SEVERAL PROBABILITY QUESTIONS WITH SIMPLE SOLUTIONS

I felt frustrated today because I spent much more time figuring out the solution with the standard method. Actually the following questions can be solved using very basic combinatorial arguments along with the concept of indicators.

The indicator function of a subset A of a set X is a function

$$\mathbb{I}_A : X \mapsto \{0, 1\}$$

defined as

$$\mathbb{I}_A(x) = \begin{cases} 1 & x \in A \\ 0 & x \notin A \end{cases}$$

It is very useful to use indicators in computing the expected number of distinct objects.

Question 1: Sample m times with replacement from a box of n labeled balls. How many distinct balls will you see on average?

Approach One: (Standard difference equation) Let X_i denote the number of distinct balls so far after i-th round. Notice that $X_1 = 1$ is a deterministic value.

Then

$$X_{i+1} = X_i + \mathbb{I}_{\{new\, ball\, selected\}}$$

Take conditional expectation on X_{i+1} given X_i (this means in round $i+1$, X_i is an observed deterministic value), we get

$$\begin{aligned} \mathbb{E}\left[X_{i+1} \left| X_i\right.\right] &= X_i + 1 \cdot \mathbb{P}\left(new\, ball\, selected\right) \\ &= X_i + \frac{n - X_i}{n} \end{aligned}$$

Take expectation on X_i on both sides, we can get

$$\begin{aligned} \mathbb{E}\left[X_{i+1}\right] &= \mathbb{E}\left[\mathbb{E}\left[X_{i+1} \left| X_i\right.\right]\right] \\ &= \left(1 - \frac{1}{n}\right) \mathbb{E}\left[X_i\right] + 1 \end{aligned}$$

The above difference equation can be easily solved for $\mathbb{E}\left[X_m\right] = n\left(1 - \left(1 - \frac{1}{n}\right)^m\right)$.

Let's do a sanity check, i.e., when $m = 1$, we can get

$$\mathbb{E}\left[X_1\right] = 1$$

when $m = 2$, we know that

$$\begin{aligned} \mathbb{E}\left[X_2\right] &= 1 + 1 \times \mathbb{P}\left(new\, ball\, selected\right) \\ &= 1 + \frac{n-1}{n} = 2 - \frac{1}{n} \end{aligned}$$

when $m \to \infty$, we get

$$\lim_{m \to \infty} \mathbb{E}\left[X_m\right] = n\left(1 - 0\right) = n$$

which also makes great sense.

Approach Two: (Indicator Function Approach) At any time, the number of distinct balls

$$X = \sum_{i=1}^{n} \mathbb{I}_{\{i-th\, ball\, ever\, selected\}}$$

thus (notice that $\mathbb{E}[\mathbb{I}_A] = \mathbb{P}(A)$)

$$\mathbb{E}[X] = \sum_{i=1}^{n} \mathbb{P}(i - th\, ball\, ever\, selected)$$

For a m-round game,

$$\mathbb{E}[X_m] = \sum_{i=1}^{n} \mathbb{P}(i - th\, ball\, ever\, selected\, in\, this\, m - round\, game)$$

By symmetry, each ball is not selected in this game with probability

$$\mathbb{P}(not\, selected) = \left(\frac{n-1}{n}\right)^m$$

thus it is selected with probability $1 - \left(1 - \frac{1}{n}\right)^m$, hence

$$\mathbb{E}[X_m] = n \cdot \left(1 - \left(1 - \frac{1}{n}\right)^m\right)$$

Question 2: Of the $2n$ people in a given collection of n couples, exactly m people die. Assuming that the m have been picked at random, find the mean number of surviving couples. This problem was formulated by Daniel Bernoulli in 1768.

For every couple, the probability for them to survive is $\binom{2n-2}{m} / \binom{2n}{m}$, therefore the mean number of surviving couples is just $n\dfrac{\binom{2n-2}{m}}{\binom{2n}{m}}$.

Question 3: Urn R contains n red balls and urn B contains n blue balls. At each stage, a ball is selected at random from each urn, and they are swapped. Show that the mean number of red balls in urn R after stage k is $\frac{1}{2}n\left(1 + \left(1 - \frac{2}{n}\right)^k\right)$. This diffusion model was described by Daniel Bernoulli in 1769.

Approach One: Let X_k be the number of red balls in urn R after stage k. Then the difference relation can be established as follows:

$$X_{k+1} = X_k + Y_k$$

where

$$Y_k = \begin{cases} 1 & \mathbb{P}(blue \longleftrightarrow red) = \frac{n-X_k}{n} \cdot \frac{n-X_k}{n} \\ 0 & \mathbb{P}(blue \longleftrightarrow blue\, or\, red \longleftrightarrow red) = \frac{n-X_k}{n} \cdot \frac{X_k}{n} + \frac{X_k}{n} \cdot \frac{n-X_k}{n} \\ -1 & \mathbb{P}(red \longleftrightarrow blue) = \frac{X_k}{n} \cdot \frac{X_k}{n} \end{cases}$$

Take expectation on both sides, we can get

$$\mathbb{E}[X_{k+1}] = \mathbb{E}[\mathbb{E}[X_{k+1}|X_k]] = \left(1 - \frac{2}{n}\right)\mathbb{E}[X_k] + 1$$

Solve for $\mathbb{E}\left[X_k\right]$, we get

$$\mathbb{E}\left[X_k\right] = \frac{1}{2} n \left(1 + \left(1 - \frac{2}{n}\right)^k\right)$$

Approach Two: Direct calculation? This is not that so obvious? I am looking for you guy's comments.