Lecture 6: April 22, 2013

Madhur Tulsiani — Scribe: Hao Tang

1 The Power of Two Choices (cont.)

We already prove the probability of having many high-load bins is small for the case that $\beta_i^2 \geq 2n \ln n$ in the last lecture. Now we look at the case of the other regime. Let i_0 be the minimum i such that $\beta_i^2 < 2n \ln n$. We know with high probability $B_{i_0} \leq \beta_{i_0}$. The probability that B_{i_0+1} is large can be bounded as

$$\begin{aligned}\mathbb{P}\left[B_{i_0+1} \geq k\right] &\leq \mathbb{P}\left[\text{Binom}\left(n, \left(\frac{B_{i_0}}{n}\right)^2\right) \geq k\right] \\ &\leq \mathbb{P}\left[\text{Binom}\left(n, \left(\frac{\beta_{i_0}}{n}\right)^2\right) \geq k\right] \\ &\leq \mathbb{P}\left[\text{Binom}\left(n, \left(\frac{2n \ln n}{n}\right)^2\right) \geq k\right],\end{aligned}$$

where we use the fact that the probability of seeing a certain amount of heads increases as we increase the probability of heads. If we set $k = 6 \ln n$, then Chernoff bound gives

$$\mathbb{P}\left[B_{i_0+1} \geq 6 \ln n\right] \geq e^{-2 \ln n} = \frac{1}{n^2}.$$

We further look at whether there even exists a bin with load more than $i_0 + 2$, and we see that

$$\mathbb{P}\left[B_{i_0+2} \geq 1\right] = \underbrace{\mathbb{P}\left[B_{i_0+2} \geq 1 | B_{i_0+1} > k\right]}_{\leq 1} \underbrace{\mathbb{P}\left[B_{i_0+1} > k\right]}_{\leq \frac{1}{n^2}} + \mathbb{P}\left[B_{i_0+2} \geq 1 | B_{i_0+1} \leq k\right] \underbrace{\mathbb{P}\left[B_{i_0+1} \leq k\right]}_{\leq 1}.$$

Because B_{i_0+1} is small enough, it suffices to bound the only term left in the above equation with Markov's inequality,

$$\mathbb{P}\left[B_{i_0+2} \geq 1 | B_{i_0+1} \leq k\right] \leq \mathbb{E}\left[B_{i_0+2} | B_{i_0+1} \leq k\right] \leq \mathbb{E}\left[\text{Binom}\left(n, \left(\frac{k}{n}\right)^2\right)\right] \leq \frac{k^2}{n}.$$

From the solution of β_i from the last lecture

$$\beta_i = \frac{n}{2^{2^{i-6}} e},$$

we have

$$i_0 = \frac{\ln \ln n}{\ln 2} + O(1).$$

This completes the proof that if we choose two bins at random instead of one, we reduce the number of high-load bins from $O(\ln n)$ to $O(\ln \ln n)$ with high probability.

2 Continuous Random Variables

We will now look at random variables that have values in $\mathbb{R}$. As in the countably infinite case, we only need to define the probability on some elementary sets and let union, intersection, and complement take care of the rest. The elementary sets are intervals $[a, b]$, and we say the probability space is generated by intervals. It is common to define a random variable X through defining a density $\gamma : \mathbb{R} \to \mathbb{R}$ such that

$$\mathbb{P}[X \in [a, b]] = \int_a^b \gamma(x)dx \,.$$

3 Gaussian Random Variables

A Gaussian random variable X is defined through the density function

$$\gamma(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \,,$$

where μ is its mean and σ^2 is its variance, and we write $X \sim \mathcal{N}(\mu, \sigma^2)$. To see the definition gives a valid probability distribution, we need to show $\int_{-\infty}^{\infty} \gamma(x)dx = 1$. It suffices to show for the case that $\mu = 0$ and $\sigma^2 = 1$. First we show the integral is bounded.

Claim 3.1 *$I = \int_{-\infty}^{\infty} e^{-x^2/2}dx$ is bounded.*

Proof: We see that

$$I = \int_{-\infty}^{\infty} e^{-x^2/2}dx = 2\int_0^{\infty} e^{-x^2/2}dx \leq 2\int_0^2 1dx + 2\int_2^{\infty} e^{-x}dx = 4 + 2e^{-2} \,,$$

where we use the fact that I is even and after $x = 2$, $e^{-x^2/2}$ is upper bounded by e^{-x}. ∎

Next we show that the normalization factor is $\sqrt{2\pi}$.

Claim 3.2 *$I^2 = 2\pi$.*

Proof:

$$\begin{aligned}
I^2 &= \int_{-\infty}^{\infty} e^{-x^2/2}dx \int_{-\infty}^{\infty} e^{-y^2/2}dy = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-(x^2+y^2)/2}dxdy \\
&= \int_0^{\infty}\int_0^{2\pi} e^{-r^2/2} r dr d\theta \quad (\text{let } x = r\cos\theta \text{ and } y = r\sin\theta) \\
&= 2\pi \int_0^{\infty} e^{-s}ds \quad (\text{let } s = r^2/2) \\
&= 2\pi \,.
\end{aligned}$$

∎

This completes the proof that the definition gives a valid probability distribution. We prove a useful lemma for later use.

Lemma 3.3 *For* $X \sim \mathcal{N}(0,1)$, $\mathbb{E}\left[e^{tX^2}\right] = \frac{1}{\sqrt{1-2t}}$.

Proof:

$$\begin{aligned}
\mathbb{E}\left[e^{tX^2}\right] &= \int_{-\infty}^{\infty} e^{tx^2} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-(1-2t)x^2/2} dx \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} \frac{dy}{\sqrt{1-2t}} \quad (\text{let } y = \sqrt{1-2t}x) \\
&= \frac{1}{\sqrt{1-2t}}
\end{aligned}$$

■

4 Johnson–Lindenstrauss Lemma

We will use concentration bounds on Gaussian random variables to prove the following important lemma.

Lemma 4.1 (Johnson–Lindenstrauss Lemma) *Let* $\mathcal{P}$ *be a set of* n *points in* $\mathbb{R}^d$*. Let* $0 < \varepsilon < 1$*. For* $k = \frac{8 \ln n}{\varepsilon^2/2 - \varepsilon^3/3}$*, there exists a mapping* $\varphi : \mathcal{P} \to \mathbb{R}^k$ *such that for all* $u, v \in \mathcal{P}$

$$(1-\varepsilon)\|u-v\|^2 \leq \|\varphi(u) - \varphi(v)\|^2 \leq (1+\varepsilon)\|u-v\|^2 .$$

The above lemma is useful for dimensionality reduction, especially when a problem has an exponential dependence on the number of dimensions.

We construct the mapping φ as follows. First choose a matrix $G \in \mathbb{R}^{k \times d}$ such that each $G_{ij} \sim \mathcal{N}(0,1)$ is independent. Define

$$\varphi(u) = \frac{Gu}{\sqrt{k}} .$$

Note that by the above construction φ is oblivious, meaning that it doesn't depend on the points in $\mathcal{P}$, and it is linear.

Before we prove the lemma, we will use the following fact several times.

Fact 4.2 *Let* $Z = c_1 X_1 + c_2 X_2$*, where* $X_1 \sim \mathcal{N}(0,1)$ *and* $X_2 \sim \mathcal{N}(0,1)$ *are independent. Then* $Z \sim \mathcal{N}(0, c_1^2 + c_2^2)$*.*

The strategy of proving the lemma is to first prove that with high probability the lemma holds for any fixed two points and then apply union bounds to get the result for all pairs of points.

Claim 4.3 *Fix* $u, v \in \mathcal{P}$*. Let* $w = u - v$*. With probability greater than* $1 - 1/n^3$*, the following inequality holds,*

$$(1-\varepsilon)\|w\|^2 \leq \|\varphi(w)\|^2 \leq (1-\varepsilon)\|w\|^2 .$$

Proof: Recall that $\varphi(u) = \frac{Gu}{\sqrt{k}}$. Let

$$Z = \frac{k\|\varphi(w)\|^2}{\|w\|^2} = \frac{\sum_{i=1}^k (Gw)_i^2}{\|w\|^2}.$$

We need to show $(1-\varepsilon)k \le Z \le (1+\varepsilon)k$. We know that the sum of Gaussian random variables is still a Gaussian random variable, so $(Gw)_i = G_i w = \sum_{j=1}^n G_{ij} w_j$ is a Gaussian variable. Besides, $\mathsf{Var}\left[\sum_{j=1}^n G_{ij} w_j\right] = \sum_j w_j^2 = \|w\|^2$ according to Fact 4.2. In other words, $G_i w \sim \mathcal{N}(0, \|w\|^2)$. As a result, $Z = \sum_{i=1}^k \frac{(Gw)_i^2}{\|w\|^2} = \sum_{i=1}^k X_i^2$, where $X_i \sim \mathcal{N}(0,1)$. The expectation of each individual element in Gw is

$$\mathbb{E}\left[(Gw)_i^2\right] = \mathbb{E}\left[(G_i w)^2\right] = \mathbb{E}\left[\left(\sum_{j=1}^n G_{ij} w_j\right)^2\right] = \mathsf{Var}\left[\sum_{j=1}^n G_{ij} w_j\right] = \|w\|^2\,.$$

In addition,

$$\mathbb{E}\left[Z\right] = \frac{\sum_{j=1}^k \mathbb{E}\left[(Gw)_i^2\right]}{\|w\|^2} = k\,.$$

Now we prove the concentration bound for Z. The proof is almost identical to Chernoff bound.

$$\begin{aligned}
\mathbb{P}\left[Z \ge (1+\varepsilon)k\right] &\le \mathbb{P}\left[e^{tZ} \ge e^{t(1+\varepsilon)k}\right] \\
&\le \frac{\mathbb{E}\left[e^{tZ}\right]}{e^{t(1+\varepsilon)k}} && \text{(by Markov's inequality)} \\
&= \frac{\mathbb{E}\left[e^{t\sum_{i=1}^k X_i^2}\right]}{e^{t(1+\varepsilon)k}} = \frac{\prod_{i=1}^k \mathbb{E}\left[e^{tX_i^2}\right]}{e^{t(1+\varepsilon)k}} && \text{(by the independence of } X_i^2\text{)} \\
&= \frac{\prod_{i=1}^k \frac{1}{\sqrt{1-2t}}}{e^{t(1+\varepsilon)k}} && \text{(by Lemma 3.3)} \\
&\le \left(\frac{e^{-2(1+\varepsilon)t}}{1-2t}\right)^{k/2} && \text{(assume } t \le 1/2\text{)} \\
&\le \left(e^{-\varepsilon}(1+\varepsilon)\right)^{k/2} && \text{(let } t = \frac{\varepsilon}{2(1+\varepsilon)}\text{)} \\
&\le \left((1-\varepsilon+\frac{\varepsilon^2}{2}-\frac{\varepsilon^3}{6})(1+\varepsilon)\right)^{k/2} && \text{(by Taylor expansion of } e^{-x}\text{)} \\
&\le e^{-\left(\frac{\varepsilon^2}{2}-\frac{\varepsilon^3}{3}\right)\frac{k}{2}} && \text{(by } 1+x \le e^x\text{)}
\end{aligned}$$

We can derive the other side of the inequality in an analogous way, so we have

$$\mathbb{P}\left[|Z-k| \ge \varepsilon k\right] \le 2e^{-\left(\frac{\varepsilon^2}{2}-\frac{\varepsilon^3}{3}\right)\frac{k}{2}} \le 2e^{-3\ln n} = \frac{2}{n^3}\,,$$

where we choose

$$k = \left\lceil \frac{6\ln n}{\frac{\varepsilon^2}{2}-\frac{\varepsilon^3}{3}} \right\rceil\,.$$

∎

To prove Johnson–Lindenstrauss Lemma, we apply the union bound and get the desired result

$$\mathbb{P}\left[\forall u, v \in \mathcal{P}, (1-\varepsilon)\|u-v\|^2 \leq \|\varphi(u)-\varphi(v)\|^2 \leq (1+\varepsilon)\|u-v\|^2\right] \geq 1 - \binom{n}{2}\frac{2}{n^3} \geq 1 - \frac{1}{n}\,.$$