MTH 224, Spring 2024

Instructor: Bahman Angoshtari

Lecture 6

Section 1.5: Bayes' rule
Section 2.1: random variables, probability mass functions, cumulative distribution functions

6.1. Law of total probability (continued)

EXAMPLE 6.1. There are 15 tennis balls in a box, nine of which are new and 6 are used. In the first round, three of the balls are randomly selected, played with, and then returned to the box. In the second round, another three balls are randomly selected from the box. Find the probability that all three balls selected for the second round are new.

SOLUTION. Let B = all 3 ball for round 2 are new. Let $A_i = i$ new balls are selected for round 1, with $i = 0, 1, 2, 3$. We have that

$$\mathbb{P}(A_0) := \frac{\binom{6}{3}}{\binom{15}{3}} = \frac{4}{91},\ \mathbb{P}(A_1) := \frac{\binom{9}{1}\binom{6}{2}}{\binom{15}{3}} = \frac{27}{91},$$
$$\mathbb{P}(A_2) := \frac{\binom{9}{2}\binom{6}{1}}{\binom{15}{3}} = \frac{216}{455},\ \mathbb{P}(A_3) := \frac{\binom{9}{3}}{\binom{15}{3}} = \frac{12}{65}$$

and

$$\mathbb{P}(B|A_0) := \frac{\binom{9}{3}}{\binom{15}{3}} = \frac{12}{65},\ \mathbb{P}(B|A_1) := \frac{\binom{8}{3}}{\binom{15}{3}} = \frac{8}{65},$$
$$\mathbb{P}(B|A_2) := \frac{\binom{7}{3}}{\binom{15}{3}} = \frac{1}{13},\ \mathbb{P}(B|A_3) := \frac{\binom{6}{3}}{\binom{15}{3}} = \frac{4}{91}.$$

Therefore, by LPT,

$$\begin{aligned}\mathbb{P}(B) &= \sum_{i=0}^{3} \mathbb{P}(A_i)\mathbb{P}(B|A_i)\\ &= \frac{4}{91} \times \frac{12}{65} + \frac{27}{91} \times \frac{8}{65} + \frac{216}{455} \times \frac{1}{13} + \frac{12}{65} \times \frac{4}{91}\\ &= \frac{528}{5915} \approx 0.0893.\end{aligned}$$

6.2. Bayes' rule

- Let A and B be two events. In general, $\mathbb{P}(A|B)$ and $\mathbb{P}(B|A)$ are different. The **Bayes' rule** helps us determine one of the conditional probabilities in terms of the other.

EXAMPLE 6.2. We know that 0.5% of the population in Miami have COVID-19. We have a PCR test that detects an infected individual 99% of times. It (falsely) produce a positive result for an uninfected individual 1% of times. If we randomly test an individual in Miami and the test is positive, what is the probability that he is infected?

SOLUTION. Define the events A = infected; and B = test is positive. We know that $\mathbb{P}(A) = 0.005, \mathbb{P}(B|A) = 0.99$, and $\mathbb{P}(B|A^c) = 0.01$. We want to know $\mathbb{P}(A|B)$.

Note that

$$\begin{aligned}\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} &= \frac{\mathbb{P}(A)\mathbb{P}(B|A)}{\mathbb{P}(A)\mathbb{P}(B|A) + \mathbb{P}(A^c)\mathbb{P}(B|A^c)} \\ &= \frac{0.005 \times 0.99}{0.005 \times 0.99 + 0.995 \times 0.01} = 0.332.\end{aligned}$$

So, there is only 33% chance that he is infected. This may seem a bit counterintuitive!

Intuition: The disease is rare so most positive tests are from uninfected individuals. For example, among 1000 people, about 5 are infected and 995 are not. Even if all the 5 infected test positive, about 10 of the uninfected will also (falsely) test positive. So the probability is around $\frac{5}{15} = \frac{1}{3}$.

Bayes rule for two events: Let A and B be two events such that $\mathbb{P}(B) \neq 0$, $\mathbb{P}(A) \neq 0$ and $\mathbb{P}(A) \neq 1$. Then,

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

Bayes' rule for multiple events: Let $A_1, A_2, ..., A_N$ be **partition** of the sample space. That is,

(1) They are disjoint, that is, $A_i \cap A_j = \emptyset$ for all $i \neq j$.
(2) $A_1 \cup A_2 \cup ... \cup A_N$ = sample space.

Furthermore, assume that $\mathbb{P}(A_i) \neq 0$ and that $\mathbb{P}(B) \neq 0$. We then have

$$\mathbb{P}(A_1|B) = \frac{\mathbb{P}(A_1)\mathbb{P}(B|A_1)}{\mathbb{P}(B)} = \frac{\mathbb{P}(A_1)\mathbb{P}(B|A_1)}{\mathbb{P}(A_1)\mathbb{P}(B|A_1) + \ldots + \mathbb{P}(A_N)\mathbb{P}(B|A_N)}.$$

6.3. Random variables

- Intuitively, **random variables** are numerical summaries of an experiment. Think of them as labeling the sample space with numbers.

EXAMPLE 6.3. We toss a coin 3 times. Let the sample space be

$$S = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}.$$

Consider the random variable X = the number of heads. In particular:

$$\begin{aligned}&X(HHH) = 3, \\ &X(HHT) = X(HTH) = X(THH) = 2, \\ &X(HTT) = X(THT) = X(TTH) = 1, \\ &X(TTT) = 0.\end{aligned}$$

We can then write various events in terms of X:

$$"X \leq 1" = \{\omega \in S : X(\omega) \leq 1\} = \{TTT, TTH, THT, HTT\}.$$

DEFINITION 6.4. A random variable is a function X that assigns the number $X(\omega)$ to the outcome ω in the sample space. We can define events in terms of the random variable. For example, the event $X \leq 1$

is actually the event

$$"X \leq 1" = \Big\{\omega \in S : X(\omega) \leq 1\Big\}.$$

We usually drop the function arguments of random variables, that is, instead of $X(\omega)$,we simply use X.

6.4. Distribution of discrete random variables

- If a random variable takes only a finite number of values, or a countably infinite number of values, then we call it a **discrete random variable**. So, the range of a discrete random variable X is either of the form $\{x_1, x_2, \ldots, x_n\}$or $\{x_1, x_2, \ldots\}$.
- Can you name a random variable that is not discrete?
- In general, we say that we know the **distribution** of a random variable X if we can determine $\mathbb{P}(X \in B)$ for **"any"** subset B of real numbers. For a discrete random variable, one way to specify its distribution is by using the **probability mass function** (pmf).

Definition 6.5. Then, the probability mass function (pmf) of a discrete random variable X, denoted by $p_X(x)$, is given by $p_X(x) = \mathbb{P}(X = x)$ for all $x \in \mathbb{R}$.
The **support** of a discrete random variable X is the set of all points $x \in \mathbb{R}$ such that $p_X(x) \neq 0$. That is:

$$\mathrm{supp}(X) = \{x \in \mathbb{R} : \mathbb{P}(X = x) \neq 0\}.$$

- Another common way of specifying the distribution of a random variable is through its **cumulative distribution function (cdf)**. The advantage of using cdf is that it works for all types of random variables, even for continuous random variables that we will encounter later.

Definition 6.6. Let X be a random variable. The cumulative distribution function of X, denoted by $F_X(x)$, is given by $F_X(x) = \mathbb{P}(X \leq x); \quad x \in \mathbb{R}$.

Example 6.7. Consider the example of tossing a coin 3 times and letting $X =$ the number of heads. The pmf of X is then given by

$$\begin{aligned}
p_X(0) &= \mathbb{P}(X = 0) = \mathbb{P}(\{TTT\}) = \frac{1}{8}, \\
p_X(1) &= \mathbb{P}(X = 1) = \mathbb{P}(\{HTT, THT, TTH\}) = \frac{3}{8}, \\
p_X(2) &= \mathbb{P}(X = 2) = \mathbb{P}(\{HHT, HTH, THH\}) = \frac{3}{8}, \\
p_X(3) &= \mathbb{P}(X = 3) = \mathbb{P}(\{HHH\}) = \frac{1}{8}.
\end{aligned}$$

We have $\mathrm{supp}(X) = \{0, 1, 2, 3\}$. We define $p_X(x) = 0$ for all $x \notin \mathrm{supp}(X)$.