590D-homework 1

Vinayak Mathur and Krishna Prasad Sankaranarayanan

September 2017

1 Question 1

We have to find the expected value of the number of times the word 'proof' appears.

The probability of the letters 'p' 'r' 'o' 'o' 'f' appearing together = $(\frac{1}{26})^5$
Let us define a random variable X = Probability that the word 'proof' starts at that index

So the expectation is defined as $\sum a * P(X = a)$

In this case: $E(X) = \sum_1^{(n-4)} 1 * (\frac{1}{26})^5$ We sum only till $(n - 4)$ because 'proof' can only start till that index

Therefore the expected value = $(n - 4) * (\frac{1}{26})^5$
The expected value of number of occurrences of proof = 0.0841649

2 Question 2

We define an indicator random variable such that
$X_i = 1$ if the i^{th} digit is a fixed point and
$X_i = 0$ otherwise.
The expectation that the i^{th} element is in its place i.e. it is a fixed element is $Ex_i = \frac{1}{n}$
Using linearity of expectations
$E[X] = E[x_i] + E[x_2] \ldots E[x_n]$
$= \frac{1}{n} + \frac{1}{n} \ldots + \frac{1}{n}$
$= n * \frac{1}{n} = 1$

3 Question 3

3.1

We define a random variable in the following manner:
$X = +1$ if the coin returns Heads, and
$X = -1$ if the result is tails

$E[X] = \sum_1^{100} +1 * \frac{1}{2} = 100 * \frac{1}{2} = 50$
Therefore the expected payoff is 50

3.2

For a biased coin:

$E[X] = \sum_1^{100} +1 * 0.3$
$= 100 * 0.3 = 30$
Therefore the expected payoff is 30

3.3

Using Markov Inequality:
$P(X \geq t) \leq \frac{E[X]}{t}$
We know for our friend to win the expectation is 70, we substitute that in the inequality which results in:
$P(X \geq 50) \leq \frac{70}{50}$
Therefore the required upper bound is $\frac{7}{5}$

4 Question 4

Consider X to be sum of rolls of dice.
X_i - number on the i^{th} roll of the dice
$X = \sum_{i=1}^{100} X_i$

By linearity of expectation , $E[x] = \sum_{i=1}^{100} E[X_i]$

$E[X_i] = \sum_{k=1}^{6} kP[X_i = l] = \frac{1}{6} * \frac{42}{2} = \frac{7}{2}$

Therefore $E[X] = n * \frac{7}{2} = 100 * \frac{7}{2} = 350$
$Var(X) = \sum_{i=1}^{100} Var(X_i)$
$Var(X_i) = E[{X_i}^2] - E[X]^2$
$E[{X_i}^2] = \sum_{k=1}^{6} k^2 P[X_i = k]$
$E[{X_i}^2] = \sum_{k=1}^{6} k^2 \frac{1}{6}$
$E[{X_i}^2] = \frac{13*42}{36}$
$E[{X_i}^2] = \frac{91}{6}$
Now, $Var(X_i) = E[{X_i}^2] - E[X]^2$
$Var(X_i) = \frac{91}{6} - (\frac{7}{2})^2 = \frac{35}{12}$
Therefore , According to Chebyshev's inequality
$P[|x - 350| \geq 50] \leq \frac{100*35}{12*2500} = \frac{7}{60}$

5 Question 5

5.1

Probability of a particular ball going into a bin = $\frac{1}{n}$ The required expectation is given by:
$E[B_i] = \sum_1^m \frac{1}{n}$
$= m * \frac{1}{n}$
$E[B_i] = \frac{m}{n}$

$m = 100n \ln n$
Applying Chernoff bound:
We define an indicator random variable X
$P[E[X] - X \geq 25 \ln n] \leq e^{-E[x]\frac{\delta^2}{3}}$
Substituting the value of δ we get
$\leq e^{-E[x]\frac{1}{48}}$
$\leq e^{-100 \ln n \frac{1}{48}}$
$\leq e^{-2 \ln n}$
$\leq \frac{1}{n^2}$

Therefore the required probability is given by:
P[Number of balls does not differ by 25ln(n)] $\leq 1 - \frac{1}{n^2}$

5.2

Here $m = n \ln n$
Show that all bins lie in the range of $\frac{n}{n} \pm \sqrt{\frac{m}{n}} * \ln n$
$\frac{n}{n} \pm \sqrt{\frac{m}{n} * \ln n} \approx \ln n \pm \ln n$

$\leq 2e^{-E[X]*\frac{1}{3}}$
$\leq \frac{2}{n^{\frac{1}{3}}}$

Hence $\delta = 1$ in Chernoff bound

$P[|X - E[X]|] \geq \delta E[X]]$ =P[all bins are out of this range in terms of balls]

$\leq \frac{2}{n^1/3}$

For n balls it is $\leq 2n^{2/3}$

So Answer is $1 - 2n^{2/3} \approx 1 - 1/n$

6 Question 6

Consider an indicator random variable X_i
$X_i = 1$ if 1st item is stored on the i^{th} run of the algorithm.

Algorithm is run t times.
Therefore $X = \sum_{i=1}^{t} X_i$
$E[X] = \sum_{i=1}^{t} E[X_i = \frac{t}{100}$
Given in the question , each item is selected in the range $\frac{t}{100}(1 \pm \frac{1}{3})$
Therefore $P(|X - E[X]| \geq \frac{t}{300}) \leq 2e^{-E[X]\frac{\delta^2}{3}}$
$\delta = \frac{1}{3}$
$P(|X - E[X]| \geq \frac{t}{300}) \leq 2e^{-\frac{t}{100*27}}$
$P(|X - E[X]| \geq \frac{t}{300}) \leq 2e^{-\frac{t}{2700}} = 0.99$
$e^{-\frac{t}{2700}} = 0.495$
$\frac{-t}{2700} = ln(0.495)$
$t = 1898.63 \approx 1899$ runs of the algorithm

7 Extra Credit

Let $1 + \delta = \frac{\ln n}{\ln \ln n}$
Consider $m = n$
$P[X \geq (1 + \delta)] \leq \frac{(e)^{\delta}}{(1+\delta)^{1+\delta})}^{E[X]}$

Expectation $E[X] = 1$
Since m=n

$P[anybins \geq \frac{\ln n}{\ln \ln n}] \leq n * (\frac{e \ln ln(n)}{\ln n})^{\frac{\ln n}{\ln \ln n}}$

$\leq n * [e^{\frac{\ln n}{\ln \ln n}} + e^{\frac{\ln n}{\ln \ln n} * \ln \frac{\ln \ln n}{\ln n}}]$