Worksheet on Probability

Benjamin Cosman, Patrick Lin and Mahesh Viswanathan

Fall 2020

Takeaways from Lecture

- Probability space:
  - *Sample space* S, the set of possible outcomes
  - *Probability distribution*: $\Pr : S \to [0,1]$ so that $\sum_{x \in S} \Pr[x] = 1$
  - For an *event* $E \subseteq S$, $\Pr[E] = \sum_{x \in E} \Pr[x]$
- "Counting" rules:
  - Sum Rule: For *disjoint* $E_1, \ldots, E_n$, $\Pr\left[\bigcup_{i=1}^{n} E_i\right] = \sum_{i=1}^{n} \Pr[E_i]$
  - Difference Rule: $\Pr[A \setminus B] = \Pr[A] - \Pr[A \cap B]$
  - Union Bound: $\Pr\left[\bigcup_{i=1}^{n} E_i\right] \leq \sum_{i=1}^{n} \Pr[E_i]$
  - Monotonicity Rule: If $A \subseteq B$, then $\Pr[A] \leq \Pr[B]$
- Conditional probability:
  - $\Pr[x|B]$: the probability of outcome x, *given* event B
  - Kolmogorov's Rule: $\Pr[A \cap B] = \Pr[A|B]\Pr[B]$
  - Bayes' Rule: $\Pr[A|B] = \frac{\Pr[B|A]\Pr[A]}{\Pr[B]}$
  - Events A and B are *independent* if and only if $\Pr[A|B] = \Pr[A]$ if and only if $\Pr[A \cap B] = \Pr[A]\Pr[B]$

Problem 1. Two standard six-sided dice, one orange and one blue, are rolled. Suppose the orange die rolled an even number. What is the (conditional) probability that the sum of the two rolls is at least 7?

Problem 2. Consider an integer grid, where we want to walk from $(0, 0)$ to (m, n), by only walking rightwards or upwards.

a) A rather annoying construction project has blocked coordinate (a, b). Suppose you pick a valid path from $(0, 0)$ to (m, n) uniformly at random.[1] What is the probability that you successfully reach (m, n) without running into construction? (You do not need to simplify your answer.)

b) The next day, construction has gotten worse. Now they have blocked off all coordinates of the form (a, b) where $a + b = m$. Now what is the probability of not running into construction? (Assume that $n > 0$.)

[1] Recall that this means that each path is chosen with equal probability.

Problem 3. Consider a graph with n vertices and m edges. Suppose we independently color each vertex with one of k colors, chosen uniformly at random. Compute upper and lower bounds on the probability that this results in a proper coloring of the graph.[2]

[2] Recall that in a proper coloring, the endpoints of each edge must have different colors.

Problem 4. You are writing an antivirus software. You find a code segment that appears in 95% of all infected files, and happily submit a rule that flags a file as infected if it contains the code segment. During code review, your colleague points out that while the probability that the code segment appears in an infected file is 95%, this does not mean that if the code segment appears, then the file is infected with probability 95%. After running some more data analysis, you find that this code segment also appears in 25% of non-infected files, and that, overall, only 1% of files are infected. Compute the (conditional) probability that a file containing this code segment is actually infected, to two or three digits of precision.[3] Should this rule pass code review?

[3] Hint: You may find the following identity useful: for sets A and B, $A = (A \cap \overline{B}) \cup (A \cap B)$.

Problem 5. Consider the following statements, and determine if each is true or false. Prove your answer.

a) For all events A and B, $\Pr[A|B] = \Pr[B|A]$.

b) For all events A and B, $\Pr[A \cap B|B] = \Pr[A|B]$.

c) For all events A and B, $\Pr[A] \leq \Pr[A|B]$.

d) For all events A and B, $\Pr[A \cap B] \leq \Pr[A|B]$.

e) For all events A and B, $\Pr[A|B] > \Pr[A]$ implies $\Pr[B|A] > \Pr[B]$.