Precept Outline

- Probability and concentration:
  - Expectation and Variance
  - Markov's and Chebyshev's inequalities
- Expected comparisons in Quicksort

Relevant Book Sections

- Book chapters: 2.3, 2.4 and 2.5

A. Probability and Concentration

Recall the definition of the expected value and variance of a non-negative integer-valued random variable: $\mathbb{E}[X] := \sum_{i=1}^{\infty} \mathbb{P}[X = i] \cdot i$ and $\mathrm{Var}[X] := \mathbb{E}[(X - \mathbb{E}[X])^2]$.

Construct a (non-negative, integer-valued) random variable X with finite expectation and infinite variance.

Prove *Markov's inequality*: for every $t > 0$,

$$\mathbb{P}[X \geq t] \leq \frac{\mathbb{E}[X]}{t}.$$

Prove *Chebyshev's inequality*: for every $t > 0$,

$$\mathbb{P}\left[|X - \mathbb{E}[X]| \geq t\right] \leq \frac{\mathrm{Var}[X]}{t^2}.$$

In other words, prove that X concentrates around the mean with a quadratic tail bound. (*Hint: apply Markov's inequality to a well-chosen random variable.*)

B. Quicksort Analysis

Let C_n be the random variable that counts the number of comparisons made by (2-way) quicksort on an array of length n. Assume the random-pivot version for simplicity, and note that $C_0 = C_1 = 0$; you may also assume distinct keys, although this assumption isn't strictly necessary.

Write a recurrence relation that expresses $\mathbb{E}[C_n]$ in terms of n and $\mathbb{E}[C_k]$ for $k < n$.

Recall that $H_n := \sum_{i=1}^{n} \frac{1}{i}$ is the n-th harmonic number. Using the recurrence you found, prove the following identity for all $n \geq 1$:

$$\mathbb{E}[C_n] = 2(n+1)H_n - 4n.$$

You may use the following identity without proof: $\sum_{k=1}^{n}(k+1)H_k = \frac{(n+1)(n+2)}{2}H_n - \frac{n(n+3)}{4}$. (But feel free to include a proof by induction for this fact!)

It is possible to find and solve for a recurrence relation for $\mathrm{Var}[C_n]$ (a *hard* challenge) to derive the inequality $\mathrm{Var}[C_n] \leq \frac{3}{10}n^2$ when n is large enough.

Show that C_n concentrates around the mean with all but vanishing probability: that is, prove that

$$\mathbb{P}[C_n \geq 2n \log_2 n] \to 0$$

as $n \to \infty$.