CMPT 210: Probability and Computing

Lecture 6

Sharan Vaswani

January 20, 2023

Recap - Axioms of Probability

Sample (outcome) space $\mathcal{S}$: Nonempty (countable) set of possible outcomes. Example: When we threw one dice, the sample space is $\{1, 2, 3, 4, 5, 6\}$.

Outcome $\omega \in \mathcal{S}$: Possible "thing" that can happen. Example: When we threw one dice, a possible outcome is $\omega = 1$.

Event E: Any subset of the sample space. Example: When we threw one dice, a possible event is $E = \{6\}$ (first example) or $E = \{3, 6\}$ (second example).

Probability function on a sample space $\mathcal{S}$ is a total function $\Pr : \mathcal{S} \to [0, 1]$. For any $\omega \in \mathcal{S}$,

$$0 \leq Pr[\omega] \leq 1 \quad ; \quad \sum_{\omega \in \mathcal{S}} \Pr[\omega] = 1 \quad ; \quad \Pr[E] = \sum_{\omega \in E} \Pr[\omega]$$

Recap - Probability rules

Union: For mutually exclusive events $E_1, E_2, \ldots, E_n$,
$\Pr[E_1 \cup E_2 \cup \ldots E_n] = \Pr[E_1] + \Pr[E_2] + \ldots + \Pr[E_n]$.

Complement rule: $\Pr[E] = 1 - \Pr[E^c]$

Inclusion-Exclusion rule: For any two events E, F, $\Pr[E \cup F] = \Pr[E] + \Pr[F] - \Pr[E \cap F]$.

Union Bound: For any events $E_1, E_2, E_3, \ldots E_n$, $\Pr[E_1 \cup E_2 \cup E_3 \ldots \cup E_n] \leq \sum_{i=1}^{n} \Pr[E_i]$.

Uniform probability space: A probability space is said to be uniform if $\Pr[\omega]$ is the same for every outcome $\omega \in \mathcal{S}$. In this case, $\Pr[E] = \frac{|E|}{|\mathcal{S}|}$.

Probability Examples

Q: From a set of n items a random sample of size k is to be selected. Given two items of interest: α and β, what is the probability that (i) both α and β will be among the k selected (ii) at least one of α or β will be among the k selected (iii) neither α nor β will be among the k selected?

(i) If we want both α and β to be in the sample, number of ways of choosing the other items = $\binom{n-2}{k-2}$. Hence, probability that both α and β will be in the sample = $\frac{\binom{n-2}{k-2}}{\binom{n}{k}} = \frac{k(k-1)}{n(n-1)}$.

(ii) Let A be the event that item α is in the selection. $\Pr[A] = \frac{k}{n}$. Similarly B be the event that item β is in the selection. $\Pr[B] = \frac{k}{n}$. We want to compute $\Pr[A \cup B]$. By the union-rule, $\Pr[A \cup B] = \Pr[A] + \Pr[B] - \Pr[A \cap B]$. Hence, probability that either α or β will be among the k selected items = $\frac{2k}{n} - \frac{k(k-1)}{n(n-1)}$.

(iii) If we want neither α nor β to be in the sample, number of ways of choosing the items = $\binom{n-2}{k}$. Hence, probability that neither α nor β will be in the sample = $\frac{\binom{n-2}{k}}{\binom{n}{k}} = \frac{(n-k)(n-k-1)}{n(n-1)}$.

Probability - Examples

Q: Let us consider random permutations of the letters (i) ABBA (ii) ABBA'. What is the probability that the third letter is B?

Ans: **(i)** $|\mathcal{S}| = \frac{4!}{2!\,2!} = 6$. $|E| = \frac{3!}{2!1!} = 3$. $\Pr[E] = \frac{1}{2}$.
(ii) $|\mathcal{S}| = \frac{4!}{2!\,1!\,1!} = 12$. $|E| = \frac{3!}{1!1!} = 6$. $\Pr[E] = \frac{1}{2}$.

Questions?

Birthday Paradox

Q: There are 54 students in the class. What is the probability that two students have their birthdays in the same week? Ans: 1. By the pigeonhole principle, there has to be a pair of students that have their birthdays in the same week.

Q: In this class, what is the probability that two students share the same birthday? Assume that (i) each student is equally likely to be born on any day of the year, (ii) no leap years and (iii) student birthdays are independent of each other.

Let n be the number of students, and let d be the number of days in the year. Let's order the students according to their ID. A birthday sequence is (11 Feb, 23 April, 31 August, $\ldots$). First let's count the number of possible birthday sequences.

The first student's birthday can be one of d days. Similarly, the second student's birthday can be one of d days, and so on. By the product rule, the total number of birthday sequences = $d \times d \times \ldots = d^n$.

Birthday Paradox

The event of interest is that two students share the same birthday. Let us compute the probability of the event that NO two students share the same birthday, and then use the complement rule.

The first birthday can be chosen in d ways, the second in $d-1$ ways, and so on. By the generalized product rule, the number of birthday sequences such that no birthday is shared = $d \times (d-1) \times (d-2) \times \ldots (d-(n-1))$.

Hence, the probability that no two students share the same birthday

$$= \frac{\text{the number of birthday sequences such that no birthday is shared}}{\text{total number of birthday sequences}} = \frac{d \times (d-1) \times (d-2) \times \ldots (d-(n-1))}{d^n}$$

$$= \left(1 - \frac{0}{d}\right) \times \left(1 - \frac{1}{d}\right) \ldots \left(1 - \frac{n-1}{d}\right) \leq \exp(-0/d) \times \exp(-1/d) \ldots \exp(-(n-1)/d)$$

$$(\text{for } x > 0,\ 1 - x \leq \exp(-x))$$

$$= \exp\left(\frac{-0}{d} + \frac{-1}{d} + \ldots \frac{-(n-1)}{d}\right) = \exp\left(-\frac{n(n-1))}{2d}\right)$$

Birthday Paradox

Probability that two students share a birthday $\geq 1 - \exp\left(-\frac{n(n-1))}{2d}\right)$. Let's plot for $d = 365$.

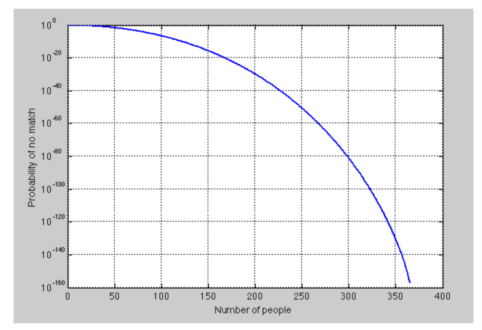

Figure 1: Plotting $\exp\left(-\frac{n(n-\mathbf{1}))}{\mathbf{2}d}\right)$ for $d = 365$

In our class, there is $> 96.4\%$ that two students have the same birthday!

Birthday Principle

If there are n pigeons and d pigeonholes, then the probability that two pigeons occupy the same hole is $\geq 1 - \exp\left(-\frac{n(n-1))}{2d}\right)$

For $n = \lceil\sqrt{2d}\rceil$, probability that two pigeons occupy the same hole is about $1 - \frac{1}{e} \approx 0.632$.

Example: If we are randomly throwing $\lceil\sqrt{2d}\rceil$ balls into d bins, then the probability that two balls land in the same bin is around 0.632.

We will see applications of this principle to hashing, load balancing.

Questions?

Conditional Probability

Conditioning is revising probabilities based on partial information (an event).

Q: Suppose we throw a "standard" dice, what is the probability of getting a 6 if we are told that the outcome is even?

We wish to compute Pr[we get a 6|the outcome is even] or Probability of getting a 6 *given* that the outcome is even or Probability of a 6 *conditioned on the event* that the outcome is even.

Sample space: $\mathcal{S} = \{1, 2, 3, 4, 5, 6\}$, Event: $E = \{6\}$. Additional information: Event $F = \{2, 4, 6\}$ has happened. With conditioning on F, *new sample space* $S' = F = \{2, 4, 6\}$. Since each outcome in $S' = \{2, 4, 6\}$ is equally likely, the new probability space is still a uniform probability space. Hence, conditioned on the event that the outcome is even, $\sum_{\omega \in S'} \Pr[\omega] = 1$ and $\Pr[\{\text{even number}\}] = \frac{1}{3}$ and $\Pr[\{\text{odd number}\}] = 0$. Hence, $\Pr[\{6\}] = \frac{1}{3}$.

Q: What is the probability of getting either a 3 or 6 if we are told that the outcome is even? $E = \{3, 6\}$, $F = \{2, 4, 6\}$. With conditioning, the new sample space $S' = \{2, 4, 6\}$. By the same reasoning as above, $\Pr[\{6\}] = \frac{1}{3}$ and $\Pr[\{3\}] = 0$. Hence, $\Pr[E] = \Pr[\{3\}] + \Pr[\{6\}] = \frac{1}{3}$.

Conditional Probability

Q: Suppose we throw two standard dice one after the other. What is the probability that the sum of the dice is 6?

Recall the sample space consists of tuples, i.e. $S = \{(1,1),(1,2),(1,3),\ldots,(6,6)\}$. The event E consists of outcomes such that the sum of the dice is 6. $E = \{(1,5),(2,4),(3,3),(4,2),(5,1)\}$. Since all outcomes are equally likely, this is a uniform probability space, and $\Pr[E] = \frac{|E|}{|S|} = \frac{5}{36}$.

Q: Suppose I tell you that the first dice came up 4. Given this information, what is the probability that the sum of the dice is 6?

Let F be the event that the first dice came up 4. We wish to compute $\Pr[E|F]$, the probability that the sum of the dice is 6 conditioned on the event that the first dice came up 4.

With conditioning, the new sample space $S' = F = \{(4,1),(4,2),(4,3),(4,4),(4,5),(4,6)\}$. For this new probability space, $E = \{(4,2)\}$. Since each outcome in S' is equally likely, $\Pr[E|F] = \frac{|E|}{|S'|} = \frac{1}{6}$.