**Full Proof**

Here are proofs about space complexity to provide error guarantee with probability.

The target is to guarantee

Chernoff bound: , for sub-Gaussian variable ERR with a variance of .

Section 3.1 Disk-Bounded Chunk Sketches

The total error variance of chunk sketches is

Invoke the Chernoff Bound, the target is to achieve

Then there should be .

Thus

When is small enough, there is satisfying above.

Thus , i.e., can provide the required error guarantee.

Note that the chunk sketch should be non-null, i.e.,

Then we have

Thus, the chunk sketch size to guarantee is at least

The I/O cost, i.e., the total size of chunk sketches, is

# **Section 3.2 Disk-Bounded SSTable Sketches**

When ，the variance of error of SSTable sketches is at most times of concatenated chunk sketches.

Now the target is to satisfy

At level L, the size of top sketch is . Then the size of largest top sketch is .

Let . Then ,，

Recall that

The space complexity is

# **Section 4.2 Memory-Constrained Chunk Sketches**

To make the merge-and-compressed chunk sketches more accurate than streaming KLL, in other words, .

That requires

For any top capacity in the streaming KLL,

Note that

Then

Recall that , the I/O cost is

When , that is

# **Section 4.2 Memory-Constrained SSTable Sketches**

The height of the top sketch in level L SSTable is

The target is to make the merge-and-compacted SSTable sketches more accurate than streaming KLL, invoke the lemmas bounding SSTable sketch error with Chunk sketch error and we have:

Let . Then ,

The target is to satisfy

Recall that

Again, the I/O cost is . However, L is determined by N and we need further analyze.

Now the target is to satisfy

Now is bounded, and with another bound of , the I/O cost is

When , T=10, Ts=5, b=0.3979, it is