Probability & Statistics for EECS: Homework #1 Solution

Problem 1

Define $\left\{ \begin{matrix} n \\ k \end{matrix} \right\}$ as the number of ways to partition $\{1, 2, \ldots, n\}$ into k non-empty subsets, or the number of ways to have n students split up into k groups such that each group has at least one student. For example, $\left\{ \begin{matrix} 4 \\ 2 \end{matrix} \right\} = 7$ because:

- $\{1\}, \{2, 3, 4\}$,
- $\{2\}, \{1, 3, 4\}$,
- $\{3\}, \{1, 2, 4\}$,
- $\{4\}, \{1, 2, 3\}$.
- $\{1, 2\}, \{3, 4\}$,
- $\{1, 3\}, \{2, 4\}$,
- $\{1, 4\}, \{2, 3\}$,

Prove the following identities:

(a)

$$\left\{ \begin{matrix} n+1 \\ k \end{matrix} \right\} = \left\{ \begin{matrix} n \\ k-1 \end{matrix} \right\} + k \left\{ \begin{matrix} n \\ k \end{matrix} \right\}.$$

(b)

$$\sum_{j=k}^{n} \binom{n}{j} \left\{ \begin{matrix} j \\ k \end{matrix} \right\} = \left\{ \begin{matrix} n+1 \\ k+1 \end{matrix} \right\}.$$

Solution

(a) The left-hand side is the number of ways to divide $n+1$ people into k nonempty groups. Now let's count this a different way. Say I'm the $(n+1)$st person. Either I'm in a group by myself or I'm not. If I'm in a group by myself, then there are $\left\{ \begin{matrix} n \\ k-1 \end{matrix} \right\}$ ways to divide the remaining n people into $k-1$ nonempty groups. Otherwise, the n people other than me form k nonempty groups, which can be done in $\left\{ \begin{matrix} n \\ k \end{matrix} \right\}$ ways, and then I can join any of those k groups. So in total, there are $\left\{ \begin{matrix} n \\ k-1 \end{matrix} \right\} + k \left\{ \begin{matrix} n \\ k \end{matrix} \right\}$ possibilities, which is the right-hand side.

(b) The right-hand side is the number of ways to divide $n+1$ people into $k+1$ nonempty groups. Say I'm the $(n+1)$st person. Let j be the number of people *not* in my group. Then $k \leq j \leq n$. The number of possible divisions with j people not in my group is $\binom{n}{j} \left\{ \begin{matrix} j \\ k \end{matrix} \right\}$ since we have $\binom{n}{j}$ possibilities for which j specific people are not in my group (and then it's determined who *is* in my group) and then $\left\{ \begin{matrix} j \\ k \end{matrix} \right\}$ possibilities for how to divide those j people into k groups that are not my group. Summing over all possible j gives the left-hand side.

Problem 2

A *norepeatword* is a sequence of at least one (and possibly all) of the usual 26 letters $a, b, c, \ldots, z$, with repetitions not allowed. For example, "course" is a norepeatword, but "statistics" is not. Order matters, *e.g.*, "course" is not the same as "source". A norepeatword is chosen randomly, with all norepeatwords equally likely. Show that the probability that it uses all 26 letters is very close to $1/e$.

Solution

The number of norepeatwords having all 26 letters is the number of ordered arrangements of 26 letters: 26!. To construct a norepeatword with $k \leq 26$ letters, we first select k letters from the alphabet $\left(\binom{26}{k} \right.$ selections) and then arrange them into a word (k ! arrangements). Hence there are $\binom{26}{k} k!$ norepeatwords with k letters, with k ranging from 1 to 26 . With all norepeatwords equally likely, we have

$$\begin{aligned} P(\text{ norepeatword having all 26 letters }) &= \frac{\# \text{ norepeatwords having all 26 letters}}{\# \text{ norepeatwords}} \\ &= \frac{26!}{\sum_{k=1}^{26} \binom{26}{k} k!} = \frac{26!}{\sum_{k=1}^{26} \frac{26!}{k!(26-k)!} k!} \\ &= \frac{1}{\frac{1}{25!} + \frac{1}{24!} + \ldots + \frac{1}{1!} + 1} \\ &\approx \frac{1}{e}. \end{aligned}$$

The denominator is the first 26 terms in the Taylor series $e^x = 1 + x + x^2/2! + \ldots$, evaluated at $x = 1$. Thus the probability is approximately $1/e$ (this is an extremely good approximation since the series for e converges very quickly; the approximation for e differs from the truth by less than 10^{-26}).

Problem 3

An academic department offers 8 lower level courses: $\{L_1, L_2, \ldots, L_8\}$ and 10 higher level courses: $\{H_1, H_2, \ldots, H_{10}\}$. A valid curriculum consists of 4 lower level courses and 3 higher level courses.

(a) How many different curricula are possible?

(b) Suppose that $\{H_1, \ldots, H_5\}$ have L_1 as a prerequisite, and $\{H_6, \ldots, H_{10}\}$ have L_2 and L_3 as prerequisites, *i.e.*, any curricula which involve, say, one of $\{H_1, \ldots, H_5\}$ must also include L_1. How many different curricula are there?

Solution

(a) Becuase we need to choose 4 courses from 8 lower level courses, and 3 courses from 10 higher level courses, so the number of different valid curricula:

$$\binom{8}{4} \times \binom{10}{3} = 8400$$

(b) There exists four valid cases in total as below:

- choose L_1; no L_2, L_3: We need to select 3 courses from $\{L_4, L_5, L_6, L_7, L_8\}$, so the total number of combinations of lower level courses is $\binom{5}{3}$. Also, we need to select 3 courses from $\{H_1, H_2, H_3, H_4, H_5\}$ (because we choose L_1 but no L_2, L_3 as prerequisite), so the total number of combinations of higher level courses is $\binom{5}{3}$. So the total number of valid curricula in this case is as below:

$$\binom{5}{3} \times \binom{5}{3} = 100$$

- choose L_1; no L_2, choose L_3:

$$\binom{5}{2} \times \binom{5}{3} = 100$$

- choose L_1; choose L_2, no L_3:

$$\binom{5}{2} \times \binom{5}{3} = 100$$

- no L_1; choose L_2, L_3:

$$\binom{5}{2} \times \binom{5}{3} = 100$$

- choose L_1; choose L_2, L_3:

$$\binom{5}{1} \times \binom{10}{3} = 600$$

The total number of different curricula is $100 \times 4 + 600 = 1000$.

Problem 4

In the birthday problem, we assumed that all 365 days of the year are equally likely (and excluded February 29). In reality, some days are slightly more likely as birthdays than others. For example, scientists have long struggled to understand why more babies are born 9 months after a holiday. Let $p = (p_1, p_2, \dots, p_{365})$ be the vector of birthday probabilities, with p_j the probability of being born on the jth day of the year (February 29 is still excluded, with no offense intended to Leap Dayers). The kth elementary symmetric polynomial in the variables $x_1, \dots, x_n$ is defined by

$$e_k(x_1, \dots, x_n) = \sum_{1 \le j_1 < j_2 < \dots < j_k \le n} x_{j_1} \dots x_{j_k}.$$

This just says to add up all of the $\binom{n}{k}$ terms we can get by choosing and multiplying k of the variables. For example, $e_1(x_1, x_2, x_3) = x_1 + x_2 + x_3$, $e_2(x_1, x_2, x_3) = x_1x_2 + x_1x_3 + x_2x_3$, and $e_3(x_1, x_2, x_3) = x_1x_2x_3$. Now let $k \ge 2$ be the number of people.

(a) Find a simple expression for the probability that there is at least one birthday match, in terms of $\mathbf{p}$ and an elementary symmetric polynomial.

(b) Explain intuitively why it makes sense that P(at least one birthday match) is minimized when $p_j = \frac{1}{365}$ for all j, by considering simple and extreme cases.

(c) The famous arithmetic mean-geometric mean inequality says that for $x, y \ge 0$

$$\frac{x+y}{2} \ge \sqrt{xy}.$$

This follows from adding $4xy$ to both sides of $x^2 - 2xy + y^2 = (x-y)^2 \ge 0$. Define $\mathbf{r} = (r_1, \dots, r_{365})$ by $r_1 = r_2 = (p_1 + p_2)/2, r_j = p_j$ for $3 \le j \le 365$. Using the arithmetic mean-geometric mean bound and the fact, **which you should verify**, that

$$\begin{aligned} e_k(x_1, \dots, x_n) = \; & x_1x_2e_{k-2}(x_3, \dots, x_n) \\ & + (x_1 + x_2)e_{k-1}(x_3, \dots, x_n) + e_k(x_3, \dots, x_n), \end{aligned}$$

show that

$$P(\text{at least one birthday match} \mid \mathbf{p}) \ge P(\text{at least one birthday match} \mid \mathbf{r})$$

with strict inequality if $\mathbf{p} \ne \mathbf{r}$, where the given $\mathbf{r}$ notation means that the birthday probabilities are given by $\mathbf{r}$. Using this, show that the value of $\mathbf{p}$ that minimizes the probability of at least one birthday match is given by $p_j = \frac{1}{365}$ for all j.

Solution

(a) If $k > 365$, then from the pigeonhole principle, there must be at least two people with the same birthday;

$$P(\text{at least one birthday match}) = 1$$

Otherwise, for $k \le 365$, we consider that choose k days from a year, then adding up all of the $\binom{365}{k}$ terms we get by choosing and multiplying k of the variables can be denoted by $e_k(\mathbf{p})$. For each combination of days we choose, there is $k!$ matching for k people and k days.

$$P(\text{at least one birthday match}) = 1 - k!e_k(\mathbf{p})$$

So we get the simple expression

$$P(\text{at least one birthday match}) \begin{cases} 1, \text{if } k > 365 \\ 1 - k! e_k(\mathbf{p}), \text{if } 2 \leq k \leq 365 \end{cases}$$

(b) An extremely extreme case would be if $p_j = 1$ for some j, i.e., everyone is always born on the same day; then a match is guaranteed if there are at least 2 people. For another simple case, suppose that there are only 2 days in a year, with probabilities p and $q = 1 - p$. For 2 people, the probability of a match is $p^2 + q^2 = p^2 + (1-p)^2$, which is minimized at $p = 1/2$. In the general case, imagine starting with probabilities $1/365$ for all days and shifting some "probability mass" from some p_i to another p_j. This makes it less likely to have a match on day i and more likely for there to be a match on day j, but from the above it makes sense that the latter outweighs the former. i.e. The shift increases the probability of 'at least one birthday match'.

Another answer: We simply consider the case $k = 2$, $p_1 + p_2 = \frac{2}{365}$ with $p_1 \neq \frac{1}{365}, p_2 \neq \frac{1}{365}$ and $p_i = \frac{1}{365}, i = [3, 4, ..., 365]$. The probability that the two persons have the same birthday is

$$P(\text{at least one birthday match}) = \sum_i p_i^2$$

From the AM–GM inequality, we know that $p_1^2 + p_2^2$ got minimum when $p_1 = p_2 = \frac{1}{365}$.
Thus, $P(\text{at least one birthday match})$ gets minimum when $p_i = \frac{1}{365}$ for all $i \in [1, 2, ..., 365]$.

(c) The identity for $e_k(x_1, \ldots, x_n)$ is true since it is just partitioning the terms into 3 cases: terms with both x_1 and x_2, terms with one but not the other, and terms with neither x_1 nor x_2. Let $n = 365$. Note that $p_1 + p_2 = r_1 + r_2$ and by the arithmetic mean-geometric mean inequality, $p_1 p_2 \leq ((p_1 + p_2)/2)^2 = r_1 r_2$. So

$$\begin{aligned} e_k(p_1, \ldots, p_n) &= p_1 p_2 e_{k-2}(p_3, \ldots, p_n) + (p_1 + p_2) e_{k-1}(p_3, \ldots, p_n) + e_k(p_3, \ldots, p_n) \\ &\leq r_1 r_2 e_{k-2}(r_3, \ldots, r_n) + (r_1 + r_2) e_{k-1}(r_3, \ldots, r_n) + e_k(r_3, \ldots, r_n) \\ &= e_k(r_1, \ldots, r_n). \end{aligned}$$

So by (a), the probability of a birthday match when the probabilities are p is at least as large was when they are r. The inequality is strict unless $p_1 = p_2$.
Now let $\mathbf{p}_0$ be a vector of birthday probabilities that minimizes the probability of at least one birthday match. If two components of $\mathbf{p}_0$ are not equal, the above shows that we could replace those two components by their average in order to obtain a smaller chance of a match, but this would contradict $\mathbf{p}_0$ minimizing the probability of a match. Thus, $\mathbf{p}_0$ has all components equal.

Problem 5

Note that for each natural number n, we have the following equation:

$$1^3 + 2^3 + \cdots + n^3 = (1 + 2 + \cdots + n)^2.$$

In this problem, we will try to prove this identity with the technique of story proof.

(a) Give a story proof of the identity

$$1 + 2 + \cdots + n = \binom{n+1}{2}.$$

(b) Give a story proof of the identity

$$1^3 + 2^3 + \cdots + n^3 = 6\binom{n+1}{4} + 6\binom{n+1}{3} + \binom{n+1}{2}.$$

It is then just basic algebra (not required for this problem) to check that the square of the right-hand side in (a) is the right-hand side in (b).

Solution

(a) Consider a chess tournament with $n+1$ players, where everyone plays against everyone else once. Λ total of $\binom{n+1}{2}$ games are played. Label the players $0, 1, \ldots, n$. Player 0 plays n games, player 1 plays $n-1$ games not already accounted for, player 2 plays $n-2$ games not already accounted for, etc. So

$$n + (n-1) + (n-2) + \cdots + 1 = \binom{n+1}{2}.$$

(b) Following the hint, let us count the number of choices of (i, j, k, l) where i is greater than j, k, l. Given i, there are i^3 choices for (j, k, l), which gives the left-hand side. On the other hand, consider 3 cases: there could be 2, 3, or 4 distinct numbers chosen. There are $\binom{n+1}{2}$ ways to choose 2 distinct numbers from $0, 1, \ldots, n$, giving, e.g., $(3, 1, 1, 1)$. There are $\binom{n+1}{4}$ ways to choose 4 distinct numbers, giving, e.g., $(5, 2, 1, 4)$, but the $(2, 1, 4)$ could be permuted in any order so we multiply by 6 . There are $\binom{n+1}{3}$ ways to choose 3 distinct numbers, giving, e.g., $(4, 2, 2, 1)$, but the $2, 2, 1$ can be in any order and could have been $1, 1, 2$ in any order also, again giving a factor of 6. Adding these cases gives the right-hand side.