2023 Spring Probability & Statistics for EECS February 12, 2023

Homework 01

Professor: Ziyu Shao & Dingzhu Wen *Due: 23:59 on February 19, 2023*

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

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 $1/e$.

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?

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 $p = (p_1, p_2, \ldots, p_{365})$ 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 $x_1, \ldots, x_n$ is defined by

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

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 $k \geq 2$ be the number of people.

(a) Find a simple expression for the probability that there is at least one birthday match, in terms of **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 $x, y \geq 0$

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

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

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

show that

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

with strict inequality if $\mathbf{p} \neq \mathbf{r}$, 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.

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).