RA2

david.pujolperich

October 2020

Problem 15

Problem statement

Let a_1, a_2, . . . , a_n a sequence of n integers without repetitions. We say that a_i and a_j are inverted when $i < j$ but $a_i > a_j$. The Bubble sort algorithm interchanges pairs of inverted elements until the sequence became sorted.

Assume that the input to Bubble sort is is a random permutation selected u.a.r and that X is a random variable counting the number of interchanges during the execution of the algorithm.

Find the expectation and the variance

$$X = \text{ of interchanges the sequence} \tag{1}$$

$$X_{i,j} = \begin{cases} 1, & \text{If pair } (a_i, a_j) \text{ are interchanged} \\ 0, & \text{o/w} \end{cases} \tag{2}$$

Then we know that:

$$X = \sum_{i=1}^{n-1} \sum_{j>i}^{n} X_{i,j} \tag{3}$$

And thus, the expectation is the following (since the permutation is chosen u.a.r, and X_i is a Bernoulli R.V):

$$\mathbb{E}[X] = \sum_{i=1}^{n-1} \sum_{j>i}^{n} \mathbb{E}[X_{i,j}] = \sum_{i=1}^{n-1} \sum_{j>i}^{n} Pr[X_{i,j} = 1] = \sum_{i=1}^{n-1} \sum_{j>i}^{n} \frac{1}{2} \tag{4}$$

Since this are all the possible pairs (without ordering), then

$$\mathbb{E}[X] = \binom{n}{2} \cdot \frac{1}{2} = \frac{n \cdot (n-1)}{4} \tag{5}$$

Find the variance

We know, by definition of variance, that:

$$Var[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 \tag{6}$$

$$Var[X] = \mathbb{E}[\sum_{1\leq i<j\leq n} X_{i,j} \cdot \sum_{1\leq k<l\leq n} X_{k,l}] - \mathbb{E}[\sum_{1\leq i<j\leq n} X_{i,j}] \cdot \mathbb{E}[\sum_{1\leq k<l\leq n} X_{k,l}] = \tag{7}$$

By linearity of expectation we get taht:

$$\sum_{\substack{1\leq i<j\leq n \\ 1\leq k<l\leq n}} \mathbb{E}[X_{i,j} \cdot X_{k,l}] - \mathbb{E}[X_{i,j}] \cdot \mathbb{E}[X_{k,l}] = \tag{8}$$

Since $X_{i,j}$ is a Bernoully R.V:

$$\sum_{\substack{1\leq i<j\leq n \\ 1\leq k<l\leq n}} Pr[X_{i,j} = 1 \wedge X_{k,l} = 1] - \frac{1}{2} \cdot \frac{1}{2} = \tag{9}$$

$$\sum_{\substack{1\leq i<j\leq n \\ 1\leq k<l\leq n}} Pr[X_{i,j} = 1 \wedge X_{k,l} = 1] - \frac{1}{4} \tag{10}$$

We know that the permutation was chosen u.a.r:
Case 1: i==k and j==l (same pair)

$$Pr[X_{i,j} = 1 \wedge X_{k,l} = 1] = Pr[X_{i,j} = 1] = \frac{1}{2} \tag{11}$$

Case 2: $i == k$ and $j \neq l$ (three distinct points/value)

$$Pr[X_{i,j} = 1 \wedge X_{i,l} = 1] = Pr[X_{i,j} = 1] \cdot Pr[X_{i,l} = 1] = \frac{1}{4} \tag{12}$$

Case 3: Similarly, $i \neq k$ and $j == l$ (three distinct points/value)

$$Pr[X_{i,j} = 1 \wedge X_{i,l} = 1] = Pr[X_{i,j} = 1] \cdot Pr[X_{k,j} = 1] = \frac{1}{4} \tag{13}$$

Case 4: $i \neq k$ and $j \neq l$ (4 distinct values)

$$Pr[X_{i,j} = 1 \wedge X_{i,l} = 1] = Pr[X_{i,j} = 1] \cdot Pr[X_{k,l} = 1] = \frac{1}{4} \tag{14}$$

So basically, we can see that for all values i, j, k, l st. $|i, j, k, l| < 4$, the value cancels out. Hence, we can simplify the equation of the variance to simply the values where $i == k$ and $j == l$. So, to all the possible pairs:

$$\sum_{1\leq i\leq j\leq n} Pr[X_{i,j} = 1] - \frac{1}{4} = \sum_{1\leq i\leq j\leq n} \frac{1}{2} - \frac{1}{4} = \sum_{1\leq i\leq j\leq n} \frac{1}{4} \tag{15}$$

As seen before, this is that:

$$Var[X] = \binom{n}{2} \cdot \frac{1}{4} = \frac{n \cdot (n-1)}{8} \tag{16}$$

Is it concentrated around the mean?

For this exercise we can use Chebyshev, from which we know that:

$$Pr[|X - mu| \geq k] \leq \frac{Var[X]}{k^2} \tag{17}$$

If we plug in our value of the variance, we get:

$$Pr[|X - mu| \geq k] \leq \frac{\frac{n\cdot(n-1)}{8}}{k^2} = \frac{n^2 - n}{8 \cdot k^2} \tag{18}$$

So, for instance, if $k = n \cdot logn$, then we get that:

$$Pr[|X - mu| \geq n] \leq \frac{n^2 - n}{8 \cdot (n \cdot logn)^2} \tag{19}$$

This, as $n \to \infty$, tends to 0. So we can assure that with high probability absolute difference between X and $\mathbb{E}[X]$ is going to be smaller than $n \cdot log(n)$.