \omega_2 \neq \omega_4\}$ belongs to $\mathcal{F}_4$ (as well as to $\mathcal{F}_k$ for every $k \geq 4$). Let A be a subset of $\{0,1\}^n$. Consider the event

$$\{\omega \in \{0,1\}^{\infty} \mid (\omega_1, \omega_2, \ldots, \omega_n) \in A\}.$$

This event belongs to $\mathcal{F}_n$ and all elements of $\mathcal{F}_n$ are of this form, for some A.

Let $A_n = \{\omega \mid \omega_n = 1\}$. The event $A = \cup_{i=1}^{\infty} A_n$ is the event that there is at least one "1" in the infinite toss sequence. We would like to be able to assign a probability to the event A. Note that $A_n \in \mathcal{F}_n$ for all n, but A does not belong to $\mathcal{F}_n$ for any n.

We define $\mathcal{F}_0 = \cup_{i=1}^{\infty} \mathcal{F}_i$. So, an event B belongs to $\mathcal{F}_0$ if and only if it belongs to $\mathcal{F}_n$ for some n. In particular, the event A above does not belong to $\mathcal{F}_0$, even though A is the union of events in $\mathcal{F}_0$. Thus, $\mathcal{F}_0$ is not a σ-field.

Note: The union $\cup_{i=1}^{\infty} \mathcal{F}_i = \mathcal{F}_0$ is not the same as the collection of sets of the form $\cup_{i=1}^{\infty} B_i$, for $B_i \in \mathcal{F}_i$. For example, the set A discussed earlier is of the latter form but is not in $\mathcal{F}_0$. For a more concrete example, if $\mathcal{F}_1 = \{\{a\}, \{b,c\}\}$ and $\mathcal{F}_2 = \{\{d\}\}$, then $\mathcal{F}_1 \cup \mathcal{F}_2 = \{\{a\}, \{b,c\}, \{d\}\}$, so that $\{b,c,d\}$ is not in $\mathcal{F}_1 \cup \mathcal{F}_2$.

We would like to have a probability model that assigns probabilities to all of the events in the collections $\mathcal{F}_n$. This means that we need a σ-field that includes $\mathcal{F}_0$. But we would like to keep our σ-field as small as possible. (If it is too large, we may run into trouble when trying to assign a legitimate probability to all of the events in the field.) This is accomplished next.

4 The σ-field generated by a collection of sets.

Theorem 2: Given a collection $\mathcal{C}$ of subsets of Ω, there exists a σ-field $\mathcal{F}$ [also denoted $\sigma(\mathcal{C})$ and called the σ-field generated by $\mathcal{C}$] which is the smallest possible σ-field that includes all elements of $\mathcal{C}$. [That is, if $\mathcal{G}$ is any other σ-field and $\mathcal{C} \subseteq \mathcal{G}$, then $\mathcal{F} \subseteq \mathcal{G}$.]

Proof: Consider all σ-fields $\mathcal{G}$ that contain $\mathcal{C}$. [There is at least one such $\mathcal{G}$, namely, the set of all subsets of Ω.] Let $\mathcal{F}$ be the intersection of all such $\mathcal{G}$. [That is, a subset A of Ω belongs to $\mathcal{F}$ if and only if it belongs to every σ-field $\mathcal{G}$ such that $\mathcal{C} \subseteq \mathcal{G}$.]

We first verify that $\mathcal{F}$ is indeed a σ-field. Fix any σ-field $\mathcal{G}$ that includes all of $\mathcal{C}$. If $A_1, A_2, \ldots \in \mathcal{F}$, then $A_i \in \mathcal{G}$, for all i. Since $\mathcal{G}$ is a σ-field, it follows that $A_i^c \in \mathcal{G}$ and that $\cup_{i=1}^{\infty} A_i \in \mathcal{G}$. Since such complements and unions belong to every σ-field $\mathcal{G}$ that includes the elements of $\mathcal{C}$, it follows that such complements and unions are also in $\mathcal{F}$, as required. Thus $\mathcal{F}$ is indeed a σ-field.

If $\mathcal{H}$ is any other σ-field that contains $\mathcal{C}$, then by definition $\mathcal{F} \subseteq \mathcal{H}$ (since $\mathcal{F}$ was defined as the intersection of various σ-fields, one of which is just $\mathcal{H}$). **q.e.d.**

5 The extension theorem

Theorem 3: Suppose that $\mathcal{F}_0$ is a field. [That is, $\emptyset \in \mathcal{F}_0$. And if $A, B \in \mathcal{F}_0$, then $A^c \in \mathcal{F}_0$ and $A \cup B \in \mathcal{F}_0$. A field is not necessarily a σ-field, as it is not closed under countable unions.]

Suppose that $\mathbf{P}_0$ is a mapping from $\mathcal{F}_0$ to $[0, 1]$ with the following properties: (i) $\mathbf{P}_0(\Omega) = 1$ and (ii) if $A_i \in \mathcal{F}_0$ are disjoint sets such that $\cup_{i=1}^{\infty} A_i \in \mathcal{F}_0$, then $\mathbf{P}_0(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbf{P}_0(A_i)$. [In other words $\mathbf{P}_0$ satisfies all the axioms required of probability measures, but is just defined on sets A in $\mathcal{F}_0$.]

Then, there exists a unique probability measure $\mathbf{P}$ on $(\Omega, \sigma(\mathcal{F}_0))$, which is consistent with $\mathbf{P}_0$ on $\mathcal{F}_0$ [that is, $\mathbf{P}(A) = \mathbf{P}_0(A)$ for every $A \in \mathcal{F}_0$].

Remark 1: The general scheme is that we start with a collection $\mathcal{F}_0$ of "interesting" sets for which probabilities are defined. We extend $\mathcal{F}_0$ to make it a σ-field, which is required in order to have a legitimate probability space. And we then assign probabilities to every event in the σ-field, in a manner which is consistent with the probabilities for events in $\mathcal{F}_0$. The theorem states that (under the stated assumptions on $\mathcal{F}_0$ — it is a field — and on $\mathbf{P}_0$ — it is countably additive on $\mathcal{F}_0$), this is can be done, and it can be done in a unique way.

Remark 2: The proof of the extension theorem is long and involved.
Renark 3: To apply the theorem one needs to verify countable additivity of $\mathbf{P}_0$ on $\mathcal{F}_0$. Alternatively, in the spirit if Theorem 1, one need only check that $\mathbf{P}_0$ is finitely additive, and that if A_i is a decreasing sequence of sets in $\mathcal{F}_0$ that converges to the empty set, then $\mathbf{P}(A_i)$ converges to zero.

6 Back to the coin tossing model

Recall that $\mathcal{F}_0$ was defined to be $\cup_{n=1}^{\infty}\mathcal{F}_n$. Thus, a set A is in $\mathcal{F}_0$ if and only if there is some n such that the occurrence of the event can be decided on the basis of the first n flips. It is easily checked that $\mathcal{F}_0$ is a field. [Indeed, if $A \in \mathcal{F}_0$, then $A \in \mathcal{F}_n$ for some n, and then $A^c \in \mathcal{F}_n$ for that same n, which implies that $A^c \in \mathcal{F}_0$. Furthermore, if $A, B \in \mathcal{F}_0$, there is some m and n such that $A \in \mathcal{F}_m$ and $B \in \mathcal{F}_n$. This implies that $A \cup B \in \mathcal{F}_{\max\{m,n\}} \subset \mathcal{F}_0$.]

If $A \in \mathcal{F}_0$, then A is an event whose occurrence can be decided by the first n coin tosses, for some n. We let every n-toss sequence be equally likely (with probability $1/2^n$). It can be verified that this leads to a well-defined probability $\mathbf{P}_0(A)$ for every $A \in \mathcal{F}_0$. [Verifying this fact involves the following: If a set A belongs to $\mathcal{F}_m$ and also to $\mathcal{F}_n$, the probability $\mathbf{P}_0(A)$ is defined in terms of the model that lets all n-toss sequences be equally likely, but also in terms of the model that lets all m-toss sequences be equally likely. Thus, we have to check that these alternative definitions are consistent, tht is, lead to the same value for $\mathbf{P}_0(A)$.] It can be checked (though it is not trivial) that $\mathbf{P}_0$ is countably additive on $\mathcal{F}_0$.

Then, the extension theorem implies that there is a well-defined probability space that agrees with the elementary coin tossing probabilities $\mathbf{P}_0(A)$ for events A that relate to a finite number of tosses. It confirms that the intuitive process of an infinite sequence of coin flips can be captured rigorously within the framework of probability theory.

7 The uniform distribution on the unit interval

As with the coin tossing model, we now wish to define a uniform probability distribution on the unit interval. For simplicity of exposition, we consider the interval $(0, 1]$ (that is, the left endpoint is missing). The uniform probability distribution is meant to correspond to $\mathbf{P}(A) = \text{length}(A)$. For this, we need a family of sets which is a σ-field, and for which "length" can be well-defined. Length is of course well-defined for nice and simple sets like $(1/4, 1/2]$ (its length is $1/4$), but what about more complicated sets?

Let $\mathcal{F}_0$ consist of all sets that are unions of finitely many intervals of the form $(a, b]$. Given a set of the form $A = (a_1, b_1] \cup \cdots \cup (a_n, b_n]$, where $0 \le a_1 < b_1 \le a_2 < b_2 \le \cdots \le a_n < b_n \le 1$ (which is the typical set in $\mathcal{F}_0$), we define $\mathbf{P}_0(A) = (b_1 - a_1) + \cdots + (b_n - a_n)$ which is its total length.

The family $\mathcal{F}_0$ is a field. Indeed, for the set A above, its complement (in

$(0,1])$ is $(b_1, a_2] \cup \cdots \cup (b_n, 1]$ which is also in $\mathcal{F}_0$. And the union of two sets that are unions of finitely many intervals is also a union of finitely many intervals. (For example, if $A = (1/8, 2/8] \cup (4/8, 7/8]$ and $B = (3/8, 5/8]$, then $A \cup B = (1/8, 2/8] \cup (3/8, 7/8]$.)

It turns out that $\mathbf{P}_0$ is countably additive on $\mathcal{F}_0$. This essentially boils down to checking the following. If $(a, b] = \cup_{i=1}^{\infty}(a_i, b_i]$, where the intervals $(a_i, b_i]$ are disjoint, then $b - a = \sum_{i=1}^{\infty}(b_i - a_i)$. This may appear intuitively obvious, but a formal proof is nontrivial.

Let $\mathcal{F} = \sigma(\mathcal{F}_0)$, which is the smallest σ-field containing $\mathcal{F}_0$. Sets in $\mathcal{F}$ are called **Borel sets** and $\mathcal{F}$ is called the **Borel σ-field**. Using the extension theorem and the earlier assertions, it follows that there is a probability measure $\mathbf{P}$ which agrees with $\mathbf{P}_0$ for sets A that are in $\mathcal{F}_0$ (union of finitely many disjoint intervals). This probability measure is called the **Lebesgue measure**.

By augmenting the sample space Ω to include 0, and assigning zero probability to it, we obtain a new probability model with sample space $\Omega = [0, 1]$. (Exercise: define formally the sigma-field on $[0, 1]$, starting from the σ-field on $(0, 1]$).

Any set that can be formed by starting with intervals $[a, b)$ using a countable number of set-theoretic operations (taking complements of previously formed sets, or forming countable unions of previously formed sets) is a Borel set. For example, single-element sets $\{a\}$ are Borel sets, and so is the set of rational numbers in $[0, 1)$. Furthermore, intervals $[a, b]$ are also Borel since they are of the form $[a, b) \cup \{b\}$, which is the union of two Borel sets. More generally, any set you may be able to construct will be a Borel set, as long as your construction is not "exotic."

Having defined Lebesgue measure ("length") for Borel subsets of [0,1], it is straightforward to define the Lebesgue measure μ_n for Borel subsets of $[n, n+1]$. This leads to a construction of Lebesgue measure on Borel subsets of the real line. (To define the Borel subsets, we consider the σ-field generated by sets of the form $(a, b]$. We then define $\overline{\mu}$, by letting

$$\overline{\mu}(A) = \sum_{n=-\infty}^{\infty} \mu\Big(A \cap (n, n+1]\Big), \qquad A \in \mathcal{F},$$

and verify that $\overline{\mu}$ is indeed a measure (need to use the countable additivity of μ_n to establish the countable additivity of $\overline{\mu}$).

8 Completion of a probability space

Consider a probability space $(\Omega, \mathcal{F}, \mathbf{P})$. Suppose that $A \subseteq B$, $B \in \mathcal{F}$, $\mathbf{P}(B) = 0$. (Any set B with this property is called a **null** set. Note that in this context, "null" is not the same as "empty.") If the set A is not in $\mathcal{F}$, it does not have a probability assigned to it, but it would seem that one can safely assign zero probability to it.

The first step is to augment our σ-field so that it contains all the subsets of null sets. This is done by letting $\mathcal{F}^* = \sigma(\mathcal{F} \cup \mathcal{N})$, where $\mathcal{N}$ is the collection of all subsets of null sets, and then extending **P** in a natural manner to obtain a new probability measure $\mathbf{P}^*$ that applies to all sets in $\mathcal{F}^*$. [This is also discussed briefly in pp. 14-15 of [GS].] The resulting probability space is said to be **complete**. It has the property that all subsets of null sets are included in the σ-field and are also null sets.

When this completion is carried out for the model of the preceding section ($\Omega = [0, 1)$, $\mathcal{F}$=Borel sets, **P**=Lebesgue measure), we obtain a new probability space with the same Ω, an augmented σ-field $\mathcal{F}^*$, and a measure on $(\Omega, \mathcal{F}^*)$. The sets in $\mathcal{F}^*$ are called **Lebesgue measurable** sets; the new measure still has the same name ("Lebesgue measure").

9 Interesting facts

(a) There exist sets that are Lebesgue measurable but not Borel measurable, so in this context $\mathcal{F}$ is a proper subset of $\mathcal{F}^*$.

(b) There are as many Borel measurable sets as there are points on the real line (this is the "cardinality of the continuum"), but there are as many Lebesgue measurable sets as there are subsets of the real line (which is a higer cardinality).

(c) There exist subsets of $(0, 1]$ that are not Lebesgue measurable. [W, pp. 14-15, 192].

(d) It is not possible to construct a probability space in which the σ-field includes all subsets of $(0, 1]$, with the property that $\mathbf{P}(\{x\}) = 0$ for every $x \in (0, 1]$. [B, pp. 45-46]

10 References

Essentially all of this discussion (plus proofs and much more) is in [B]. Reference [W] (Chapters 1 and 2) is more concise and closer to our development. It includes most of the proofs (hard ones are in appendices). But the countable additivity assertion for $\mathbf{P}_0$ in Section 5 of this handout is not in [W].

[B] Billingsley, *Probability and Measure.*
[W] Williams, *Probability with Martingales.*