ECE 361 Probability for Engineers (Fall, 2016)
Lecture 2b

Sample problems that use strategies similar to those required in HW 2

Problem. Fix events A, B in a sample space Ω. Suppose you are given three probabilities: $p_A = \mathbb{P}(A)$, $p_B = \mathbb{P}(B)$, and $p_o = \mathbb{P}(A \cap B | A \cup B)$. Find an expression for $\mathbb{P}(A|B)$ in terms of these three numbers.

Solution. By Bayes' rule:

$$1 = \mathbb{P}(A \cup B | A \cap B) = \frac{\mathbb{P}(A \cap B | A \cup B)\mathbb{P}(A \cup B)}{\mathbb{P}(A \cap B)} \Rightarrow \mathbb{P}(A \cap B) = p_o \mathbb{P}(A \cup B), \tag{1}$$

and thus

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B) = p_A + p_B - p_o \mathbb{P}(A \cup B), \tag{2}$$

which means

$$\mathbb{P}(A \cup B) = \frac{p_A + p_B}{1 + p_o}. \tag{3}$$

Then:

$$\mathbb{P}(A \cap B) = \frac{(p_A + p_B)p_o}{1 + p_o} \tag{4}$$

and so

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{(p_A + p_B)p_o}{p_B(1 + p_o)}. \tag{5}$$

Problem. There are three coins in a box. One is a two-headed coin, another is a fair coin, and the third is a biased coin that comes up heads with probability p. When one of the three coins is selected at random and flipped, it shows heads. What is the probability that it was the two-headed coin?

Solution. Let T, F, B be the event that we selected the two-headed coin, the fair coin, and the biased coin, respectively. Then:

$$\mathbb{P}(T|H) = \frac{\mathbb{P}(H|T)\mathbb{P}(T)}{\mathbb{P}(H|T)\mathbb{P}(T) + \mathbb{P}(H|F)\mathbb{P}(F) + \mathbb{P}(H|B)\mathbb{P}(B)} = \frac{1 \cdot \frac{1}{3}}{1 \cdot \frac{1}{3} + \frac{1}{2} \cdot \frac{1}{3} + p \cdot \frac{1}{3}} = \frac{1}{\frac{3}{2} + p} = \frac{2}{3 + 2p}. \tag{6}$$

Problem. There are two urns. Urn 1 has w_1 white and b_1 black balls, while urn 2 has w_2 white and b_2 black balls. We flip a fair coin. If the outcome is heads, then a ball from urn 1 is selected, while if the outcome is tails, then a ball from urn 2 is selected. Suppose that a white ball is selected. What is the probability that the coin landed tails?

Solution. Let H, T denote the events that the coin flip is heads and tails, respectively. Let W, B denote the events that a white and black ball is selected. Then

$$\mathbb{P}(T|W) = \frac{\mathbb{P}(W|T)\mathbb{P}(T)}{\mathbb{P}(W|T)\mathbb{P}(T) + \mathbb{P}(W|H)\mathbb{P}(H)} = \frac{\frac{w_2}{w_2+b_2} \cdot \frac{1}{2}}{\frac{w_2}{w_2+b_2} \cdot \frac{1}{2} + \frac{w_1}{w_1+b_1} \cdot \frac{1}{2}} = \frac{w_2(w_1 + b_1)}{w_2(w_1 + b_1) + w_1(w_2 + b_2)}. \tag{7}$$

§1.6 Counting

The counting principle

The counting principle gives the number of outcomes for a random experiment with r **independent** stages.

> **The counting principle.** In a random experiment of r independent stages where there are n_i options at stage $i \in [r]$, there are a total of $|\Omega| = n_1 \times n_2 \times \cdots \times n_r$ outcomes.

Example. $|\mathcal{P}(\Omega)| = 2^{|\Omega|}$. If $\Omega = \{\omega_1, \ldots, \omega_r\}$ then think of an r-stage experiment where at each stage we decide whether to include ω_i. There are two possible outcomes at each stage.

k-permutations

Given a set of n distinct objects S ($|S| = n$), we differentiate between ordered and unordered sets. When order matters we consider permutations, when order doesn't matter we consider combinations. A k-permutation of an n-set is an ordered subset $R = (s_1, \ldots, s_k)$; recall braces denote an unordered set and parentheses denote an ordered set. The number of possible permutations is

$$(n-0) \times (n-1) \times (n-2) \times \cdots \times (n-k+1). \tag{8}$$

There are $n-0$ choices for position 1 in R, then there are $n-1$ choices for position 2, and finally $n-(k-1) = n-k+1$ choices for position k. Note we can write this as

$$(n-0) \times (n-1) \times (n-2) \times \cdots \times (n-k+1) = \frac{n(n-1)(n-2)\cdots(n-k+1)(n-k)\cdots 2 \cdot 1}{(n-k)\cdots 2 \cdot 1}. \tag{9}$$

The number of permutations. The number of permutations (ordered subsets) of size k from a set of n distinct objects is

$$\frac{n!}{(n-k)!} \tag{10}$$

Note the number of permutations of n objects is then $n! = n(n-1)\cdots 2 \cdot 1$.

Example. The number of words consisting of 4 distinct letters is $26 \times 25 \times 24 \times 23 = 358,800$.

Combinations

Combinations are unordered subsets. The question of interest is the number of combinations of size k from a set of n distinct objects. For example, given the set $\{A, B, C, D\}$ the set of 2-permutations is

$$AB, BA, AC, CA, AD, DA, BC, CB, BD, DB, CD, DC, \tag{11}$$

while the combinations are

$$AB, AC, AD, BC, BD, CD. \tag{12}$$

In fact for any given combination of size k there are $k!$ permutations that correspond to the possible orderings of the elements of the combination. Given that there are $n!/(n-k)!$ permutations, it follows there are $\frac{n!}{k!(n-k)!}$ combinations of the same set. This explains the definition.

The binomial coefficient. The number of combinations (unordered subsets) of size k from a set of n distinct objects is

$$\binom{n}{k} \equiv \frac{n!}{k!(n-k)!}. \tag{13}$$

Recall the binomial formula $(x+y)^n = \sum_{k=0}^n \binom{n}{k} x^k y^{n-k}$ was used to verify the binomial probabilities p_k summed to one using $x = p$ and $y = 1-p$. In fact setting $p = 1/2$ we see this gives

$$\sum_{k=0}^n \binom{n}{k} = 2^n. \tag{14}$$

Recall given a set of n distinct objects there are 2^n possible subsets. The number of subsets of size k is $\binom{n}{k}$. The above identity observes that if you add up the number of subsets of each cardinality $k \in \{0, \ldots, n\}$ then you have accounted for all possible subsets.

Partitions

A combination is a subset: it divides the original n set into two parts: the subset of size k and its complement of size $n-k$. A partition generalizes this concept to dividing an n set into r parts.

The multinomial coefficient. Given integers $n_1, \ldots, n_r$ such that $n_1 + \cdots + n_r = n$, the number of ways of dividing an n set into r parts of sizes $n_1, \ldots, n_r$ is given by the multinomial coefficient:

$$\binom{n}{n_1, n_2, \ldots, n_r} \equiv \frac{n!}{n_1! n_2! \cdots n_r!}. \quad (15)$$

Example: anagrams. How many different words can be obtained by rearranging the letters in the word TATTOO? There are six positions so 6! permutations, but there are 2 O's and 3 T's, so any reordering of the Os or the Ts leaves the word unchanged, hence $6!/(1!2!3!) = 60$.

§2.1 Basic concepts

Random variables. A random variable is a function mapping each outcome of a random experiment to a real number.

Formally, if $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability triple for a random experiment, then random variable $X : \Omega \to \mathbb{R}$ is a function with the interpretation that $X(\omega) \in \mathbb{R}$ is the "value" assigned to outcome ω for each $\omega \in \Omega$. See Fig. 2.1 in text. We will use RV to denote "random variable" throughout the course.

Examples.

1. Experiment: sequence of five tosses of a coin. RV: number of heads in the sequence.
2. Experiment: two rolls of a die. RV: sum of the two rolls, number of sixes, second roll raised to the fifth power.
3. Experiment: transmission of a message. RV: time needed to transmit, number of symbols received in error, messsage delay.

The main concepts to be clarified in this chapter:

Main concepts related to RVs.

- An RV is a real-valued function of the outcome of the experiment.
- A function of an RV defines another RV.
- We can associate with each RV certain "averages" of interest, such as the mean and variance.
- An RV can be conditioned on an event or another RV.
- There is a notion of independence of an RV from an event or from another RV.

In general RVs may take on a finite, countably infinite, or uncountably infinite set of values. Simple examples of these include – the face of a rolled die, the number of coin flips until a head, or a real number chosen uniformly at random from $[0, 1]$, respectively. This chapter we focus on **discrete** RVs where the set of values is either discrete or countably infinite. The following concepts hold for discrete RVs.

Concepts related to discrete RVs.

- A discrete RV is a real-valued function of the outcome of the experiment that can take a finite or countably infinite number of values.
- A discrete RV has an associated probability mass function (PMF) which gives the probability of each numerical value that the RV can take.
- A function of a discrete RV defines another discrete RV, whose PMF can be obtained from the PMF of the original RV.

§2.2 Probability mass functions (PMF)

Let X be a discrete RV. Let $\mathcal{X}$ be the set of values that the RV takes. We call $\mathcal{X}$ the support. Formally: $\mathcal{X} = \{x \in \mathbb{R} : \exists \omega \in \Omega : X(\omega) = x\}$ – the set of values $x \in \mathbb{R}$ for which there exists at least one outcome $\omega \in \Omega$ where $X(\omega) = x$.

The PMF is defined on the support. It is a probability vector, denoted $\mathbf{p}$, with elements $p(x)$ for each $x \in \mathcal{X}$. More succinctly, the vector $\mathbf{p} = (p(x), x \in \mathcal{X})$ is a valid PMF if

$$p(x) \geq 0, \ \forall x \in \mathcal{X}, \ \text{ and } \sum_{x \in \mathcal{X}} p(x) = 1. \tag{16}$$

We say a vector $\mathbf{p}$ satisfying the above is a probability vector on $\mathcal{X}$.

Example. Toss a fair coin twice and let X be the number of heads. X has PMF

$$p(0) = 1/4, \ p(1) = 1/2, \ p(2) = 1/4. \tag{17}$$

Note that $p(x) = \mathbb{P}(\{X = x\})$ is the probability of the event $X = x$, and this in turn is the probability of all outcomes that map to x, i.e., $p(x) = \mathbb{P}(\{\omega : X(\omega) = x\})$. In other words, the RV X viewed as a function partitions Ω into events E_x, one for each $x \in \mathcal{X}$ where $E_x = \{\omega : X(\omega) = x\}$, and $p(x) = \mathbb{P}(E_x) = \sum_{\omega \in E_x} \mathbb{P}(\omega)$. As with all partitions, $E_x \cap E_y = \emptyset$ for $x \neq y$, and $\bigcup_{x \in \mathcal{X}} E_x = \Omega$. These ideas are summarized below.

Calculation of the PMF of an RV X. For each possible value of X:

- Collect all the possible outcomes that give rise to the event $\{X = x\}$.
- Add their probabilities to obtain $p(x)$.

The Bernoulli RV

A Bernoulli RV takes two values. In other words, the outcomes Ω are partitioned into two parts, with each part assigned a distinct value under the Bernoulli RV X. We often, although it is not necessary, assign values 1 and 0. For example, in the case of a coin flip we assign a value 1 to a head and a value 0 to a tail:

$$X = \begin{cases} 1, & \text{head} \\ 0, & \text{tail} \end{cases} \tag{18}$$

and the corresponding PMF is

$$p(1) = \mathbb{P}(\text{head}) = p, \quad p(0) = \mathbb{P}(\text{tail}) = 1 - p, \tag{19}$$

where $p \in [0, 1]$ is the fixed probability of a head (a biased coin). I will use the notation $X \sim \text{Ber}(p)$ to denote that X is a Bernoulli RV with bias p. Note we often speak of Bernoulli RVs as RVs for random coin flips, but in fact they model any dichotomous situation, e.g., success or failure. For example:

- The state of a telephone at a given time that can be either free or busy
- A person who can be either healthy or sick with a certain disease
- The preference of a person who can be either for or against a certain political candidate

The binomial RV

A coin (with probability of heads p) is tossed n times and each toss results in either heads or tails. Let X be the number of heads that result from the n tosses. We say X is a binomial RV with parameters n and p. I will use the notation $X \sim \text{Bin}(n, p)$ to denote that X is a binomial RV for n trials with success probability p. As discussed previously, the event $\{X = k\}$ is the union of $\binom{n}{k}$ distinct outcomes, i.e., there are $\binom{n}{k}$ distinct length n binary sequences with k ones. Each such sequence is equally likely with probability $p^k (1 - p)^{n-k}$, and as such

$$p(k) = \mathbb{P}(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}, \ k \in \{0, \ldots, n\}. \tag{20}$$

Again, as we discussed in the previous chapter, the binomial theorem assures us of the proper normalization:

$$\sum_{k=0}^{n} p(k) = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} = (p + (1-p))^n = 1. \tag{21}$$

The geometric RV

A coin (wth probability of heads $p \in (0,1)$) is tossed repeatedly until a head comes up. Here the sample space is the set of all binary sequences that end in a head, and the event $X = k$ corresponds to the outcome $\omega = (T, \ldots, T, H)$ where there are $k-1$ tails (Ts). I will use the notation $X \sim \mathrm{Geo}(p)$ to denote that X is a geometric RV with probability of success p. The PMF of X is clearly

$$p(k) = (1-p)^{k-1} p, \ k \in \mathbb{N}. \tag{22}$$

Let us check that this is a valid PMF:

$$\sum_{k \in \mathbb{N}} p(k) = \sum_{k \in \mathbb{N}} (1-p)^{k-1} p = p \sum_{k=0}^{\infty} (1-p)^k = p \frac{1}{1-(1-p)} = 1. \tag{23}$$

Here we have used the expression for the summation of a geometric series

$$\sum_{k=0}^{\infty} a^k = \frac{1}{1-a}, \ 0 < a < 1. \tag{24}$$

The Poisson RV

A Poisson RV X with parameter λ, denoted $X \sim \mathrm{Po}(\lambda)$, has PMF

$$p(k) = \mathrm{e}^{-\lambda} \frac{\lambda^k}{k!}, \ k = 0, 1, 2, \ldots \tag{25}$$

for $\lambda > 0$. This is a valid PMF since:

$$\sum_{k=0}^{\infty} p(k) = \sum_{k=0}^{\infty} \mathrm{e}^{-\lambda} \frac{\lambda^k}{k!} = \mathrm{e}^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = \mathrm{e}^{-\lambda} \mathrm{e}^{\lambda} = 1, \tag{26}$$

where we have used the power series expansion of $\mathrm{e}^x = (1 + x + x^2/2! + x^3/3! + \cdots)$. The Poisson RV is a good model for a certain limit of a binomial RV where n grows large and $p = p(n)$ grows small such that $np(n) \to \lambda$. The computational benefit of this approximation is that the calculation of the binomial coefficients $\binom{n}{k}$ is cumbersome (although Stirling's approximation is helpful), while the Poisson approximation does not involve these coefficients.

Example. Let $n = 100$ and $p = 0.01$. Consider $X \sim \mathrm{Bin}(n, p)$ and $Y \sim \mathrm{Po}(\lambda)$ for $\lambda = np = 1$. Then for $k = 5$ we find

$$\mathbb{P}(X = 5) = \binom{100}{5} (1/100)^5 (99/100)^{95} \approx 0.00290, \ \mathbb{P}(Y = 5) = \mathrm{e}^{-1} \frac{1}{5!} \approx 0.00306. \tag{27}$$

References

[1] *Introduction to Probability, 2nd Edition* by Dimitri P. Bertsekas and John N. Tsitsiklis, Athina Scientific Press, 2008.