CPSC 536N: Algorithms That Matter (Term 2, 2016-17)
Assignment 0

Your solution must be written up in LaTeX. You may work in groups of at most 3.

Question 1: Let $A[1, \ldots, n]$ be a sorted array. Let S be a subsequence (also sorted) obtained by sampling each element with probability $q = T/n$, where $T = 3\log(1/\beta)/\epsilon^2$. We have already shown that, with probability at least $1 - \beta$, $|S| \leq (1 + \epsilon)T$.

Fix any $j \leq (1 + \epsilon)T$. Suppose that $S[j]$ was created by sampling $A[J]$. (So J is a random index that depends on the sample.) Let us say that j is good if $J \in j/q + [-\epsilon n, \epsilon n]$; otherwise, j is bad.

Prove that $\Pr[\, j \text{ is bad}\,] \leq 2\beta$. (Possibly my constants are a bit off.) You should use the Chernoff bound, for which you can find a statement my Randomized Algorithms lecture notes:
http://www.cs.ubc.ca/~nickhar/W15/Lecture3Notes.pdf.

Question 2: QTWIDKTA (Question To Which I Don't Know The Answer)

The "TeraSort" algorithm that we discussed in class gives perfect parallelism: $O(n/p)$ data on each of p machines, so long as $p \leq n^{1/3}/\log n$. Can you change the algorithm or analysis so that it works up to $p \leq n^{1/2}/\log^c(n)$?