1 Basics

Let \(X_1, \ldots, X_n\) be random variables, and let \(X = X_1 + \cdots + X_n\).

- \(E[X] = E[X_1] + \cdots + E[X_n]\)
- \(Var[X] = Var[X_1] + \cdots + Var[X_n]\) if \(X_1, \ldots, X_n\) are pairwise independent.
- \(E[\prod_i X_i] = \prod_i E[X_i]\) if \(X_1, \ldots, X_n\) are mutually independent.

2 Tail bounds

Markov inequality: For nonnegative random variable \(X\) and any \(a > 0\),
\[
\Pr[X \geq a] \leq \frac{E[X]}{a}.
\]

Chebyshev’s inequality: For any \(X\) with \(E[X] = \mu\) and \(Var[X] = \sigma^2\), for any \(a > 0\),
\[
\Pr[|X - \mu| \geq a\sigma] \leq \frac{1}{a^2}.
\]

Chernoff bound: Let \(X_1, X_2, \ldots\) be independent random Bernoulli variables. Let \(X = \sum X_i\) and \(\mu = E[X]\). For any \(\delta > 0\),
\[
\Pr[X \geq (1 + \delta)\mu] \leq \left(\frac{e^\delta}{(1 + \delta)^{1 + \delta}}\right)^\mu.
\]

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

The convenient form: For any \(0 < \delta < 1\),
\[
\Pr[X \geq (1 + \delta)\mu] \leq \exp(-\mu\delta^2/3),
\]
\[
\Pr[X \leq (1 - \delta)\mu] \leq \exp(-\mu\delta^2/2).
\]

2.1 Illustrating example: estimating the number of unique items

Problem: Out of the universe \([n]\), suppose \(u\) of them are unique. How to estimate \(u\) in one pass? For for example let’s assume \(u > n/2\).
Algorithm: Random sample each item \(i \in [n]\) with probability \(p\). Let \(X\) be the number of unique items in the sample. Return \(X/p\).

Expectation and Markov inequality: For any unique \(i\), let \(X_i = 1\) if \(i\) is sampled, and 0 otherwise. Then
\[
E[X] = \sum_{i \text{ is unique}} E[X_i] = up.
\]
So \(E[X/p] = u\). Now with Markov inequality, we have \(\Pr[X \geq (1 + \epsilon)up] \leq 1/(1 + \epsilon)\). This is a one-sided bound and error is large.
Using Chebyshev. The variance of $X$ is

$$\text{Var}[X] = \sum_{i \text{ is unique}} \text{Var}[X_i] = \sum_{i \text{ is unique}} p(1-p) \leq up.$$ 

By Chebyshev’s inequality,

$$\Pr[|X - \mu| \geq a\sqrt{up}] \leq \frac{1}{a^2}.$$ 

If we want the estimator to be within a relative error $a\sqrt{up} = \epsilon up$ with $1-\delta$ probability, we need to set $a = 1/\sqrt{\delta}$ and $p = \frac{1}{\delta^{2/3} u}$. Since we don’t know $u$ in advance, we use a larger $p = \frac{1}{\delta^{2/3} n/2}$, and the space cost is $O(\frac{1}{\delta^{2/3}})$.

Using Chernoff. The $X_i$ are indeed independent Bernoulli, so by the Chernoff bound,

$$\Pr[|X - up| \geq \epsilon up] \leq 2 \exp(-upe^2/2).$$

We need $2 \exp(-upe^2/2) = \delta$, i.e., $p = 2^{\frac{1}{e^2}} \ln \frac{2}{\delta}$. Again we can afford to use a larger $p = \frac{2^{\frac{1}{e^2}}}{\epsilon \sqrt{n/2}} \ln \frac{2}{\delta}$, which implies a space bound of $O(\frac{1}{\delta^{2/3}} \log \frac{1}{\delta})$.

3 Random hash functions

Let $[N]$ be a universe with $N \geq M$. A family of hash functions $\mathcal{H}$ from $[N]$ to $[M]$ is said to be $k$-universal if, for any items $x_1, \ldots, x_k \in [N]$ and for a hash function $h$ chosen uniformly at random from $\mathcal{H}$, we have

$$\Pr[h(x_1) = h(x_2) = \cdots = h(x_k)] \leq \frac{1}{M^{k-1}}.$$ 

A family of hash functions from $[N]$ to $[M]$ is said to be strongly $k$-universal if, for any items $x_1, \ldots, x_k \in [N]$, any values $y_1, \ldots, y_k \in [M]$, and for a hash function $h$ chosen uniformly at random from $\mathcal{H}$, we have

$$\Pr[h(x_1) = y_1 \land h(x_2) = y_2 \land \cdots \land h(x_k) = y_k] = \frac{1}{M^k}.$$ 

For a random chosen $h$ from a $k$-strongly universal family, $h(0), \ldots, h(N-1)$ are $k$-wise independent, and each $h(x)$ is uniform over $[M]$.

If $N = M = p$ is a prime, a 2-strongly universal family can be constructed as follows: Define the family $\mathcal{H} = \{h_{a,b} \mid a, b \in [p]\}$, where

$$h_{a,b}(x) = (ax + b) \mod p.$$ 

For general $N, M$, one can pick some prime $p > N$, and use the family $\mathcal{H} = \{h_{a,b} \mid 1 \leq a \leq p-1, b \in [p]\}$, where

$$h_{a,b}(x) = ((ax + b) \mod p) \mod M,$$

which is almost 2-strongly universal.

These can be extended to $k$-strongly universal by using a degree-$(k - 1)$ polynomial.

3.1 Illustrating example: Two-point sampling