Collaborators : Anmol Ahuja, Vinod Dubey

1: Balls and Bins...

(a) Lets suppose that according to the required scenario in the Question, There are n bins present, considering a case when the bin is not empty, probability of a particular ball going inside a bin is $\frac{1}{n}$.
Given probability that the bin is not containing any ball or the bin is empty is $1 - \frac{1}{n}$.
According to our requirements, when all the m balls are thrown a particular bin should be empty.
Thus, probability for this scenario is $(1 - \frac{1}{n})^m$.
As we know, we are having no. of such kind of bins present = n, thus are going to take in account the Union Bound. According to union bounds it defines that " probability of at least 1 bin being empty is not more than the addition of probabilities of single bins being empty. "
Now taking into account all the n bins that are present, we can conclude that the probability of one bin being empty is $\leq n * (1 - \frac{1}{n})^m$.
$m = 4nlogn$
$\leq n * (1 - \frac{1}{n})^{(4nlogn)}$
As we already know that $(1 - \frac{1}{n})^n$ is equal to $\frac{1}{e}$
$\leq n * (\frac{1}{e^{(4logn)}})$
$\leq n * (\frac{1}{e^{(logn^4)}})$
$\leq n * \frac{1}{n^4}$
$\leq \frac{1}{n^3}$
As we can say, $n > 1$ that is the number of bins present is more than single unit.
This concludes $\frac{1}{n^3} < \frac{1}{n}$
Therefore it proves the required statement.

(b) Extending the previous conclusions we can substitute values of the two parts in above mentioned equation (as present in the initial or 1st part):
Using the values $(1 - 1/n)^{100nlogn}$
$= e^{-100logn}$
$=((1/e)^{logn^{1/100}}$
Extending our approach, further calculations gives values as:
$=(1/n^{100})$ Similar procedure for 2nd part, substituting it:
$(1 - 1/n)^{1/2nlogn}$
$= e^{-1/2logn}$
$=(1/e)^{logn^{1/2}}$
After applying further steps we get the part as:
$= 1/\sqrt{n}$

Therefore proved.

(c) Let us assume that a random variable exists, a_i. As seen in the above concluded result that probability for i^{th} bin to NULL or not containing any ball is given by:
$(1 - 1/n)^m \leq 1/e$.
As we can see for this particular scenario:

$Pr(a_i) \approx 1/e$.
Thus, we can say that expected E that a_i event will take place is:
$E(a) = \sum_{i=1}^{n} a_i * P(a_i)$
$= n * 1/e$
Thus the probability that we can get for a total of 9/10 ratio of bin to be present as empty or NULL is:
$P(a > 9/10) = E(a)/\frac{9}{10}$
After substitution of values of $E(a)$, the final answer we can obtain is:
$= \frac{n/e}{9/10}$

(d) We know according to the Bayes theorem, $P(A|B) = \frac{P(A \cap B)}{P(B)}$

1st step calculation: $P(X_{j1} = 1, X_{j2} = 1, \ldots, X_{jk} = 1)$ We here need to put n balls as such k bins will remian empty. Prob that we put ball in any one of the present k bins is k/n. So probability that ball will no go inside, any k bins is $(1 - \frac{k}{n})$. At this point we nee dto put in n such balls. Thus probab that k of those bins will remain empty is $(1 - \frac{k}{n})^n$.

Computing further:
$P(X_{j2} = 1, \ldots, X_{jk} = 1) = (\frac{(n-(k-1)}{n})^n$

According to bayes rule:

$P(X_{j1} = 1|X_{j2} = 1, \ldots, X_{jk} = 1) = P(X_{j1} = 1, X_{j2} = 1, \ldots, X_{jk} = 1)$ / $P(X_{j2} = 1, \ldots, X_{jk} = 1)$
$= \frac{(1-\frac{k}{n})^n}{(\frac{(n-(k-1)}{n})^n} = (1 - \frac{1}{n-(k-1)})^n$

So now $P(X_{j1} = 1) = (1 - 1/n)^n$ [because is we consider every ball Prob of it going inside j1 is $1/n$ and thus prob of not inside into j1 is $(1 - 1/n)$ there = n such balls]

Thus, $1/n \leq 1/(n - (k - 1))$ so $P(X_{j1} = 1|X_{j2} = 1, \ldots, X_{jk} = 1) \leq P[X_{j1} = 1]$

Thus = $P(X_{j1} = 1, X_{j2} = 1, \ldots, X_{jk} = 1) =$
$P(X_{j1} = 1|X_{j2} = 1, \ldots, X_{jk} = 1)P(X_{j2} = 1, \ldots, X_{jk} = 1)$

$\leq P(X_{j1} = 1)P(X_{j2} = 1, \ldots, X_{jk} = 1)$

Mentioning this recursively and maintaining the inequality as:
$\leq P(X_{j1} = 1)P(X_{j2} = 1) \ldots P(X_{jk} = 1)$

We even see, $P(X_{j1} = 1) = P(X_{j2} = 1) = \cdots = P(X_{jk} = 1) = (1 - 1/n)^n$

Thus this effects the inequality as, $P(X_{j1} = 1, X_{j2} = 1, \ldots, X_{jk} = 1) \leq ((1 - 1/n)^n)^k$

$P(X_{j1} = 1, X_{j2} = 1, \ldots, X_{jk} = 1) \leq e^{-k}$ ..... (a)

After this step, we just have to find the prob such that, 90% of bin remain empty.
$P(90\%\ of\ the\ bins\ remain\ empty) = P(\cup P(X_{j1} = 1, X_{j2} = 1, \ldots, X_{j0.9n} = 1)$ [this expression shows that to cal the probab such that 90% of the bins will remain empty we just need to show the union over all the possible combination of 90% bins remaining empty]

Union Bound $P(90\%\ of\ the\ bins\ remain\ empty) \leq \sum P(X_{j1} = 1, X_{j2} = 1, \ldots, X_{j0.9n} = 1)$

[summation over all possible combination of 90% bins]

Now no of ways to choose 90% bins from n bins is $\binom{n}{0.9n}$ & equation (a) above we get:

$P(90\%\ of\ the\ bins\ are\ empty) \leq \binom{n}{0.9n} e^{-0.9n}$ [now using binomial coffient is less the total ways which is 2^n]

$P(90\%\ of\ the\ bins\ are\ empty) \leq 2^n e^{-0.9n}$

$P(90\%\ of\ the\ bins\ are\ empty) \leq (\frac{2}{e^{0.9}})^n$ and $(\frac{2}{e^{0.9}}) \leq (0.9)^n$

So $P(90\%\ of\ the\ bins\ are\ empty) \leq (0.9)^n$

2: Estimating Mean and Median

(a) Let us suppose that $\hat{\mu}$ = mean for the given samples series.
Let us also consider that μ = actual mean present for the distribution.
Due to the symmetricity of the range that is provided, we can observe that μ =0.
Variance value is established to be $E((X - E(X))^2)$.
As we observe, E(x) = zero and boundary values that may be existing for X - E(X) are -1 & 1.
Here after taking square for these values the max value will remain as +1 for all X.
Therefore $E((X - E(X))^2)$ will automatically get reduced to n. We can say that this is the maximum devaiation that may exist for this particular sequence.
According to the definition of Chernoff's equation:
$P(|X| \geq k * \sigma) \leq 2 * e^{\frac{-k^2}{4*n}}$.
$P((\hat{\mu} - \mu) > \epsilon)$ = the term we are looking for.
Let's represent ϵ as a multiple values of σ in which σ can be considered as the SD or Standard Devaiation of the whole series.
Therefore $\epsilon = k * \sigma$.
$P((\hat{\mu} - \mu) > \epsilon) = P((\hat{\mu} - \mu) > k * \sigma)$
$= P((\hat{\mu} - \mu) > k * \sigma) = 2 * e^{\frac{-k^2}{4*n}}$
$= P((\hat{\mu} - \mu) > \epsilon) = 2 * e^{\frac{-\epsilon^2}{\sigma^2 * 4 * n}}$. Here we see that $variance_{max}$ is n as that was the case in previous proof.
Therefore
$= P((\hat{\mu} - \mu) > \epsilon) \leq 2 * e^{\frac{-\epsilon^2}{4*n^2}}$
Lets say that $P((\hat{\mu} - \mu) > \epsilon)$ as δ.
Here again $P((\hat{\mu} - \mu) \leq \epsilon)$ = 1 - $P((\hat{\mu} - \mu) > \epsilon)$
$= P((\hat{\mu} - \mu) \leq \epsilon) > 1 - \delta$
So further extending this δ we will try to find the desired probability values.
Therefore $\delta \leq 2 * e^{\frac{-\epsilon^2}{4*n^2}}$
In the previous equation, if we take logarithm on two sides
$log\delta \leq log(2 * e^{\frac{-\epsilon^2}{4*n^2}})$
$\leq log(2) + log(e^{-}(\frac{\epsilon}{2*n})^2)$
$\leq log(2) - (\frac{\epsilon}{2*n})^2$
$=> log\delta \leq log(2) - (\frac{\epsilon}{2*n})^2$
$= (\frac{\epsilon}{2*n})^2 \leq log(2) - log\delta$
$\frac{1}{n^2} \leq (\frac{2}{\epsilon})^2 * (log(2) - log(\delta))$
$n \geq \frac{\epsilon}{2*\sqrt{log(2) - log(\delta)}}$
$n \geq \frac{\epsilon}{2*\sqrt{log(\frac{2}{\delta})}}$
Thus as we can observe this will give the desired no. of samples.

(b) The above mentioned proof by me wont work if we do sampling without replacement. In case where sampling is done without replacement, the events rely on each other. They are no longer independent in nature. We can say that sampling of a event at a particular time t, is eventually dependent on the sampling of th events before that particular time t, and no similar element should have been sampled in the previous time.
In case where th events became dependent, binomial distribution will not be present. And as we know Markov's, Cernoff's Markov's and Chebyshev's inequalities only work on binomial distribution.

(c) Lets assume that we are dividing the full dataset by M. Eventually the range of the particular series will get reduced from the value of -1 to +1.
Now, we can observe that the question has got a similar case as that of part a above.
If we consider μ to be the mean of the previous series, the mean that will be present for th emodified series will became $\frac{\mu}{M}$. However, we see that in part a the modified series had answer as zero.
If σ was the variance for the previous series, the new variance for modified series will be $\frac{\sigma^2}{M^2}$. As we know the maximum variation of new series is n, maximum variation of previou series must be $M^2 * n^2$.
$P((\hat{\mu} - \mu) > \epsilon) = 2 * e^{\frac{-\epsilon^2}{\sigma^2 * M^2 * 4 * n}}$
If we try to solve this in part a, the outcome will contain the following values:
$n \geq \frac{\epsilon}{2 * M * \sqrt{log(2) - log(\delta)}}$
$n \geq \frac{\epsilon}{2 * M * \sqrt{log(\frac{2}{\delta})}}$

(d) Let us suppose that n=7 and set for a_i may be A=(0,0,0,3/4,1,1,1). Thus median here is 3/4. If sample used is even, it will became impossible to get the estimation for true median. So we can say that closest an even sampling will get to the true median will be 7/8, in this case sample will be same to s = (0,3/4,1,1). Thus avg of 3/4 and 1. Thus 7/8 - 3/4 is the smallest error that is possible.
If we take odd sample then only 3 possible samples more than one that will be able to predict the true median. The prob b that sample chosen that will be able to detect the true median can be estimated by applying this formula: $p = \frac{\frac{k}{m}\frac{n-k-1}{m}}{\frac{n}{2m+1}}$ This probab is very small so there is no means to find median based on sample with e <1/8.

3: Quick sort with Optimal Comparisons

(a) We can say that A is that particular set from where we are sampling elements, and M can be regarded as the sampled elements set.
Lets assume that a particular element a_i is sampled from the set A. We can even assume that the sampled a_i is the k^{th} minimum element of the sample set A.
Lets consider a_i to be pivot, so for this case it must be median of set M. In the same manner, if a_i has to be median of M, there must be present m samples in M such that it is lower than value of a_i and similarly m samples in M such that it is more than a_i.
We can even say that smapled a_i is the k^{th} smallest element of A. Therefore, if we want the median to be sampled a_i we need to select m elements from k-1 small set of elements of A. The no. of ways to do this is :C_m^{k-1}.
Therefore if we want sampled a_i to be median of M, we must select m elements from n-k greater

elements of A, this can be accomplished in C_m^{n-k} ways.
Therefore probability that sampled a_i will only be pivot is:
$= \frac{C_m^{k-1} * C_m^{n-k}}{Total-no-of-ways-in-which-n-elements-can-be-sampled-from-2m+1-elements-of-set-A}$
$= \frac{C_m^{k-1} * C_m^{n-k}}{C_n^{2m+1}}$

(b) After getting a pivot element, we will make $n-1$ comparisons to break or divide the whole array in 2 parts. The broken part will be dependent on element k. For any k, we can with probab p_k say that the broken part will became recurrent $T(k-1)$ & $T(n-k)$. We can represent it as:

$$Comparisons : (n-1) + \sum_{k=1}^{n} p_k(T(k-1) + T(n-k))$$

.

4: Randomized Min-Cut

(a) As we know by Karger's algorithm, edge lying in between two different vertices is collapsed as such that that edge is not in minimum cut.
In the other case Karger's algo will not be successful in returning the proper min cut.
Thus, we can conclude that collapsing that particular edge that is not present in min cut can help in balancing or maintaining min cut value.
Therefore, the partciular size of the min cut in the generated graph will be same as the G.

(b) Here we suppose that G is a graph.
No. of vertices present in $G = V$
No of edges present in $G = E$
No. of vertices present in $G = n$
Thus we see that the sumof degree of n vertices $= 2|E|$
Now this further implies that avg. degree of vertices is $2|E|/n$. A further point to be considered here is that the min degree of a vertex can max be $2|E|/n$. Now lets assume that there is a cut present in the G, in such a manner that the vertex having min degree will be present in S and rest $\overline{S}$. So we see that such kind of cut can have size at max $2|E|/n$ & if E' is 1 such kind of min cut, then in this case the size of min cut $|E'| \leq 2|E|/n$.

(c) Let us consider G to be a graph.
No. of vertices $= V$
No. of edges $= E$
We concluded in the above results that the no. of edges present in min cut is $\leq 2|E|/n$, in which E denotes the total no. of edges present in Graph.
P[randomly picked edge is in min cut] $< (2|E|/n)/|E|$
P[randomly picked edge is in min cut] $< 2/n$ In a case if we select an edge randomly and then also the min cut value is maintained, we can conclude that edge is not in min cut.
Thus,
P[randomly picked 1st edge is *not* in min cut] $\geq (1 - 2/n)$ Hence,
P[min cut is maintained] = P[randomly picked 1st edge is *not* in min cut]
$\geq (1 - 2/n)$

(d) Here we will try to find the probability for a specific min-cut that is $2/n^2$.

Each min-cut here as we see is not dependent on each other, so finding one min-cut does not have impact on probability of finding other. Therefore, the chances of finding every min cut, will add to one and there k will denote the no. of min cuts possible. Each probability is $\leq 2/n^2$, thus k times $2/n^2$ will became ≤ 1. Therefore, k would be $\leq n^2/2$.

5: Valiant Vazirani Lemma

Lets assume that there are just m-1 elements. So in this scenario, we can have 2 different cases:

For the first case:
As we can observe there is just 1 single unique element present in the set of $m-1$ elements.
We assume the least or minimum element to be a_j.
Now suppose for selecting the mth (a_m)element, the requiremnet is that it should not be same as a_j.
Thus we can see that the probability of selecting this particular element will be $(1-\frac{1}{N})$.

For the second case:
There are more than one minimum element present in $m-1$ elements.
Let us suppose that the 2 minimum elements present are a_i & a_j.
So in this case while we need to select the mth element, we can select the mth element in such a way that it can be considered as the new min element, for which the mth element will depict a value that is more than a_i & a_j and this particular mth element will be same as the a_i & a_jth element.
Thus we observe that this probability value is more than Zero, So we can ignore it.
P(m)= P(Case1) + P(Case2)
So this can be:
P(m)=P(m-1)P($a_j \neq a_m$) + P(Case2)
We observe that probability of choosing a element from a set of N elements will be $1/N$.
Probability that this particular value is not same as min element value is $(1-\frac{1}{N})$.
P($a_j \neq a_m$)= $(1-\frac{1}{N})$

As seen above for case 2 P(Case2) is more than zero, thus we can conclude:
$P(m) \geq P(m-1)(1-\frac{1}{N})$

$P(m) \geq P(m-2)(1-\frac{1}{N})(1-\frac{1}{N})$

After performing continuous or recursive iteration we get:
P(m)$\geq (1-\frac{1}{N})^{m-1}$

Thus, it proves the required statement.