PROBLEM 1. The digits $1, 2, 3, 4$ are randomly arranged into two two-digit numbers $\overline{AB}$ and $\overline{CD}$—each of the four digits is used exactly once. In this problem you will ultimately determine the expected value of $\overline{AB} \cdot \overline{CD}$.

(a) Randomly choose two digits from the set $\{1, 2, 3, 4\}$ without replacement (for example, we cannot choose 1 twice). What is their expected product? [To get started: create an appropriate sample space S and random variable $X\colon S \to \mathbb{R}$.]

(b) Note that $\overline{AB}$ is a linear combination of A and B: namely, $\overline{AB} = 10A + B$. A similar statement holds for $\overline{CD}$. Use this fact along with part (i) and linearity of expectation to determine the expected value $E(\overline{AB} \cdot \overline{CD})$.

(c) (Return to this after you are done with problem 2.) What is the expected value of drawing a single number from $\{1, 2, 3, 4\}$? Given your answer to part (a), is the expected value multiplicative?

SOLUTION:

(a) We can take the sample space to be

$$S = \{\{1, 2\}, \{1, 3\}, \{1, 4\}, \{2, 3\}, \{2, 4\}, \{3, 4\}\}.$$

Let $X\colon S \to \mathbb{R}$ be the random variable $X(\{i, j\}) = ij$. The values for X are $2, 3, 4, 6, 8, 12$, each achieved once (with equal probability. Therefore,

$$E(X) = 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + \cdots + 12 \cdot \frac{1}{6} = \frac{35}{6} = 5.8333\ldots.$$

(b) By linearity of expectation,

$$\begin{aligned} E(\overline{AB} \cdot \overline{CD}) &= E((10A + B)(10C + D)) \\ &= E(100A \cdot C + 10A \cdot D + 10B \cdot C + B \cdot D) \\ &= 100E(A \cdot C) + 10E(A \cdot D) + 10E(B \cdot C) + E(B \cdot D). \end{aligned}$$

From the previous problem, we know that $E(A \cdot C) = \cdots = E(B \cdot D) = 35/6$. Therefore,

$$E(\overline{AB} \cdot \overline{CD}) = (100 + 10 + 10 + 1)\frac{35}{6} = \frac{4235}{6} = 705.8333\ldots.$$

(c) The expected value is

$$E(X) = 1 \cdot \frac{1}{4} + 2 \cdot \frac{1}{4} + 3 \cdot \frac{1}{4} + 4 \cdot \frac{1}{4} = \frac{5}{2} = 2.5.$$

Since

$$2.5^2 = 6.25 \neq 5.8333...,$$

the expected value is not multiplicative.

PROBLEM 2. (The coupon collector problem.) Safeway is running a promotion in which they have produced n coupons and you randomly receive a coupon each time you check out. You passionately hope to one day collect all n coupons. What is the expected number of times T you'll have to check out at the store in order to collect all n? There's a very clever way to solve this problem with linearity of expectation!

(a) Label the coupons $C_1, C_2, \ldots, C_n$. If $n = 4$, a successful collection of all 4 coupons might look like $C_2\ C_2\ C_4\ C_2\ C_1\ C_3$. Break the sequence into segments where a segment ends when you receive a new coupon. In the example sequence, the segments are:

$$C_2, \quad C_2\, C_4, \quad C_2\, C_1, \quad C_3.$$

Because it will make our lives easier, consider these the 0-th, 1-st, …, 3-rd segments (as opposed to 1-st through 4-th). Let X_k be the length of the k-th segment, and note that k ranges from 0 through $n-1$. In the example, $X_0 = 1$, $X_1 = 2$, $X_2 = 2$, and $X_3 = 1$. Express T, the total number of checkouts needed to collect all coupons, as a linear combination of the X_k.

(b) Compute p_k, the probability that you will collect a new coupon given that you have already collected k of them. After studying the geometric distribution, we will learn that $E(X_k) = 1/p_k$. Compute this value.

(c) Use your answers to (i) and (ii) to determine $E(T)$.

(d) Can you say anything about the asymptotic behavior of $E(T)$?

SOLUTION:

(a) $T = X_0 + X_1 + \cdots + X_{n-1}$.

(b) We are seeking to collect one of the $n-k$ uncollected coupons out of the n total coupons, so $p_k = \frac{n-k}{n}$ and $E(X_k) = \frac{1}{p_k} = \frac{n}{n-k}$.

(c) By linearity of expectation,

$$\begin{aligned} E(T) &= \sum_{k=0}^{n-1} E(X_k) \\ &= \sum_{k=0}^{n-1} \frac{n}{n-k} \\ &= n \sum_{k=0}^{n-1} \frac{1}{n-k} \\ &= n \sum_{i=1}^{n} \frac{1}{i}. \end{aligned}$$

(d) It is beyond the scope of this course to prove so, but

$$E(T) = n(\log n + \gamma) + O(1/n)$$

where $\gamma \approx 0.577$ is the *Euler-Mascheroni constant*.