Lecture 2b

ECE 361
Probability for Engineers
Fall, 2016
Steven Weber

Thursday September 29, 2016

DREXEL UNIVERSITY
Electrical and Computer Engineering
College of Engineering

Outline

Bayes' rule

Bayes' rule says: given events A, B

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

Alternate form: given events $(A_1, \ldots, A_N)$ that partition the sample space Ω, and an event B, then, for any $n \in [N]$:

$$\mathbb{P}(A_n|B) = \frac{\mathbb{P}(B|A_n)\mathbb{P}(A_n)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B|A_n)\mathbb{P}(A_n)}{\sum_{k=1}^{N} \mathbb{P}(B|A_k)\mathbb{P}(A_k)}.$$

Key point: the probability of interest, here $\mathbb{P}(A_n|B)$, is expressed in terms of the probabilities $\mathbb{P}(B|A_k)$ and $\mathbb{P}(A_k)$, for each $k \in [N]$.

Problem 1: manipulating probabilities

Problem. Fix events A, B in a sample space Ω. Suppose you are given three probabilities:

- $p_A = \mathbb{P}(A)$
- $p_B = \mathbb{P}(B)$
- $p_o = \mathbb{P}(A \cap B | A \cup B)$

Find an expression for $\mathbb{P}(A|B)$ in terms of these three numbers.

Solution to problem 1

By Bayes' rule:

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

This implies:

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

Solution to problem 1 (continued)

Recall the general result:

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

In our notation, using the result on the previous slide:

$$\mathbb{P}(A \cup B) = p_A + p_B - p_o \mathbb{P}(A \cup B).$$

Solving yields:

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

Solution to problem 1 (continued)

Combining the last two results:

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

and so we come to our goal:

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

Main point: familiarity with the Bayes rule, conditional probability, the probability of a union of two events, etc., is important.

Problem 2: coins in a box

Problem. There are three coins in a box:

- A two-headed coin
- A fair coin,
- 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 to problem 2

Define events:

- T is the event we select the two-headed coin
- F is the event we select the fair coin,
- B is the event we select the biased coin
- H is the event that the selected coin shows heads

We are told: H is true. We are asked to find the probability of T given H, denoted $\mathbb{P}(T|H)$.

Solution to problem 2 (continued)

- The sample space in this problem (recall c denotes complement):

$$\Omega = \{(T,H),(F,H),(B,H),(T,H^c),(F,H^c),(B,H^c)\}.$$

- Here H^c, the complement of heads, means tails.
- Observe $\mathbb{P}(T,H^c) = 0$ as a two-headed coin can not show tails.
- Use Bayes rule since it is more direct to deduce $\mathbb{P}(H|T)$ than to deduce $\mathbb{P}(T|H)$.
- Observe that (T,F,B) partition the sample space.

Solution to problem 2 (continued)

Use Bayes rule and use the partition (T, F, B) of Ω:

$$\begin{aligned} \mathbb{P}(T|H) &= \frac{\mathbb{P}(H|T)\mathbb{P}(T)}{\mathbb{P}(H|T)\mathbb{P}(T) + \mathbb{P}(H|F)\mathbb{P}(F) + \mathbb{P}(H|B)\mathbb{P}(B)} \\ &= \frac{1 \cdot \frac{1}{3}}{1 \cdot \frac{1}{3} + \frac{1}{2} \cdot \frac{1}{3} + p \cdot \frac{1}{3}} \\ &= \frac{1}{\frac{3}{2} + p} = \frac{2}{3 + 2p} \end{aligned}$$

- It should be straightforward for you to confirm this expression gives the intuitive response for special values of p, say $p \in \{0, 1/2, 1\}$.
- Why should we expect that $\mathbb{P}(T|H)$ is a decreasing function of p?

Problem 3: white and black balls in urns

Problem. There are two urns:

- Urn 1 has w_1 white and b_1 black balls
- 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
- 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 to problem 3

Define events:

- H: the event that the coin flip shows heads
- T: the event that the coin flip shows tails
- W: the event that a white ball is selected
- B: the event that a black ball is selected

We are told: W is true. We are asked to find the probability of T given W, denoted $\mathbb{P}(T|W)$.

Solution to problem 3 (continued)

Observe the following are immediate from the problem:

$$\mathbb{P}(W|H) = \frac{w_1}{w_1 + b_1}, \ \mathbb{P}(W|T) = \frac{w_2}{w_2 + b_2}.$$

Apply Bayes' rule, since we want $\mathbb{P}(T|W)$, but we know $\mathbb{P}(W|T)$ and $\mathbb{P}(W|H)$:

$$\begin{aligned} \mathbb{P}(T|W) &= \frac{\mathbb{P}(W|T)\mathbb{P}(T)}{\mathbb{P}(W|T)\mathbb{P}(T) + \mathbb{P}(W|H)\mathbb{P}(H)} \\ &= \frac{\frac{w_2}{w_2+b_2} \cdot \frac{1}{2}}{\frac{w_2}{w_2+b_2} \cdot \frac{1}{2} + \frac{w_1}{w_1+b_1} \cdot \frac{1}{2}} \\ &= \frac{w_2(w_1 + b_1)}{w_2(w_1 + b_1) + w_1(w_2 + b_2)}. \end{aligned}$$

Outline

Outline

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 of the counting principle

- Recall the fact we established in the first lecture: $|\mathcal{P}(\Omega)| = 2^{|\Omega|}$.
- In words: the cardinality of the power set of a finite set Ω is two to the cardinality of Ω.
- Proof using the counting principle: 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.

Outline

Permutations, combinations, lists, and subsets

- Fix a set of n distinct objects and call the set S ($|S| = n$).
- We differentiate between ordered and unordered subsets of S. Recall braces denote an unordered set and parentheses denote an ordered set.
- Combinations from a set S are subsets of S: combinations do not have an order imposed on them: $\{a, b\}$ is the same as $\{b, a\}$.
- If we take a subset of S and impose an order on it, then that subset becomes a list. If the elements of S are numerical then this list coincides with the traditional notion of a vector.
- In general an ordered subset can have repeated elements, e.g., (a, b, a) is a valid list.
- When the ordered subset has *distinct* elements it is called a permutation: (a, b) is distinct from (b, a).

Permutations (continued)

- A k-permutation of an n-set is an ordered subset $R = (s_1, \ldots, s_k)$, where $\{s_1, \ldots, s_k\}$ is a subset of S.
- The number of possible permutations is

$$(n-0) \times (n-1) \times (n-2) \times \cdots \times (n-(k-1)).$$

- Proof: there are n elements and k positions.
  - Each possible permutation is obtained by serially placing an element from S in the k positions
  - There are $n-0$ unplaced elements available for the first position
  - There are $n-1$ unplaced elements available for the second position
  - There are $n-(k-1)$ unplaced elements available for the kth (last) position

Permutations (continued)

- Recall that $n!$ (read as "n-factorial") is by definition:

$$n! = n \times (n-1) \times (n-2) \times \cdots \times 3 \times 2 \times 1.$$

- Observe that the number of possible n-permutations of an n-set S (i.e., the number of elements equals the number of positions) is

$$(n-0) \times (n-1) \times (n-2) \times \cdots \times (n-(n-1)) = n!.$$

Thus, $n!$ is the number of permutations of an n-set where every element of the set is given a position.

Permutations (continued)

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

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

Proof:

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

Permutations (continued)

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

Outline

Combinations

- Combinations are unordered subsets.
- For example, given the set $\{A, B, C, D\}$ the set of 12 possible 2-permutations is

$$AB, BA, AC, CA, AD, DA, BC, CB, BD, DB, CD, DC,$$

while the set of 6 possible 2-combinations is

$$AB, AC, AD, BC, BD, CD.$$

Observe there are 2 permutations for every combination, corresponding to the two orderings of the subset.

Combinations (continued)

- One question of interest is the number of combinations of size k from a set of n distinct objects.
- Key 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.

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

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

This symbol is read as "n choose k". It is *not* a quotient.

Combinations (continued)

A fundamental identity of binomial coefficients is

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

A simple proof uses basic logic: there are 2^n possible subsets of an n-set. Each such subset has a size between 0 and n. There are $\binom{n}{k}$ subsets of size k. Adding up the number of subsets of each possible size gives the total number of subsets.

Combinations (continued)

A second, equally fundamental, identity is

$$\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}.$$

Proof: fix any element, say e, in the base set $[n]$. The collection of all k-subsets can be divided into those k-subsets that contain e and those that don't. There are $\binom{n-1}{k-1}$ k-subsets that contain e (do you see why?). Similarly, there are $\binom{n-1}{k}$ such k-subsets that don't contain e. As all k-subsets either do or do not contain e, the identity follows.

Outline

1. Sample problems related to HW 2
2. §1.6 Counting
   - The counting principle
   - Permutations
   - Combinations
   - Partitions
3. §2.1 Random variables – basic concepts
4. §2.2 Probability mass functions (PMF)
   - The Bernoulli RV
   - The binomial RV
   - The geometric RV
   - The Poisson RV

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 and r, and r additional 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:

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

Observe this reduces to the binomial coefficient when $r = 2$ (with $n_1 = k$ and $n_2 = n - k$).

Partitions (continued)

Example: anagrams. How many different words can be obtained by rearranging the letters in the word TATTOO?

Solution. 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 $6!/(1!2!3!) = 60$.

Partitions (continued)

Example: balls and bins. How many ways can the junior class of the ECE Department, of size n, form senior design teams, each of size k, where k divides n?

Solution. Note n/k is the number of teams. Then the number of different ways to form the teams is

$$\frac{1}{(n/k)!}\binom{n}{k!, \cdots, k!} = \frac{n!}{(n/k)!(k!)^{\frac{n}{k}}},$$

where there are n/k factors in the denominator on the left side. The $(n/k)!$ means we don't care about the ordering of the teams, i.e., the teams are unlabeled.

Partitions (continued)

Recall we just derived:

$$\frac{1}{(n/k)!}\binom{n}{k!,\cdots,k!} = \frac{n!}{(n/k)!(k!)^{\frac{n}{k}}},$$

For $n = 9$ and $k = 3$ (a senior class of 9 students forming three groups of size 3), the answer is 280 distinct ways of forming the teams. The first part of the list is shown below.

$$\begin{pmatrix} 1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1 \\ 2&2&2&2&2&2&2&2&2&2&5&3&2&8&3&2&6&3&2&5&3&2&5&3&2&7&3&2&6&3 \\ 3&3&3&3&3&3&3&3&3&3&6&4&4&9&4&4&7&4&4&7&4&4&9&4&4&8&4&4&9&4 \end{pmatrix}$$

$$\begin{pmatrix} 4&4&4&4&4&4&4&4&4&4&2&2&3&2&2&3&2&2&3&2&2&3&2&2&3&2&2&3&2&2 \\ 5&8&6&5&5&7&6&5&7&6&3&5&5&3&8&8&3&6&6&3&5&5&3&5&5&3&7&7&3&6 \\ 6&9&7&7&9&8&9&8&9&8&4&6&6&4&9&9&4&7&7&4&7&7&4&9&9&4&8&8&4&9 \end{pmatrix}$$

$$\begin{pmatrix} 7&5&5&6&6&5&5&6&5&5&7&7&7&5&5&5&5&5&5&6&6&6&6&6&6&5&5&5&5&5 \\ 8&6&8&8&7&6&7&7&6&7&8&8&8&6&6&6&8&8&8&8&8&8&7&7&7&6&6&6&7&7 \\ 9&7&9&9&8&9&8&9&8&9&9&9&9&7&7&7&9&9&9&9&9&9&8&8&8&9&9&9&8&8 \end{pmatrix}$$

Outline

Random variables – 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 $\omega \in \Omega$.
- We will use RV to denote "random variable" throughout the course.

Random variables – basic concepts (continued)

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

Random variables – basic concepts (continued)

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.

Random variables – basic concepts (continued)

Additional concepts related to RVs.

- In general, RVs may take on a finite, countably infinite, or uncountably infinite set of values.
- Simple examples of these include:
  - Finite: the number of pips on a die roll
  - Countably infinite: the number of coin flips until a head
  - Uncountably infinite: a real number chosen uniformly at random from $[0, 1]$
- In this chapter we focus on *discrete* RVs where the set of values is either discrete or countably infinite.

Discrete RVs – basic concepts

Concepts specific 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.

Outline

§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: $\mathcal{X} = X(\Omega) \equiv \{x \in \mathbb{R} : \exists \omega \in \Omega : X(\omega) = x\}$
  - This is the range of the function 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

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

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

PMF: example

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

$$p(0) = 1/4,\ p(1) = 1/2,\ p(2) = 1/4.$$

- Note that $p(x) = \mathbb{P}(\{X = x\})$ is the probability of the event $X = x$.
- 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 $\bigcup_{x \in \mathcal{X}} E_x = \Omega$.

Calculation of the PMF

Calculation of the PMF of an RV X. For each possible value x of X in $\mathcal{X}$:

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

Outline

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:

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

The corresponding PMF is

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

where $p \in [0, 1]$ is the fixed probability of a head (a biased coin).

The Bernoulli RV (continued)

- I will use the notation $X \sim \mathrm{Ber}(p)$ 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.
- Examples:
  - 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

Outline

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 $X \sim \mathrm{Bin}(n, p)$ to denote that X is a binomial RV for n trials with success probability p.
- 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

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

The binomial RV (continued)

- Recall the binomial theorem: for any $x, y \in \mathbb{R}$ and any $n \in \mathbb{N}$:

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}.$$

- E.g,. for $n = 2$ we have $(x+y)^2 = x^2 + 2xy + y^2$.
- The binomial theorem allows us to verify the binomial distribution sums to one:

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

Outline

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
- 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

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

The geometric RV (continued)

- Let us check that this is a valid PMF:

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

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

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

Outline

The Poisson RV

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

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

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

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

  where we have used the power series expansion of $\mathrm{e}^x = (1 + x + x^2/2! + x^3/3! + \cdots)$.

The Poisson RV (continued)

- 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.

The Poisson RV (continued)

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

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