CSC484 Assignment #1

Andrii Osipa

January 2017

Problem *1.1*

Solution. Let X_i be random variable with the following definition:

$$X_i = \begin{cases} 1, \text{there is no balls in } i^{th} \text{ bin} \\ 0, \text{otherwise} \end{cases}$$

. Then $\sum\limits_{i=1}^{n} X_i$ is the number of empty bins. If we have no balls in i^{th} bin then on each ball it was put into any other bin and this event has probability $\frac{n-1}{n}$. We have m balls therefore $P(X_i = 1) = \left(\frac{n-1}{n}\right)^m$. So $E[X_i] = \left(\frac{n-1}{n}\right)^m$. Therefore $E[X] = \sum\limits_{i=1}^{n} E[X_i] = \sum\limits_{i=1}^{n} \left(\frac{n-1}{n}\right)^m = n\left(\frac{n-1}{n}\right)^m = \frac{(n-1)^m}{n^{m-1}}$.

Problem *1.2*

Solution. **Probability that i^{th} and j^{th} elements are compared:**
$X_{ij} = \begin{cases} 1, i^{th} \text{ and } j^{th} \text{ elements were compared} \\ 0, \text{otherwise} \end{cases}$.
Lets take a look at three cases:

1. $i < k' < j$: in this case two elements will be compared if one of them is chosen as pivot. Therefore $P(X_{ij} = 1) = \frac{2}{j-i+1}$.

2. $i < j < k'$: in this case we have the following: if pivot is between i^{th} and j^{th} elements it is obvious that they never will be compared as they will be in the different partitions.
If pivot is between j^{th} and k^{th} elements then they also will not be compared as after partition algorithm will not run Select(...) for subarray where both i^{th} and j^{th} elements are.
If pivot is k^{th} element then those two also will not be compared as algorithm will just return pivot.
If pivot is i^{th} or j^{th} element then they will be compared. Therefore
$$P(X_{ij} = 1) = \frac{2}{k-i+1}.$$

3. $k < i < j$: this case is similar to previous one. Here we have

$$P(X_{ij} = 1) = \frac{2}{j - k + 1}.$$

Then for total number of comparations we have

$$X = \sum_{i<j} X_{ij} = \sum_{k \le i < j} X_{ij} + \sum_{i<k<j} X_{ij} + \sum_{i<j\le k} X_{ij}.$$

Lets calculate each of the sums:

$$\sum_{k\le i<j} E[X_{ij}] = \sum_{i=k}^{n} \sum_{j=i+1}^{n} \frac{2}{j-k+1} = \sum_{j=k+1}^{n} \sum_{i=k}^{j-1} \frac{2}{j-k+1} = \sum_{j=k+1}^{n} \frac{2(j-k)}{j-k+1} =$$

$$= \sum_{j=1}^{n-k} \frac{2j}{j+1} = \sum_{j=1}^{n-k} \left(2 - 2\frac{1}{j+1}\right) = 2(n-k+1) - 2\sum_{j=1}^{n-k+1} \frac{1}{j} = O(n) - O(\ln n) = O(n)$$

$\sum\limits_{i<j\le k} E[X_{ij}] = O(n)$ and proof is very similar to the previous one.

$$\sum_{i<k<j} E[X_{ij}] = \sum_{i=1}^{k-1} \sum_{j=k+1}^{n} \frac{2}{j-i+1} = \sum_{i=1}^{k-1} \sum_{j=k+1-i}^{n-i} \frac{2}{j+1} \le \sum_{i=1}^{k-1} \sum_{j=2}^{n-1} \frac{2}{j+1} \le$$

$$\le (k-1)\ln(n-1) = O(\ln n).$$

Therefore for X we have $E[X] = O(n) + O(n) + O(\ln n) = O(n)$.

Problem *1.3*

Solution. Let A_i be event that i was chosen during step of updating X.

$$A_i = \begin{cases} \text{1,i was selected on some step} \\ 0, \text{otherwise} \end{cases}$$

It is easy to see that each value for X can be selected only once. If some $k \in \{0, ..., n-1\}$ was selected then on every following step randomly selected number will be smaller than k. We have i from 0 to $n-1$. Number of steps in our algorithm is 1 plus number of updates of X. $A = \sum\limits_{i=0}^{n-1} A_i$ is total number of updates of X.
$P(A_i = 1) = \dfrac{1}{i+1}$: if was selected any number smaller that i then i will never be selected. Therefore $E[A_i] = \dfrac{1}{i+1}$. And $E[A] = \sum\limits_{i=0}^{n-1} E[A_i] = \sum\limits_{i=0}^{n-1} \dfrac{1}{i+1} = O(\ln n)$.

Problem *1.4. Bonus.*

Solution. Let A_i be event that i was chosen during step of updating X.

$$A_i = \begin{cases} 1\text{,i was selected on some step} \\ 0, \text{otherwise} \end{cases}$$

Each X can be chosen many times during algorithm, therefore $\sum\limits_{i=0}^{\infty} E[A_i]$ is much less then algorithm runtime. The only possibility that some specific value of X was not chosen at all means that 0 was chosen. Because in all other cases it is still possibility to select X in further steps. Therefore $P(A_i = 1) = \dfrac{i}{i+1}$. Then $E[A] = \sum\limits_{i=1}^{\infty} E[A_i] = \sum\limits_{i=1}^{\infty} \dfrac{i}{i+1} = \infty <$ runtime.

Problem *1.5*

Solution. $a_0, ..., a_{k-1} \in \{0, ..., p-1\}$
$X_g = a_0 + a_1 g + a_2 g^2 + ... + a_{k-1} g^{k-1} \mod p, g$ from 0 to $n-1$.
$P(X_0 = t_0, ..., X_{n-1} = t_{n-1}) = P(X_0 = t_0)...P(X_{n-1} = t_{n-1})$?
Obvious fact that $P(X_0 = t_0) = P(a_0 = t_0) = \dfrac{1}{p}$. For any $i > 0$ also holds that $P(X_i = t_i) = \dfrac{1}{p}$. We have $X_i = a_0 + a_1 i + a_2 i^2 + ... + a_{k-1} i^{k-1} \mod p$. Suppose $a_0, ..., a_{r-1}, a_{r+1} ..., a_{k-1}$ are fixed. Then only for $a_r * r^i = t_i - a_1 i - a_2 i^2 - ... - a_{k-1} i^{k-1} \mod p$ we have that $X_i = t_i$. And here exists only one solution for a_r.
Proof: suppose there are no solution $\Leftrightarrow \exists l \in \{0, ..., p-1\} : x r^i \neq l \mod p \Rightarrow \exists y : x r^i = y r^i \mod p \Rightarrow (x-y) r^i = 0 \mod p$ and p is not divisible by any $r : r \neq 1$ and we have that $r \leq p$ therefore $x = y$ and we have contradiction. Now we proved that solution exists. From same proof it easy to see that solution is the only one.
Therefore same fact holds for any index: we can have randomly selected $k-2$ indexes and the last one can be picked in the only way that $X_i = t_i$ holds. So we have $P(X_i = t_i) = \dfrac{1}{p}$.
Now suppose $X_0 = t_0, ..., X_{n-1} = t_{n-1}$.

$$\begin{bmatrix} 1 & 0 & 0^2 & \dots & 0^{k-1} \\ 1 & 1 & 1^2 & \dots & 1^{k-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & n-1 & (n-1)^2 & \dots & (n-1)^{k-1} \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_{k-1} \end{bmatrix} = \begin{bmatrix} X_0 \\ X_1 \\ \vdots \\ X_{n-1} \end{bmatrix} = \begin{bmatrix} t_0 \\ t_1 \\ \vdots \\ t_{n-1} \end{bmatrix}$$

Lets look at this as a system for $a_0, ..., a_{k-1}$.
First matrix is Vandermonde matrix. Determinant of this matrix nonzero as every row contains powers of different numbers from 0 to $n-1$ and if $\exists k, l \in \{0, ..., n-1\} : k = l \mod p \Rightarrow k = l$ as both k and l are less then p.

Det is nonzero $\Rightarrow$ system is defined and has only one solution $a'_0, ..., a'_{k-1}$. Therefore $X_0 = t_0, ..., X_{n-1} - t_{n-1} \Leftrightarrow a_0 = a'_0, ..., a_{k-1} = a'_{k-1}$. And $P(a_0 = a'_0, ..., a_{k-1} = a'_{k-1}) = \left(\frac{1}{p}\right)^k$.

As we showed before $\prod_{i=0} n - 1P(X_i = t_i) = \left(\frac{1}{p}\right)^k$. Therefore $X_0, ..., X_{n-1}$ are k-wise independent.

Problem *2.1*

Solution. $A, B \subseteq U$, $|A| = \Theta(n), |B| = \Theta(n)$ and $A \cap B = \emptyset$. Let's denote by c_A and c_B constants such that $|A| \leq c_A n$ and $|B| \leq c_B n$. $P(x \in R) = \Theta(\frac{1}{n})$ and let c_R be constant s.t. $P(x \in R) \geq \frac{c_R}{n}$.
$P(A \cap R = \emptyset \wedge B \cap R = \emptyset) = \prod_{x \in A \cup B} P(x \notin R) = \prod_{x \in A \cup B} (1 - P(x \in R)) \geq \prod_{x \in A \cup B} \left(1 - \frac{c_R}{n}\right) = \left(1 - \frac{c_R}{n}\right)^{|A|+|B|} \geq \left(1 - \frac{c_R}{n}\right)^{c_A n + c_B n} = \left(1 - \frac{c_R}{n}\right)^{n(c_A + c_B)}$.
The latter is strictly decreasing with the following property, known from calculus:

$$\lim_{n \to \infty} \left(1 - \frac{c_R}{n}\right)^{n(c_A + c_B)} = e^{-c_R(c_A + c_B)} =: c.$$

Therefore we showed that exists some constant $c > 0$ which is lower bound for $P(A \cap R = \emptyset \wedge B \cap R = \emptyset)$.

Problem *2.2* *$X_1, X_2, ...$ random variables. $P(X_i = 1) = p$ and $P(X_i = -1) = 1 - p$. T is smallest t s.t. $\sum_{i=1}^{t} X_i < 0$. $E[T] =$?*

Solution. It is easy to see that T can not be even. Suppose $T = 2k$ for some k. $\sum_{i=1}^{T} X_i < 0$ therefore there must be at least $k + 1$ negative ones and k positive ones and so $\sum_{i=1}^{T} X_i = -2m$ for some m. If X_T is -1 then $\sum_{i=1}^{T-1} X_i = -2m + 1 < 0$ therefore T is not the smallest one. If X_T is 1 then $\sum_{i=1}^{T-1} X_i = -2m - 1 < 0$ therefore T is not the smallest one. This gives us next fact: $P(T = 2k) = 0$.
For odd Ts we have: obviously, that X_T is -1, otherwise T is not the smallest. Therefore $\sum_{i=1}^{T-1} X_i = 0$, otherwise T also not the smallest one.
$E[T] = \sum_{k=1}^{\infty} c_k p^{k-1}(1 - p)^k$, where c_k is number of sequences of 1 and -1 of length $2k - 1$ and such that at each point $< 2k - 1$ sum of sequence is ≥ 0 and total sum is -1. $c_1 = 1$, $c_2 = 1$, $c_3 = 2$, $c_4 = 5$, etc. $c_k =$?. ...