Math 525: Lecture 2

January 9, 2018

For the entirety of this lecture, we will always assume an underlying probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Recall that $\mathcal{F}$ is a σ-algebra on Ω and $\mathbb{P}\colon \mathcal{F} \to [0,1]$ is a probability measure.

1 Conditional probability

Definition 1.1. We say $A \in \mathcal{F}$ is a *null event* (or simply *null*) if $\mathbb{P}(A) = 0$.

Given $A, C \in \mathcal{F}$ such that C is not null, how do we define the "probability of A given C"? Let's do a thought experiment. If A and C are two (possibly overlapping) regions of a dartboard, we may throw N darts at a dart board and count $n(A \cap C)$, the number of darts which land in $A \cap C$ and $n(C)$, the number of darts which land in C. Now, let $\mathbb{P}(A \mid C)$ denote the probability that a dart lands in region A given that it lands in region C and $\mathbb{P}(X)$ denote the probability that it lands in region X. Then, at least intuitively,

$$\mathbb{P}(A \mid C) = \lim_{N\to\infty} \frac{n(A \cap C)}{n(C)} = \lim_{N\to\infty} \frac{\frac{n(A\cap C)}{N}}{\frac{n(C)}{N}} = \frac{\mathbb{P}(A \cap C)}{\mathbb{P}(C)}.$$

This suggest the following definition of condition probability.

Definition 1.2. Let $A, C \in \mathcal{F}$ such that C is not null. The *conditional probability of A given C* is defined as

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

Equivalently, the conditional probability $\mathbb{P}(A \mid C)$ is the unique number satisfying

$$\mathbb{P}(A \mid C)\mathbb{P}(C) = \mathbb{P}(A \cap C). \tag{1}$$

Proposition 1.3 (Bayes' Rule)**.** *Let $B_1, \ldots, B_n \in \mathcal{F}$ be a partition of Ω such that each B_i is not null. Then, for any non-null $A \in \mathcal{F}$,*

$$\mathbb{P}(B_j \mid A) = \frac{\mathbb{P}(A \mid B_j)\mathbb{P}(B_j)}{\sum_i \mathbb{P}(A \mid B_i)\mathbb{P}(B_i)} = \left(1 + \frac{\sum_{i\neq j} \mathbb{P}(A \mid B_i)\mathbb{P}(B_i)}{\mathbb{P}(A \mid B_j)\mathbb{P}(B_j)}\right)^{-1}.$$

Proof. The proof is just two applications of (1):

$$\mathbb{P}(B_j \mid A) = \frac{\mathbb{P}(B_j \cap A)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A \mid B_j)\mathbb{P}(B_j)}{\sum_i \mathbb{P}(A \cap B_j)} = \frac{\mathbb{P}(A \mid B_j)\mathbb{P}(B_j)}{\sum_i \mathbb{P}(A \mid B_i)\mathbb{P}(B_i)}.$$

□

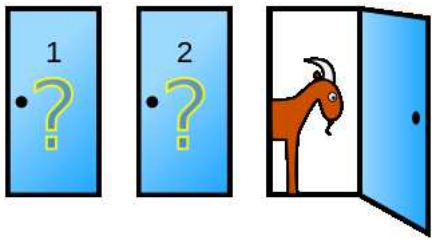

Figure 1: Monty Hall Problem

A famous example in which conditional probability produces unexpected results is the Monty Hall Problem:

Example 1.4 (Monty Hall Problem)**.** Suppose you're on a game show, and you're given the choice of three doors: Behind one door is a car; behind the others, goats. You pick a door, say No. 1, and the host, who knows what's behind the doors, opens **another** door, say No. 3, which has a goat. He then says to you, "Do you want to pick door No. 2?" Is it to your advantage to switch your choice?

There are some things to keep in mind before you make your choice:

- The car is just as likely to be behind door No. 1 as it is to be behind any other door (as such, we may as well always assume you initially choose door No. 1).
- Once you have made your initial choice, the host will not choose to open your door. Moreover, the host will not open the door with the car behind it.

Let C_i be the event that the car is behind door No. i. We have $\mathbb{P}(C_i) = 1/3$. Let H be the event that the host chooses door No. 3. We have,

$$\begin{aligned}\mathbb{P}(H \mid C_1) &= 1/2\\ \mathbb{P}(H \mid C_2) &= 1\\ \mathbb{P}(H \mid C_3) &= 0.\end{aligned}$$

By Bayes' rule,

$$\mathbb{P}(C_2 \mid H) = \left(1 + \frac{\mathbb{P}(H \mid C_1)\mathbb{P}(C_1) + \mathbb{P}(H \mid C_3)\mathbb{P}(C_3)}{\mathbb{P}(H \mid C_2)\mathbb{P}(C_2)}\right)^{-1} = \frac{1}{1 + 1/2} = \frac{2}{3}.$$

It is in your best interests to **switch** doors!

2 Independence

Definition 2.1. Let $A, B \in \mathcal{F}$. We say A and B are *independent* if

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

The definition of independence is motivated by the following observation:

Proposition 2.2. *Let $A, B \in \mathcal{F}$ with B not null. Then, A and B are independent if and only if $\mathbb{P}(A \mid B) = \mathbb{P}(A)$.*

Proof. Suppose A and B are independent. Then,

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

Suppose now that $\mathbb{P}(A \mid B) = \mathbb{P}(A)$. Then,

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

Note that if at least one of A and B are null, then A and B are trivially independent. For example, if B is null, we have

$$\mathbb{P}(A \cap B) \leq \mathbb{P}(B) = 0.$$

Since $\mathbb{P}(\cdot) \geq 0$, it follows that $\mathbb{P}(A \cap B) = 0$.

Corollary 2.3. *Let $A, B \in \mathcal{F}$ be non-null. Then,*

$$\mathbb{P}(A \mid B) = \mathbb{P}(A) \iff \mathbb{P}(B \mid A) = \mathbb{P}(B)$$

Proof. If $\mathbb{P}(A \mid B) = \mathbb{P}(A)$, then A and B are independent. Reversing the roles of A and B in Proposition 2.2, we obtain $\mathbb{P}(B \mid A) = \mathbb{P}(B)$. $\square$

3 Counting

In this section, we review some basic facts about counting that you may have previously encountered in a combinatorics class.

Definition 3.1. A *permutation* of a (possibly finite) sequence of elements is simply a reordering of that sequence.

Example 3.2. The sequence (a, b, c) has 6 permutations: (a, b, c), (a, c, b), (b, a, c), (b, c, a), (c, a, b), (c, b, a).

Given a sequence $(1, \ldots, n)$ with n elements, we want to determine the number of permutations $(i_1, \ldots, i_n)$. We argue as follows: to pick the first item i_1 in the permutation, we have n items available to us. Once we have picked the first item, we have $n - 1$ options for the second item i_2, and so forth. Therefore, there are

$$n\,(n-1)\,(n-2)\cdots 1 \equiv \boxed{n!}$$

permutations of a sequence with n elements.

Exercise 3.3. We have an urn with n **distinct** balls. We first pull out a ball and set it down. We next pull out another ball and place it to the right of the first ball, and continue in this way until we have $r \leq n$ balls in a row:

$$b_1 \quad b_2 \quad \cdots \quad b_r$$

There are exactly

$$n\,(n-1)\,(n-2)\cdots(n-r+1) = \boxed{\frac{n!}{(n-r)!}}$$

ways we can do this.

The above is referred to as sampling **without replacement**, since we do not return a ball to the urn after having drawn it. If we are sampling **with replacement**, we would have $\boxed{n^r}$ possibilities. What if we sample **without replacement, but without regards to order**? That is, how many ways are there to *choose* r objects out of n distinct ones? Since there are $r!$ permutations of a sequence of r objects, this is simply

$$\frac{\#\text{ of ways to sample } r \text{ objects from } n \text{ w/o replacement}}{\#\text{ of ways to permute } r \text{ objects}} = \frac{1}{r!}\frac{n!}{(n-r)!} \equiv \boxed{\binom{n}{r}}.$$

4 Linear homogeneous recurrence relations (optional)

This section reviews linear homogeneous recurrence relations, which are used in the example in the next section involving the gambler's ruin. This material is mostly for your interest, and won't be tested.

Consider the equation

$$a_n + c_1 a_{n-1} + \cdots + c_d a_{n-d} = 0 \tag{2}$$

where $c_1, \ldots, c_d$ are (real or complex) constants. This is called a *linear homogeneous recurrence relation*. A solution of this equation is a sequence $(a_n)_{n\geq 1}$ that satisfies it. How do we find the solutions?

Let's proceed by guessing that a solution is of the form

$$a_n = r^n$$

where $r \neq 0$. Indeed, if this is the case, (2) suggests that

$$r^n + c_1 r^{n-1} + \cdots + c_d r^{n-d} = 0.$$

Multiplying both sides of the above by r^{d-n},

$$r^d + c_1 r^{d-1} + \cdots + c_d r^0 = 0 \tag{3}$$

(note that $r^0 = 1$). The quantity on the left hand side of the equation (3) is called the *characteristic polynomial* associated with the recurrence.

By the fundamental theorem of algebra, the characteristic polynomial has d roots, which we label $r_1, \ldots, r_d$. By the argument in the previous paragraph, for any $1 \leq i \leq d$, defining the sequence $(a_n)_{n \geq 1}$ by

$$a_n = r_i^n$$

yields a solution of the recurrence.

In fact, we can do even better than that. If we assume that the roots are distinct (i.e., $r_i \neq r_j$ whenever $i \neq j$), then the sequence $(a_n)_{n \geq 1}$ defined by

$$a_n = C_1 r_1^n + \cdots + C_d r_d^n$$

is also a solution of the recurrence, where $C_1, \ldots, C_d$ are arbitrary (real or complex) constants.

5 Gambler's ruin (optional)

Consider a gambler who repeatedly plays a game against an opponent in which they receive a dollar with probability p and lose a dollar with probability $1 - p$. Both the gambler and their opponent start off with initial stakes of n and m dollars, respectively. The game ends when either the gambler or the opponent are broke.

Let $N = n + m$ be the total amount of money in the game. Let

$$f(k) = \mathbb{P}(\text{Gambler goes broke if initial stake is } k).$$

If the gambler has no money left, the game ends with the gambler broke. Therefore, $f(0) = 1$. The game also ends if the gambler has all the money in the game. Therefore, $f(N) = 0$. Moreover,

$$f(k) = pf(k+1) - (1-p)\, f(k-1) \qquad \text{if } 0 < k < N.$$

We rewrite the above as

$$\boxed{f(k+1) - \frac{1}{p} f(k) + \left(\frac{1}{p} - 1\right) f(k-1) = 0 \qquad \text{if } 0 < k < N}.$$

Ignoring the quantifier "if $0 < k < N$", the above becomes

$$f(k+1) - \frac{1}{p} f(k) + \left(\frac{1}{p} - 1\right) f(k-1) = 0.$$

This is nothing other than a linear homogeneous recurrence relation! Its characteristic polynomial is

$$r^2 - \frac{1}{p} r + \left(\frac{1}{p} - 1\right) r,$$

which has roots 1 and $a = \frac{1}{p} - 1$. Assuming $p \neq \frac{1}{2}$, the roots are distinct, and we can use the strategy of the previous section to conclude that a solution of this recurrence is

$$f(k) = A + Ba^k \tag{4}$$

where A and B are arbitrary real constants. To determine the specific value of the constants A and B, we employ the boundary conditions. Namely,

$$f(0) = 1 = A + B \qquad \text{and} \qquad f(N) = 0 = A + Ba^N.$$

If we solve the above equations for A and B, we get

$$A = -\frac{a^N}{1 - a^N} \qquad \text{and} \qquad B = \frac{1}{1 - a^N}.$$

Plugging the above back into (4), we get

$$\boxed{f(k) = \frac{a^k - a^N}{1 - a^N} \qquad \text{if } 0 \leq k \leq N}.$$

Recall that our calculations were only accurate in the case of $p \neq \frac{1}{2}$. If $p = \frac{1}{2}$ (equivalently, $a = 1$), we can take limits to obtain the answer. By L'Hopital's rule,

$$\lim_{a \to 1} f(k) = \lim_{a \to 1} \frac{a^n - a^N}{1 - a^N} = \lim_{a \to 1} a^n \frac{1 - a^{N-n}}{1 - a^N} = \lim_{a \to 1} \frac{1 - a^{N-n}}{1 - a^N} = \frac{N - n}{N}.$$

6 Generating a σ-algebra and Borel sets

Often times, we only have a small set of events $\mathcal{G}$ that do not necessarily form a σ-algebra. Since the probabilistic framework requires a σ-algebra, we need a procedure to "generate" a σ-algebra from $\mathcal{G}$.

Definition 6.1. Let $\mathcal{G} \subset 2^\Omega$ be a set. Then, the *σ-algebra generated from $\mathcal{G}$* is

$$\sigma(\mathcal{G}) = \bigcap_{\substack{\mathcal{F} \text{ is a } \sigma\text{-algebra on } \Omega \\ \mathcal{G} \subset \mathcal{F}}} \mathcal{F}.$$

In the first assignment, you are asked to prove the following:

Proposition 6.2. *$\sigma(\mathcal{G})$ is the "smallest" σ-algebra on Ω which contains $\mathcal{G}$. That is, for any σ-algebra $\mathcal{F}$ satisfying $\mathcal{G} \subset \mathcal{F}$, it follows that $\sigma(\mathcal{G}) \subset \mathcal{F}$.*

In the previous lecture, we mentioned that in the case of a countable sample space Ω, we could simply take the σ-algebra to be the powerset (i.e., $\mathcal{F} = 2^\Omega$). In this case, any probability space has the form $(\Omega, 2^\Omega, \mathbb{P})$. Recall that in the case of an uncountable sample space Ω, this approach fails. In this case, however, we can often take the σ-algebra $\mathcal{F}$ to be the something called the Borel σ-algebra. Let's first discuss what the Borel σ-algebra is in the case of $\Omega = \mathbb{R}$, generalizing later to metric and topological spaces.

Definition 6.3. Let

$$\mathcal{G} = \{(-\infty, x] \colon x \in \mathbb{R}\}$$

be the set of all intervals of the form $(-\infty, x]$. Note that $\mathcal{G} \subset 2^\mathbb{R}$. Let $\mathcal{B}(\mathbb{R}) = \sigma(\mathcal{G})$. We call $\mathcal{B}(\mathbb{R})$ the *Borel σ-algebra* on $\mathbb{R}$, and refer to any element of $\mathcal{B}(\mathbb{R})$ as a Borel set.

Example 6.4. Any open interval is a Borel set. To see this, let (a, b) be an open interval with $a < b$. First, note that

$$\bigcup_{n\geq 1}(-\infty, b - 1/n] = (-\infty, b).$$

Therefore,

$$(-\infty, a]^c \cap (-\infty, b) = (a, b)$$

and hence $(a, b) \in \mathcal{B}(\mathbb{R})$. A similar argument yields that any closed interval $[a, b]$ with $a < b$ is also a Borel set.

Example 6.5. Any open set is a Borel set. This is a direct consequence of the fact that any open set $G \subset \mathbb{R}$ is a countable union of open intervals $G = \cup_{n\geq 1}(a_n, b_n)$. Moreover, since a closed set F is a complement of an open set $F = G^c$, any closed set is also a Borel set (remember, the Borel σ-algebra is closed under complements).

7 Generalizing the Borel sets (optional)

Example 6.5 implies that we could have also defined the Borel sets on $\mathbb{R}$ by

$$\mathcal{B}(\mathbb{R}) = \sigma(\{G \subset \mathbb{R} \colon G \text{ is open}\}).$$

This suggests that we can generalize Borel sets to metric spaces (or even more generally, topological spaces):

Definition 7.1 (Borel sets on metric spaces)**.** Let (X, d) be a metric space. Then,

$$\mathcal{B}(X) = \sigma(\{G \subset X \colon G \text{ is open with respect to } d\})$$

is the set of Borel sets on X.

Definition 7.2 (Borel sets on topological spaces)**.** Let (X, τ) be a topological space. Then,

$$\mathcal{B}(X) = \sigma(\{G \subset X \colon G \in \tau\})$$

is the set of Borel sets on X.