# 15-451/651 Algorithms, Spring 2019
# Recitation #12 Worksheet

---

## Chebyshev's inequality

Chebyshev's inequality states that for a random variable $X$ with expectation $\mathbf{E}[X]$ and variance $\mathbf{Var}[X]$, for any $\lambda > 0$, $\mathbf{Pr}[|X - \mathbf{E}[X]| \geq \lambda] \leq \frac{\mathbf{Var}[X]}{\lambda^2}$. Prove Chebyshev's inequality. Hint: use Markov's inequality.

**Solution:** Markov's inequality says for a non-negative random variable $X$ and any $t \geq 1$, that

$$\mathbf{Pr}[X \geq t\mathbf{E}[X]] \leq 1/t.$$

Applying Markov's inequality to the random variable $|X - \mathbf{E}[X]|^2$, we have

$$\mathbf{Pr}[(X - \mathbf{E}[X])^2 \geq \lambda^2] \leq \frac{\mathbf{E}[(X - \mathbf{E}[X])^2]}{\lambda^2}.$$

But using the definition of variance, this is the same as

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

**Variance of the Sum.** Suppose we take $k$ independent copies of a random variable $X$ with expectation $\mathbf{E}[X]$ and variance $\mathbf{Var}[X]$. Let the $k$ copies be denoted $X_1, \ldots, X_k$. Let $T = \frac{1}{k}\sum_{i=1}^{k} X_i$. Then $\mathbf{E}[T] = \mathbf{E}[X]$ by linearity of expectation. Argue that $\mathbf{Var}[T] = \frac{1}{k}\mathbf{Var}[X]$.

**Solution:** We have $\mathbf{Var}[T] = \mathbf{E}[T^2] - (\mathbf{E}[T])^2 = \frac{1}{k^2}\mathbf{Var}[\sum_{i=1}^{k} X_i]$. We will show that if $U$ and $V$ are independent random variables, then $\mathbf{Var}[U + V] = \mathbf{Var}[U] + \mathbf{Var}[V]$. It follows by induction that $\mathbf{Var}[\sum_{i=1}^{k} X_i] = \sum_{i=1}^{k} \mathbf{Var}[X_i] = k \cdot \mathbf{Var}[T]$. Consequently, $\mathbf{Var}[T] = \frac{1}{k^2} \cdot k \cdot \mathbf{Var}[X] = \frac{1}{k}\mathbf{Var}[X]$.

For the claim, note that $\mathbf{Var}[U+V] = \mathbf{E}[(U+V)^2] - (\mathbf{E}[U+V])^2 = \mathbf{E}[U^2 + 2UV + V^2] - (\mathbf{E}[U])^2 - (\mathbf{E}[V])^2 - 2\mathbf{E}[U]\mathbf{E}[V]$. Since $U$ and $V$ are independent, $\mathbf{E}[2UV] = 2\mathbf{E}[U]\mathbf{E}[V]$. We thus have $\mathbf{Var}[U+V] = \mathbf{E}[U^2] - (\mathbf{E}[U])^2 + \mathbf{E}[V^2] - (\mathbf{E}[V])^2 = \mathbf{Var}[U] + \mathbf{Var}[V]$, as desired.

**Flipping Coins.** Suppose you flip $n$ independent coins, each with heads probability $p$. Let $X$ be the number of heads. What is the expectation $\mu = \mathbf{E}(X)$ of $X$? What is the variance $sigma^2 = \mathbf{Var}(X)$? What is the probability that the number of heads differs from its expectation by more than $\lambda$? For the case where $p = 1/2$, what is the probability that the number of heads you see lies outside $n/2 \pm 10\sqrt{n}$?

**Solution:** The expectation is $pn$, by linearity of expectations. Formally, let $X_i$ be the indicator that the $i^{th}$ coin comes up heads. Then $\mathbf{E}[X] = \mathbf{E}[\sum_i X_i] = \sum_i \mathbf{E}[X_i] = np$.

For the variance, here's the fact we proved in the previous part: for *independent* random variables $\mathbf{var}[X] = \mathbf{var}(\sum_i X_i) = \sum_i \mathbf{var}(X_i)$. But then $\mathbf{var}(X_i) = \mathbf{E}[(X_i - \mathbf{E}[X_i])^2] = p(1-p)^2 + (1-p)(0-p)^2 = p(1-p)$. So $\mathbf{var}(X) = np(1-p)$. Now by Chebyshev,

$$\mathbf{Pr}[|X - \mathbf{E}[X]| \geq \lambda] \leq \frac{\mathbf{var}(X)}{\lambda^2}.$$

And when $p = 1/2$, the $\mathbf{E}[X] = n/2$, and $\mathbf{var}(X) = n/4$, so

$$\mathbf{Pr}[|X - n/2| \geq 10\sqrt{n}] \leq \frac{n/4}{(10\sqrt{n})^2} \leq \frac{1}{400}.$$

## CountSketch

**CountSketch probability.** For a random CountSketch matrix $S$ with $k$ rows and a fixed vector $x$, we showed in lecture and last recitation that $\mathbf{E}[\|Sx\|^2] = \|x\|^2$ and $\mathbf{Var}[\|Sx\|^2] = O(\|x\|^4/k)$. Assuming this, for what value of $k$ do we have $\mathbf{Pr}[|\|Sx\|^2 - \|x\|^2| \geq \epsilon\|x\|^2] \leq 1/10$? Big-oh notation for $k$ is fine. Please justify your answer.

**Solution:** Plugging into Chebyshev we get

$$\mathbf{Pr}[|\|Sx\|^2 - \|x\|^2| \geq \epsilon\|x\|^2] \leq \frac{\mathbf{var}(\|Sx\|^2)}{\epsilon^2\|x\|^4} = \frac{1}{k\epsilon^2}.$$

So set $k = 10/\epsilon^2$ to get the bound.

**CountSketch vs. CountMin.** How is the CountSketch data structure different than the earlier CountMin hashing data structure we saw in class?

**Solution:** CountSketch has a 2-universal hash function $h$ mapping coordinates $x_i$ of a vector $x$ into buckets, where $x_i$ goes to $h(i)$, whereas for CountMin the corresponding hash function $h$ just needed to be universal. However, in bucket $j$ CountMin stores $\sum_{i \text{ s.t. } h(i)=j} x_i$, whereas CountSketch stores $\sum_{i \text{ s.t. } h(i)=j} \sigma(i)x_i$, where $\sigma : \{1, 2, 3, \ldots, n\} \to \{-1, 1\}$ is drawn from a 4-universal family.

The guarantees are also different. CountMin gives an estimate of individual element counts (to within some accuracy) under inserts and deletes. CountSketch maintains the squared length of the vector $x$ to within some accuracy.

**Sketch and Solve.** Here is the sketch-and-solve paradigm for the approximate regression problem of outputting an $x' \in \mathbb{R}^d$ for which $\|Ax' - b\|_2^2 \leq (1+\epsilon)\min_x \|Ax - b\|_2^2$ with probability at least $9/10$.

1. Draw $S$ from a $k \times n$ random family of matrices for a value $k = O(d^2/\epsilon^2)$.
2. Compute $S \cdot A$ and $S \cdot b$.

3. Output the solution $x'$ to $\min_{x'} \|(SA)x - (Sb)\|_2$.

What is the overall running time of this algorithm? Note you will need to account for the time for computing $S \cdot A$ and $S \cdot b$, as well as the time to solve the smaller regression problem $\min_x \|SAx - Sb\|_2^2$. You can assume the columns of $SA$ are linearly independent. You can also assume each row of $A$ has at least one non-zero entry (otherwise you can throw out the row without affecting the objective function). Assume $A$ is represented in a way that the non-zero entries are stored in a list so can be accessed without reading the zero entries.

**Solution:** As done in lecture, computing $S \cdot A$ can be done in nnz$(A)$ time where nnz$(A)$ denotes the number of non-zero entries of $A$. Similarly computing $S \cdot b$ can be done in $O(n)$ time. Assuming each row of $A$ has at least one non-zero entry, we have nnz$(A) \geq n$. Now to solve the small regression problem $\min_x \|SAx - Sb\|_2^2$, the solution is $x = ((SA)^T SA)^{-1}(SA)^T Sb$. Computing $(SA)^T(SA)$ takes $O(kd^2)$ time, where $k = O(d^2/\epsilon^2)$ is the number of rows of $S$. Then computing $((SA)^T(SA))^{-1}$ takes $O(d^3)$ time from lecture. Then $(SA)^T Sb$ can be computed in $O(dk)$ time. Finally, computing $((SA)^T(SA))^{-1} \cdot (SA)^T Sb$ can be done in $O(d^2)$ time. Thus, the overall time is nnz$(A) + O(d^4/\epsilon^2)$ time.