CSE 613: Parallel Programming

Lecture 6
(High Probability Bounds)

Rezaul A. Chowdhury

Department of Computer Science
SUNY Stony Brook
Spring 2012

Markov's Inequality

Theorem 1: Let X be a random variable that assumes only nonnegative values. Then for all $\delta > 0$,

$$\Pr[X \geq \delta] \leq \frac{E[X]}{\delta}.$$

Proof: For $\delta > 0$, let

$$I = \begin{cases} 1 & \text{if } X \geq \delta; \\ 0 & \text{otherwise.} \end{cases}$$

Since $X \geq 0$, $I \leq \frac{X}{\delta}$.

We also have, $E[I] = \Pr[I = 1] = \Pr[X \geq \delta]$.

Then $\Pr[X \geq \delta] = E[I] \leq E\left[\frac{X}{\delta}\right] \leq \frac{E[X]}{\delta}$.

Example: Coin Flipping

Let us bound the probability of obtaining more than $\frac{3n}{4}$ heads in a sequence of n fair coin flips.

Let

$$X_i = \begin{cases} 1 & \text{if the } i\text{th coin flip is heads;} \\ 0 & \text{otherwise.} \end{cases}$$

Then the number of heads in n flips, $X = \sum_{i=1}^{n} X_i$.

We know, $E[X_i] = \Pr[X_i = 1] = \frac{1}{2}$.

Hence, $E[X] = \sum_{i=1}^{n} E[X_i] = \frac{n}{2}$.

Then applying Markov's inequality,

$$\Pr\left[X \geq \frac{3n}{4}\right] \leq \frac{E[X]}{3n/4} = \frac{n/2}{3n/4} = \frac{2}{3}.$$

Chebyshev's Inequality

Theorem 2: For any $\delta > 0$,

$$\Pr[|X - E[X]| \geq \delta] \leq \frac{Var[X]}{\delta^2}.$$

Proof: Observe that $\Pr[|X - E[X]| \geq \delta] = \Pr[(X - E[X])^2 \geq \delta^2]$.

Since $(X - E[X])^2$ is a nonnegative random variable, we can use Markov's inequality,

$$\Pr[(X - E[X])^2 \geq \delta^2] \leq \frac{E\left[(X - E[X])^2\right]}{\delta^2} = \frac{Var[X]}{\delta^2}.$$

Example: n Fair Coin Flips

$$X_i = \begin{cases} 1 & \text{if the } i\text{th coin flip is heads;} \\ 0 & \text{otherwise.} \end{cases}$$

Then the number of heads in n flips, $X = \sum_{i=1}^{n} X_i$.

We know, $E[X_i] = \Pr[X_i = 1] = \frac{1}{2}$ and $E[(X_i)^2] = E[X_i] = \frac{1}{2}$.

Then $Var[X_i] = E[(X_i)^2] - (E[X_i])^2 = \frac{1}{2} - \frac{1}{4} = \frac{1}{4}$.

Hence, $E[X] = \sum_{i=1}^{n} E[X_i] = \frac{n}{2}$ and $Var[X] = \sum_{i=1}^{n} Var[X_i] = \frac{n}{4}$.

Then applying Chebyshev's inequality,

$$\Pr\left[X \geq \frac{3n}{4}\right] \leq \Pr\left[|X - E[X]| \geq \frac{n}{4}\right] \leq \frac{Var[X]}{(n/4)^2} = \frac{n/4}{(n/4)^2} = \frac{4}{n}.$$

Preparing for Chernoff Bounds

Lemma 1: Let $X_1, \ldots, X_n$ be independent Poisson trials, that is, each X_i is a 0-1 random variable with $\Pr[X_i = 1] = p_i$ for some p_i. Let $X = \sum_{i=1}^{n} X_i$ and $\mu = E[X]$. Then for any $t > 0$,

$$E[e^{tX}] \leq e^{(e^t-1)\mu}.$$

Proof: $E[e^{tX_i}] = p_i e^{t\times 1} + (1-p_i)e^{t\times 0} = p_i e^t + (1-p_i)$

$$= 1 + p_i(e^t - 1)$$

But for any y, $1 + y \leq e^y$. Hence, $E[e^{tX_i}] \leq e^{p_i(e^t-1)}$.

Now, $E[e^{tX}] = E\left[e^{t\sum_{i=1}^{n} X_i}\right] = E[\prod_{i=1}^{n} e^{tX_i}] = \prod_{i=1}^{n} E[e^{tX_i}]$

$$\leq \prod_{i=1}^{n} e^{p_i(e^t-1)} = e^{(e^t-1)\sum_{i=1}^{n} p_i}$$

But, $\mu = E[X] = E[\sum_{i=1}^{n} X_i] = \sum_{i=1}^{n} E[X_i] = \sum_{i=1}^{n} p_i$.

Hence, $E[e^{tX}] \leq e^{(e^t-1)\mu}$.

Chernoff Bound 1

Theorem 3: Let $X_1, \ldots, X_n$ be independent Poisson trials, that is, each X_i is a 0-1 random variable with $\Pr[X_i = 1] = p_i$ for some p_i. Let $X = \sum_{i=1}^{n} X_i$ and $\mu = E[X]$. Then for any $\delta > 0$,

$$\Pr[X \geq (1+\delta)\mu] \leq \left(\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}}\right)^{\mu}.$$

Proof: Applying Markov's inequality for any $t > 0$,

$$\Pr[X \geq (1+\delta)\mu] = \Pr\left[e^{tX} \geq e^{t(1+\delta)\mu}\right] \leq \frac{E\left[e^{tX}\right]}{e^{t(1+\delta)\mu}}$$

$$\leq \frac{e^{(e^t-1)\mu}}{e^{t(1+\delta)\mu}} \quad \text{[Lemma 1]}$$

Setting $t = \ln(1+\delta) > 0$, i.e., $e^t = 1+\delta$, we get,

$$\Pr[X \geq (1+\delta)\mu] \leq \left(\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}}\right)^{\mu}.$$

Chernoff Bound 2

Theorem 4: For $0 < \delta < 1$, $\Pr[X \geq (1+\delta)\mu] \leq e^{-\frac{\mu\delta^2}{3}}$.

Proof: From Theorem 3, for $\delta > 0$, $\Pr[X \geq (1+\delta)\mu] \leq \left(\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}}\right)^{\mu}$.

We will show that for $0 < \delta < 1$, $\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}} \leq e^{-\frac{\delta^2}{3}}$

$$\Rightarrow \delta - (1+\delta)\ln(1+\delta) \leq -\frac{\delta^2}{3}$$

That is, $f(\delta) = \delta - (1+\delta)\ln(1+\delta) + \frac{\delta^2}{3} \leq 0$

We have, $f'(\delta) = -\ln(1+\delta) + \frac{2}{3}\delta$, and $f''(\delta) = -\frac{1}{1+\delta} + \frac{2}{3}$

Observe that $f''(\delta) < 0$ for $0 \leq \delta \leq \frac{1}{2}$, and $f''(\delta) > 0$ for $\delta > \frac{1}{2}$.

Chernoff Bound 2

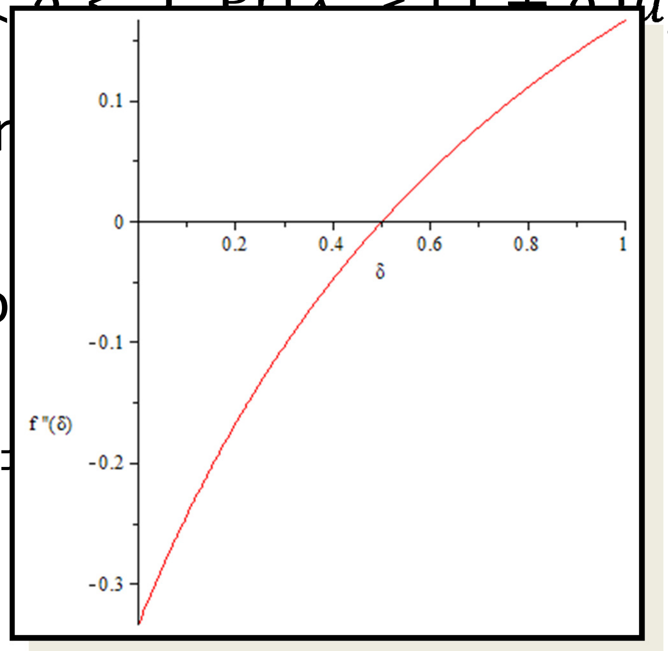
Theorem 4: For $0 < \delta < 1$, $\Pr[X \geq (1 + \delta)\mu] \leq e^{-\frac{\mu\delta^2}{3}}$.

Proof: From Theorem $+ \delta)\mu] \leq \left(\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}}\right)^{\mu}$.

We will show that fo $e^{-\frac{\delta^2}{3}}$

$\leq -\frac{\delta^2}{3}$

That is, $f(\delta) = \delta -$ 0



We have, $f'(\delta) = -\ln(1 + \delta) + \frac{2}{3}\delta$, and $f''(\delta) = -\frac{1}{1+\delta} + \frac{2}{3}$

Observe that $f''(\delta) < 0$ for $0 \leq \delta \leq \frac{1}{2}$, and $f''(\delta) > 0$ for $\delta > \frac{1}{2}$.

Chernoff Bound 2

Theorem 4: For $0 < \delta < 1$, $\Pr[X \geq (1+\delta)\mu] \leq e^{-\frac{\mu\delta^2}{3}}$.

Proof: From Theorem 3, for $\delta > 0$, $\Pr[X \geq (1+\delta)\mu] \leq \left(\frac{e^\delta}{(1+\delta)^{(1+\delta)}}\right)^\mu$.

We will show that for $0 < \delta < 1$, $\frac{e^\delta}{(1+\delta)^{(1+\delta)}} \leq e^{-\frac{\delta^2}{3}}$

$$\Rightarrow \delta - (1+\delta)\ln(1+\delta) \leq -\frac{\delta^2}{3}$$

That is, $f(\delta) = \delta - (1+\delta)\ln(1+\delta) + \frac{\delta^2}{3} \leq 0$

We have, $f'(\delta) = -\ln(1+\delta) + \frac{2}{3}\delta$, and $f''(\delta) = -\frac{1}{1+\delta} + \frac{2}{3}$

Observe that $f''(\delta) < 0$ for $0 \leq \delta \leq \frac{1}{2}$, and $f''(\delta) > 0$ for $\delta > \frac{1}{2}$.

Hence, $f'(\delta)$ first decreases and then increases over $[0,1]$.

Since $f'(0) = 0$ and $f'(1) < 0$, we have $f'(\delta) \leq 0$ over $[0,1]$.

Chernoff Bound 2

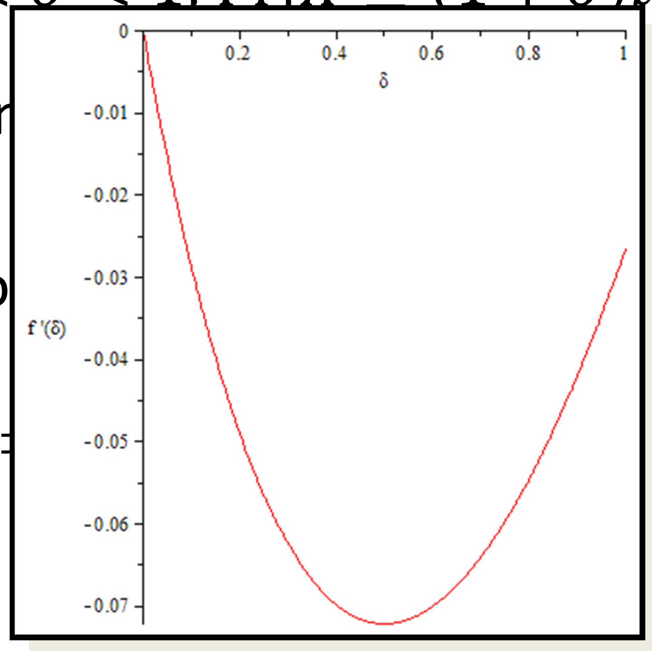
Theorem 4: For $0 < \delta < 1$, $\Pr[X \geq (1+\delta)\mu] \leq e^{-\frac{\mu\delta^2}{3}}$.

Proof: From Theorem $+ \delta)\mu] \leq \left(\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}}\right)^{\mu}$.

We will show that fo $e^{-\frac{\delta^2}{3}}$

$\leq -\frac{\delta^2}{3}$

That is, $f(\delta) = \delta -$ 0



We have, $f'(\delta) = -\ln(1+\delta) + \frac{2}{3}\delta$, and $f''(\delta) = -\frac{1}{1+\delta} + \frac{2}{3}$

Observe that $f''(\delta) < 0$ for $0 \leq \delta \leq \frac{1}{2}$, and $f''(\delta) > 0$ for $\delta > \frac{1}{2}$.

Hence, $f'(\delta)$ first decreases and then increases over $[0,1]$.

Since $f'(0) = 0$ and $f'(1) < 0$, we have $f'(\delta) \leq 0$ over $[0,1]$.

Chernoff Bound 2

Theorem 4: For $0 < \delta < 1$, $\Pr[X \geq (1+\delta)\mu] \leq e^{-\frac{\mu\delta^2}{3}}$.

Proof: From Theorem 3, for $\delta > 0$, $\Pr[X \geq (1+\delta)\mu] \leq \left(\frac{e^\delta}{(1+\delta)^{(1+\delta)}}\right)^\mu$.

We will show that for $0 < \delta < 1$, $\frac{e^\delta}{(1+\delta)^{(1+\delta)}} \leq e^{-\frac{\delta^2}{3}}$

$$\Rightarrow \delta - (1+\delta)\ln(1+\delta) \leq -\frac{\delta^2}{3}$$

That is, $f(\delta) = \delta - (1+\delta)\ln(1+\delta) + \frac{\delta^2}{3} \leq 0$

We have, $f'(\delta) = -\ln(1+\delta) + \frac{2}{3}\delta$, and $f''(\delta) = -\frac{1}{1+\delta} + \frac{2}{3}$

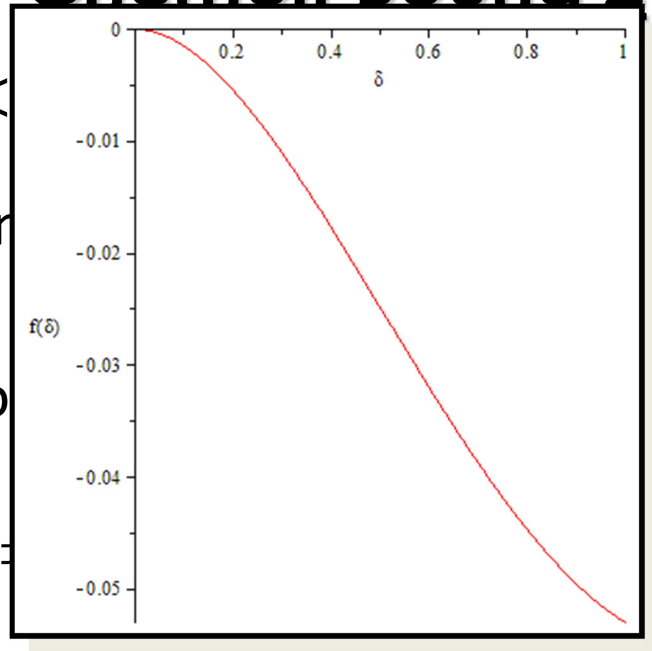
Observe that $f''(\delta) < 0$ for $0 \leq \delta \leq \frac{1}{2}$, and $f''(\delta) > 0$ for $\delta > \frac{1}{2}$.

Hence, $f'(\delta)$ first decreases and then increases over $[0,1]$.

Since $f'(0) = 0$ and $f'(1) < 0$, we have $f'(\delta) \leq 0$ over $[0,1]$.

Since $f(0) = 0$, it follows that $f(\delta) \leq 0$ in that interval.

Chernoff Bound 2



Theorem 4: For $0 <$ … $] \leq e^{-\frac{\mu\delta^2}{3}}$.

Proof: From Theorem … $+ \delta)\mu] \leq \left(\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}}\right)^{\mu}$.

We will show that fo … $e^{-\frac{\delta^2}{3}}$

… $\leq -\frac{\delta^2}{3}$

That is, $f(\delta) = \delta - (1+\delta)\ln(1+\delta) + \frac{\delta^2}{3} \leq 0$

We have, $f'(\delta) = -\ln(1+\delta) + \frac{2}{3}\delta$, and $f''(\delta) = -\frac{1}{1+\delta} + \frac{2}{3}$

Observe that $f''(\delta) < 0$ for $0 \leq \delta \leq \frac{1}{2}$, and $f''(\delta) > 0$ for $\delta > \frac{1}{2}$.

Hence, $f'(\delta)$ first decreases and then increases over $[0,1]$.

Since $f'(0) = 0$ and $f'(1) < 0$, we have $f'(\delta) \leq 0$ over $[0,1]$.

Since $f(0) = 0$, it follows that $f(\delta) \leq 0$ in that interval.

Chernoff Bound 2

Theorem 4: For $0 < \delta < 1$, $\Pr[X \geq (1+\delta)\mu] \leq e^{-\frac{\mu\delta^2}{3}}$.

Proof: From Theorem 3, for $\delta > 0$, $\Pr[X \geq (1+\delta)\mu] \leq \left(\frac{e^\delta}{(1+\delta)^{(1+\delta)}}\right)^\mu$.

We will show that for $0 < \delta < 1$, $\frac{e^\delta}{(1+\delta)^{(1+\delta)}} \leq e^{-\frac{\delta^2}{3}}$

$$\Rightarrow \delta - (1+\delta)\ln(1+\delta) \leq -\frac{\delta^2}{3}$$

That is, $f(\delta) = \delta - (1+\delta)\ln(1+\delta) + \frac{\delta^2}{3} \leq 0$

We have, $f'(\delta) = -\ln(1+\delta) + \frac{2}{3}\delta$, and $f''(\delta) = -\frac{1}{1+\delta} + \frac{2}{3}$

Observe that $f''(\delta) < 0$ for $0 \leq \delta \leq \frac{1}{2}$, and $f''(\delta) > 0$ for $\delta > \frac{1}{2}$.

Hence, $f'(\delta)$ first decreases and then increases over $[0,1]$.

Since $f'(0) = 0$ and $f'(1) < 0$, we have $f'(\delta) \leq 0$ over $[0,1]$.

Since $f(0) = 0$, it follows that $f(\delta) \leq 0$ in that interval.

Chernoff Bound 3

Corollary 1: For $0 < \gamma < \mu$, $\Pr[X \geq \mu + \gamma] \leq e^{-\frac{\gamma^2}{3\mu}}$.

Proof: From Theorem 2, for $0 < \delta < 1$, $\Pr[X \geq (1+\delta)\mu] < e^{-\frac{\mu\delta^2}{3}}$.

Setting $\gamma = \mu\delta$, we get, $\Pr[X \geq \mu + \gamma] \leq e^{-\frac{\gamma^2}{3\mu}}$ for $0 < \gamma < \mu$.

Example: n Fair Coin Flips

$$X_i = \begin{cases} 1 & \text{if the } i\text{th coin flip is heads;} \\ 0 & \text{otherwise.} \end{cases}$$

Then the number of heads in n flips, $X = \sum_{i=1}^{n} X_i$.

We know, $E[X_i] = \Pr[X_i = 1] = \frac{1}{2}$.

Hence, $\mu = E[X] = \sum_{i=1}^{n} E[X_i] = \frac{n}{2}$.

Now putting $\delta = \frac{1}{2}$ in Chernoff bound 2, we have,

$$\Pr\left[X \geq \frac{3n}{4}\right] \leq e^{-\frac{n}{24}} = \frac{1}{e^{\frac{n}{24}}}.$$

Chernoff Bounds 4, 5 and 6

Theorem 5: For $0 < \delta < 1$, $\Pr[X \leq (1-\delta)\mu] \leq \left(\frac{e^{-\delta}}{(1-\delta)^{(1-\delta)}}\right)^{\mu}$.

Theorem 6: For $0 < \delta < 1$, $\Pr[X \leq (1-\delta)\mu] \leq e^{-\frac{\mu\delta^2}{2}}$.

Corollary 2: For $0 < \gamma < \mu$, $\Pr[X \leq \mu - \gamma] \leq e^{-\frac{\gamma^2}{2\mu}}$.

High Probability Bound on Steal Attempts

Theorem: The number of steal attempts is $O\left(p\left(T_\infty + \log\frac{1}{\epsilon}\right)\right)$ with probability at least $1 - \epsilon$, for $0 < \epsilon < 1$.

Proof: Suppose the execution takes $n = 32T_\infty + m$ phases. Each phase succeeds with probability $\geq \frac{1}{4}$. Then $\mu = n \times \frac{1}{4} = 8T_\infty + \frac{m}{4}$.

Let's use Chernoff bound 6 with $\gamma = \frac{m}{4}$ and $m = 32T_\infty + 16\ln\frac{1}{\epsilon}$.

$$Pr[X \leq 8T_\infty] < e^{-\frac{(m/4)^2}{16T_\infty + m/2}} \leq e^{-\frac{(m/4)^2}{m/2 + m/2}} = e^{-\frac{m}{16}} \leq e^{-\frac{16\ln\frac{1}{\epsilon}}{16}} = \epsilon$$

Thus the probability that the execution takes $64T_\infty + 16\ln\frac{1}{\epsilon}$ phases or more is less than ϵ.

Hence, the number of steal attempts is $O\left(p\left(T_\infty + \log\frac{1}{\epsilon}\right)\right)$ with probability at least $1 - \epsilon$.

Parallel Sample Sort

Task: Sort an array $A[1, \dots, n]$ of n distinct keys using $p \le n$ processors.

Steps:

1. **Pivot Selection:** Select and sort $m = p - 1$ pivot elements $e_1, e_2, \dots, e_m$. These elements define $m + 1 = p$ buckets: $(-\infty, e_1), (e_1, e_2), \dots, (e_{m-1}, e_m), (e_m, +\infty)$
2. **Local Sort:** Divide A into p segments of equal size, assign each segment to different processor, and sort locally.
3. **Local Bucketing:** Each processor inserts the pivot elements into its local sorted sequence using binary search, and thus divide the keys among $m + 1 = p$ buckets.
4. **Merge Local Buckets:** Processor i $(1 \le i \le p)$ merges the contents of bucket i from all processors through a local sort.
5. **Final Result:** Each processor copies its bucket to a global output array so that bucket i $(1 \le i \le p - 1)$ precedes bucket $i + 1$ in the output.

Pivot Selection & Load Balancing

In step 4 of the algorithm each processor works on a different bucket. If bucket sizes are not reasonably uniform some processors may become overloaded while some others may mostly sit idle in that step.

We need to select the pivot elements carefully so that the bucket sizes are as balanced as possible.

We may choose the $m = p - 1$ pivots uniformly at random so that the expected size of a bucket is $\frac{n}{m}$.

But if we use such a scheme the largest bucket can still have $\frac{n}{m}\log m$ keys with significant probability leading to significant load imbalance.

A better approach is to use *oversampling*.

Oversampling for Pivot Selection

Steps:

1. **Pivot Selection:**
   a) For some oversampling factor s, select sm keys uniformly at random. Each processor can choose $\frac{sm}{p}$ keys in parallel.
   b) Sort the selected keys on a single processor.
   c) Select the every s-th key as a pivot from the sorted sequence.

Bound on Bucket Sizes

Theorem: If sm keys are initially sampled then no bucket will contain more than $\frac{4n}{m}$ keys with probability at least $1 - \frac{1}{n^2}$, provided $s = 12 \ln n$.

Proof: Split the sorted sequence of n keys into $\frac{m}{2}$ blocks of size $\frac{2n}{m}$ each.

Since every s-th sample is retained as a pivot, at least one pivot will be chosen from a block provided at least $s + 1$ samples are drawn from it.

Fix a block. Let random variable X_i be 1 if the i-th key of the block is sampled, and 0 otherwise. All X_i's are independent. Let $X = \sum_{i=1}^{n} X_i$.

Then $\mu = E[X] = E\left[\sum_{i=1}^{2n/m} X_i\right] = \sum_{i=1}^{2n/m} E[X_i] = \frac{2n}{m} \times \frac{sm}{n} = 2s$.

Bound on Bucket Sizes

Theorem: If sm keys are initially sampled then no bucket will contain more than $\frac{4n}{m}$ keys with probability at least $1 - \frac{1}{n^2}$, provided $s = 12 \ln n$.

Proof (continued): Now putting $\delta = \frac{1}{2}$ in Chernoff bound 5, we get,

$$\Pr[X < s + 1] = \Pr[X \leq s] = \Pr\left[X \leq \left(1 - \frac{1}{2}\right)\mu\right] \leq e^{-\frac{\mu}{2}\left(\frac{1}{2}\right)^2} = e^{-\frac{s}{4}}.$$

For $s = 12 \ln n$, $\Pr[X < s + 1] \leq e^{-\frac{s}{4}} = e^{-3 \ln n} = \frac{1}{n^3}$.

Hence, $\Pr[\textit{there is a block without a pivot}] \leq \frac{m}{2} \cdot \frac{1}{n^3} \leq \frac{1}{n^2}$.

Thus, $\Pr[\textit{no block without a pivot}] \geq 1 - \frac{1}{n^2}$.

Size of a bucket in that case $< 2 \times \frac{2n}{m} = \frac{4n}{m}$.

Parallel Running Time of Parallel Sample Sort

Steps:

1. **Pivot Selection:**

   a) $\mathrm{O}\left(\frac{sm}{p}\right) = \mathrm{O}(\log n)$ [worst case]

   b) $\mathrm{O}(sm \log(sm)) = \mathrm{O}(p \log^2 n)$ [worst case]

   c) $\mathrm{O}(m) = \mathrm{O}(p)$ [worst case]

2. **Local Sort:** $\mathrm{O}\left(\frac{n}{p} \log \frac{n}{p}\right)$ [worst case]

3. **Local Bucketing:** $\mathrm{O}\left(m \log \frac{n}{p}\right) = \mathrm{O}\left(p \log \frac{n}{p}\right)$ [worst case]

4. **Merge Local Buckets:** $\mathrm{O}\left(\frac{4n}{m} \log \frac{4n}{m}\right) = \mathrm{O}\left(\frac{n}{p} \log \frac{n}{p}\right)$ [w.h.p.]

5. **Final Result:** $\mathrm{O}\left(\frac{4n}{m}\right) = \mathrm{O}\left(\frac{n}{p}\right)$ [w.h.p.]

Overall: $\mathrm{O}\left(\frac{n}{p} \log \frac{n}{p} + p \log^2 n\right)$ [w.h.p.]