1: Balls and Bins

$$\boxed{Markov's\ Inequality:\ Pr(X \geq a) \leq \frac{E[x]}{a}} \tag{1}$$

(a) Given n bins and $4n \log n$ balls, we want to prove that the probability that there exists an empty bin is $< 1/n$.

Proof We will prove this using Markov's Inequality. Let X be a random variable representing the number of empty bins. We will show that the probability that at least one bin is empty is $< 1/n$.

$$Pr(X \geq 1) \leq E[X]$$

From class, we saw the expectation of X is equivalent to the summation of each of the bins. Let y_i represent each bin where a 1 represents an empty bin and a 0 otherwise.

$$Pr(X \geq 1) \leq E[X] = E[\sum_{i=1}^{n} y_i]$$

The probability a bin is empty is given by $(1 - 1/n)^m$ where m is the number of balls. That is, there are $1 - 1/n$ other bins each ball can be placed in.

$$Pr(X \geq 1) \leq n * (1 - \frac{1}{n})^{4nlogn}$$

From Bernoulli's inequality we know that $1 + y \leq e^y$ for all for all y. Letting $y = \frac{-1}{n}$ will allow us to substitute $1 + y$ for e^y. This is valid as $1 + y \leq e^y$ so Markov's Inequality still holds.

$$Pr(X \geq 1) \leq n * (1 + y)^{4nlogn}$$
$$Pr(X \geq 1) \leq n * (e^y)^{4nlogn}$$
$$Pr(X \geq 1) \leq n * (e^{-1/n})^{4nlogn}$$

Applying exponent and natural log identities.

$$Pr(X \geq 1) \leq n * e^{-4logn}$$
$$Pr(X \geq 1) \leq \frac{1}{n^3} < \frac{1}{n}$$

Thus proving the probability that at least one bin is empty is $< 1/n$.

(b.a) When $m = \frac{1}{2}nlogn$, the logic follows as in part a.

$$Pr(X \geq 1) \leq E[X] = E[\sum_{i=1}^{n} y_i] = n * (1 - \frac{1}{n})^{\frac{1}{2}nlogn}$$

Where y_i is a bin and 1 represents an empty bin and 0 otherwise. Making the similar substitution and applying similar identities as part a.

$$Pr(X \geq 1) \leq n * (e^{-1/n})^{\frac{1}{2}nlogn}$$
$$Pr(X \geq 1) \leq n * \frac{1}{\sqrt{n}}$$
$$Pr(X \geq 1) \leq \sqrt{n}$$

Given $\frac{1}{2}nlogn$ balls, we can say that the probability that at least one bin is empty is $\leq \sqrt{n}$.

$\boxed{Pr(X \geq 1) \leq \sqrt{n}}$

(b.b) When $m = 100nlogn$, a similar logic follows.

$$Pr(X \geq 1) \leq E[X] = E[\sum_{i=1}^{n} y_i] = n * (1 - \frac{1}{n})^{100nlogn}$$
$$Pr(X \geq 1) \leq n * (e^{-1/n})^{100nlogn}$$
$$Pr(X \geq 1) \leq n * \frac{1}{n^{100}}$$
$$Pr(X \geq 1) \leq \frac{1}{n^{99}}$$

Given $100nlogn$ balls, we can say that the probability that at least one bin is empty is $\leq \frac{1}{n^{99}}$.

$\boxed{Pr(X \geq 1) \leq \frac{1}{n^{99}}}$

(c) Given n bins and n balls we want to bound the probability that 90% of the bins are empty. Using Markov's Inequality we will derive such a bound.

$$Pr(X \geq a) \leq \frac{E[X]}{a}$$
$$Pr(X \geq 0.9n) \leq \frac{E[X]}{0.9n}$$

Where X is a random variable representing the number of empty bins. From class we the expectation of X is equivalent to the summation of each of the bins. Let y_i represent each bin where a 1 represents an empty bin and a 0 otherwise.

$$Pr(X \geq 0.9n) \leq \frac{E[\sum_{i=1}^{n} y_i]}{0.9n}$$
$$Pr(X \geq 0.9n) \leq \frac{n * (1 - \frac{1}{n})^n}{0.9n}$$
$$Pr(X \geq 0.9n) \leq \frac{n * (e^{\frac{-1}{n}})^n}{0.9n}$$
$$Pr(X \geq 0.9n) \leq \frac{\frac{1}{e}}{0.9} = \frac{1}{0.9e}$$

Given n balls and n bins, we can say that the probability that 90% of the bins are empty is $\leq \frac{1}{0.9e}$.

$\boxed{Pr(X \geq 0.9n) \leq \frac{1}{0.9e}}$

(d)

2: Estimating the Mean and Median

As by the suggestion of the TA Michael Matteny I will be using Hoeffding's Inequality to provide bounds for the following questions rather than the suggested Chernoff bounds. They are related and provide similar strength in bounds as they both provide bounds for **independent** random variables. This Hoeffding's Inequality stronger than Markov's or Chebyshev's inequalities.

$$\boxed{Hoeffding's\ Inequality:\ P(|X - E[X]| \geq t) \leq 2exp(\frac{-2n^2t^2}{\sum_{i=1}^{n}(b_i - a_i)^2})} \tag{2}$$

(a) This variation of Hoeffding's Inequality is used when we know that X_i's are strictly in the intervals $[a_i, b_i]$. For our purposes a_i is -1 and b_i is 1. Let X be the sample mean after j random indices are sampled or $\hat{\mu}$. The expected value of X is the true mean or μ.

Hoeffding's Inequality tells us the probability that the difference between X and $E[X]$ will be $\geq t$ is less than the RHS of the inequality. That is, if t is ϵ we want the RHS to be $\leq \delta$. This will give us the number of required samples to ensure that $|\hat{\mu} - \mu| \leq \epsilon$ with probability $1 - \delta$, where n is the number of samples.

$$P(|\hat{\mu} - \mu| \geq \epsilon) \leq 2exp(\frac{-2n^2\epsilon^2}{\sum_{i=1}^{n}(1+1)^2}) \leq \delta$$
$$exp(\frac{-2n^2\epsilon^2}{4n}) \leq \frac{\delta}{2}$$
$$\frac{-n\epsilon^2}{2} \leq \ln(\frac{\delta}{2})$$
$$n\epsilon^2 \geq ln(\frac{4}{\delta^2})$$
$$n \geq ln(\frac{4}{\delta^2}) * \frac{1}{\epsilon^2}$$

Given a δ and ϵ, the derived equation produces the n number of indices we must sample to satisfy $|\hat{\mu} - \mu| \leq \epsilon$ with probability $1 - \delta$.

(b) No. If we were to sample without replacement, the sampled a_j's are no longer independent events. They become **dependent** on what has already been sampled as the sample space is changing with each sample. Hoeffding's Inequality is only applicable for **independent** random variables.

(c) Given the value of each a_i has the constraint $a_i \in [-M, M]$, we can still use Hoeffding's Inequality to derive the require number of samples. The setup will be similar to part a, with the exception that now $a = -M$ and $b = M$. Recall that the generalization of Hoeffding's Inequality is used when we know that X_i's are strictly bounded by the intervals $[a_i, b_i]$.

$$P(|\hat{\mu} - \mu| \geq \epsilon) \leq 2exp(\frac{-2n^2\epsilon^2}{\sum_{i=1}^{n}(M+M)^2}) \leq \delta$$

$$exp(\frac{-n\epsilon^2}{2M^2}) \leq \frac{\delta}{2}$$
$$\frac{-n\epsilon^2}{2M^2} \leq ln(\frac{\delta}{2})$$
$$\frac{n\epsilon^2}{M^2} \geq ln(\frac{4}{\delta^2})$$

(d)

3: Quick-sort with Optimal Comparisons

(a)

(b)

(c)

4: Randomized Min-Cut

(a)

(b)

(c)

(d)

5: Valiant-Vazirani Lemma