DISCRETE PROBABILITY

OUTLINE:

1) Basic probability theory
2) Axionms of probability and independence
3) Bayes’s theorem

1. BASIC PROBABILITY THEORY

The set-up

- If we want to rigorously study random phenomena, and have a way of computing how likely the possible outcomes of these phenomena are, we have to formalize them in mathematical terms.
- An experiment is a procedure that produces one outcome from a given set of possible outcomes.
- The sample space of the experiment is the set of possible outcomes of the experiment, that is, the possible outcomes are the elements of the sample space.
- An event is a subset of the sample space, that is, an element of the power set of the sample space.

The set-up

- EX: the roll of a d6 die is an experiment, whose sample space is the set of all the faces the die can possibly show, i.e., S = {1,2,3,4,5,6}. The events are thus the elements of P({1,2,3,4,5,6}). Some instances of events are
  - {1} = ‘the rolled die shows a 1’; {2} = ‘the rolled die shows a 2’
  - {2,4,6} = ‘the rolled die shows an evenly numbered face’
  - {1,2,3,4,5,6} = ‘the rolled die shows any face’ that is, simply, ‘the die is rolled (iacta alea est)’
  - Ø = ‘the rolled die shows no face’

The set-up

- Intuitively, some events are more likely than others. In the previous example (roll of a d6 die), we acknowledge that (if the die is honest) the event {1} is equally likely to the event {2}, and less likely than the event {2,4,6}. The event {1,2,3,4,5,6} is certain (i.e., it has “full likelihood”), as it encompasses all the possible outcomes of the roll. On the opposite end, the event Ø is impossible (i.e., it has “empty likelihood”), as it encompasses no outcome at all.
- We can try to assign to each event a measure of how likely the event is. This measure is called the probability of the event.

Uniform probability

- In general, if an experiment has a finite sample space S and all the outcomes (elements) in the sample space are equally likely, then the probability of an event E ∊ P(S) is the number of favourable outcomes (i.e., the outcomes contained in E) over the number of possible cases (i.e., all the outcomes in S). In set-theoretic language,

$$p(E) = \frac{|E|}{|S|}$$

- This probability is called the uniform probability on S.

Uniform probability

- EX: In the roll of an honest d6 die, with sample space S = {1,2,3,4,5,6}, the possible outcomes 1,2,3,4,5, and 6 are equally likely (that’s the definition of ‘honest d6 die’), therefore the probability of any event E can be computed as p(E) = |E| / |S|:

$$p(\{1\}) = \frac{|\{1\}|}{|S|} = \frac{1}{6} = \frac{|\{2\}|}{|S|} = p(\{2\})$$

$$p(\{2,4,6\}) = \frac{|\{2,4,6\}|}{|S|} = \frac{3}{6} = \frac{1}{2}$$

$$p(\{1,2,3,4,5,6\}) = \frac{|\{1,2,3,4,5,6\}|}{|S|} = \frac{6}{6} = 1$$

$$p(\emptyset) = \frac{|\emptyset|}{|S|} = \frac{0}{6} = 0$$

Uniform probability

- EX: What is the probability that, when two (honest d6) dice are rolled, the sum of the numbers on the two dice is 7?
  - Experiment: roll of 2 dice.
  - Sample space: If we define $D1 = \{1,2,3,4,5,6\}$ to be the outcome of the 1st die and $D2 = \{1,2,3,4,5,6\}$ to be the outcome of the 2nd die, then the sample space is $S = D1 \times D2 = \{1,2,3,4,5,6\} \times \{1,2,3,4,5,6\}$. Note that $|S| = |D1| \cdot |D2| = 36$ (product rule).
  - Event we are interested in: $E = \{(a,b) \in S \mid a+b=7\} = \{(1,6), (2,5), (3,4), (4,3), (5,2), (6,1)\}$
  - $p(E) = |E|/|S| = 6/36 = 1/6$

Uniform probability

- EX: In a famous Canadian lottery, 6 numbers are randomly drawn from the set of integers from 1 to 49. To win, you must buy a ticket matching all 6 numbers. What is the probability of winning?
  - Experiment: draw 6 numbers (without repetitions) from a set of 49.
  - Sample space: If we define A = {1,2,...,49}, the sample space is S = {6-combinations of A}, so

$$|S| = C(49,6) = \frac{49!}{6!(49-6)!} = 13983816$$

  - The event E we are interested in contains exactly 1 element, the only winning 6-combination (it does not matter which one, just that there is 1 and only 1)
  - $p(E) = \frac{|E|}{|S|} = \frac{1}{C(49,6)} = \frac{1}{13983816} \approx 0.0000000715112384$

Uniform probability

- EX: What if, instead, you have to guess the right 6-permutation (correct numbers in the correct order)?
  - Experiment: draw 6 numbers (without repetitions) from a set of 49.
  - Sample space: If we define A = {1,2,...,49}, the sample space is
    S = {6-permutations of A}, so

$$|S| = P(49,6) = 49 \cdot 48 \cdot 47 \cdot 46 \cdot 45 \cdot 44 = \frac{49!}{(49-6)!} = 10068347520$$

  - The event E we are interested in contains exactly 1 element, the only winning 6-permutation (it does not matter which one, just that there is 1 and only 1)
  - $p(E) = \frac{|E|}{|S|} = \frac{1}{P(49,6)} = \frac{1}{10068347520} \approx 0.0000000000993212$

Uniform probability

- EX: What if, instead, the drawn numbers are immediately returned to the pool and you have to guess the correct numbers in the correct order?
  - Experiment: draw 6 numbers (with possible repetitions) from a set of 49.
  - Sample space: If we define A = {1,2,...,49}, the sample space is $S = A^6$, so $|S| = 49^6 = 13841287201$
  - The event E we are interested in contains exactly 1 element
  - $p(E) = \frac{|E|}{|S|} = \frac{1}{49^6} = \frac{1}{13841287201} \approx 0.0000000000722476$

2. AXIOMS OF PROBABILITY AND INDEPENDENCE

Axioms of probability

- We can more generally define probability measures on sample spaces in which the outcomes need not have the same likelihood.
- Let S be the sample space of an experiment with a finite number of outcomes. We first define a probability mass function, a function pmf : S → **R** which assigns a "probability value" to each outcome s∈S, and such that
  - $\forall s \in S\ (0 \leq pmf(s) \leq 1)$
  - $\sum_{s \in S} pmf(s) = 1$
- We then extend pmf to a function p defined on all events, that is, p : $\mathcal{P}$(S) → [0,1], p(E) = $\sum_{s \in E} pmf(s)$. This function is called a probability distribution on S.

Axioms of probability

- The probability mass function is defined based on the experimental observations of the frequency with which each outcome tends to happen. This is part of the modelling of the experiment.
- If $|S| = n$ and each $s \in S$ happens with the same frequency, it makes sense to define $pmf(s) = 1/n$ for any $s \in S$. This defines the uniform probability distribution, which we have encountered earlier.
- EX: If we toss a coin many times and we observe that we get approximately the same number of heads ('H') and tails ('T'), we can be reasonably confident that the coin is fair, and so we can define the probability mass function as follows:
  - Sample space: $S = \{H,T\}$; $pmf(H) = pmf(T)$; as $pmf(H)+pmf(T) = 1$, this forces $pmf(H) = pmf(T) = 1/2$.

Axioms of probability

- The probability mass function is defined based on the experimental observations of the frequency with which each outcome tends to happen. This is part of the modelling of the experiment.
- EX: If we toss a coin many times and we observe that we get approximately 3 times as many heads (‘H’) as tails (‘T’), we can be reasonably confident that the coin is unfair, and a good probability mass function would be:
  - Sample space: S = {H,T}; pmf(H) = 3·pmf(T); as pmf(H)+pmf(T) = 1, this forces pmf(H) = 3/4, pmf(T) = 1/4.

Axioms of probability

- EX: you roll a d6 die many times and you observe that all the numbers are rolled with approximately the same frequency, with the exception of 2, which seems to be rolled half as often as the others. Find a good probability mass function for the experiment and compute the probability of rolling an even number and the probability of rolling an odd one.
- Sample space S = {1,2,3,4,5,6}; the observed frequencies suggest: pmf(1) = pmf(3) = pmf(4) = pmf(5) = pmf(6) = 2·pmf(2)
- Subbing into pmf(1)+pmf(2)+pmf(3)+pmf(4)+pmf(5)+pmf(6) = 1, we get 2·pmf(2)+pmf(2)+2·pmf(2)+2·pmf(2)+2·pmf(2)+2·pmf(2) = 1, i.e. 11·pmf(2) = 1, so
- pmf(2) = 1/11 and pmf(1) = pmf(3) = pmf(4) = pmf(5) = pmf(6) = 2/11
- P({2,4,6}) = pmf(2)+pmf(4)+pmf(6) = 1/11+2/11+2/11 = 5/11
- P({1,3,5}) = pmf(1)+pmf(3)+pmf(5) = 2/11+2/11+2/11 = 6/11

Properties of probability distributions

- For any finite sample space S, equipped with any probability distribution $p : \mathcal{P}(S) \rightarrow [0,1]$, the following properties hold
- $p(S) = 1$ ('total probability': follows directly from the 2nd axiom of probability mass functions)
- $p(E^C) = p(S \setminus E) = 1 - p(E)$ ('probability of the complement': follows from the 2nd axiom of probability mass functions and the fact that $\{E, E^C\}$ is a partition of S)
- $p(E \cup F) = p(E) + p(F) - p(E \cap F)$ ('probability of unions': follows from the inclusion-exclusion principle)

Properties of probability distributions

- For any finite sample space S, equipped with any probability distribution $p : \mathcal{P}(S) \to [0,1]$,
- If $E_1, E_2, ..., E_k$ are pairwise disjoint events, then $p(E_1 \cup E_2 \cup ... \cup E_k) = p(E_1)+p(E_2)+...+p(E_k)$ (that's again, in essence, an application of the inclusion-exclusion principle)

Conditional probability

- Let S be a sample space and let E, F be events with p(F) > 0. The conditional probability of E given F is

$$p(E|F) = \frac{p(E \cap F)}{p(F)}$$

- This probability represents the likelihood of event E to happen, knowing that event F has surely happened.

Conditional probability

- EX: An honest d6 die is rolled. An oracle reveals that the outcome is an even number.
- What is the probability that a 2 is rolled?
  - In the above notations, we have S = {1,2,3,4,5,6}, E = {2}, F = {2,4,6} (note that p(F) = 1/2 ≠ 0). The probability that 2 is rolled, knowing that an even number is rolled, is

$$p(E|F) = \frac{p(E \cap F)}{p(F)} = \frac{p(\{2\} \cap \{2,4,6\})}{p(\{2,4,6\})} = \frac{p(\{2\})}{p(\{2,4,6\})} = \frac{1/6}{1/2} = \frac{1}{3}$$

  - Note that P(E|F) ≠ P(E) = 1/6: the knowledge that the roll produced an even number modifies the probability of having rolled a 2.
- In the same set-up, what is the probability that a 1 is rolled?
  - This time we have S = {1,2,3,4,5,6}, E = {1}, F = {2,4,6}. Therefore

$$p(E|F) = \frac{p(E \cap F)}{p(F)} = \frac{p(\{1\} \cap \{2,4,6\})}{p(\{2,4,6\})} = \frac{p(\emptyset)}{p(\{2,4,6\})} = \frac{0}{1/2} = 0$$

3. BAYES’S THEOREM