University of Washington Math 523A Lecture 7

Lecturer: Eyal Lubetzky

Monday, April 20, 2009

1 Another application of Hoeffding-Azuma

Here we discuss an application of the Hoeffding-Azuma inequality in which it's important to use nonuniform bounds on the increments (as opposed to our previous applications, which used a uniform bound). We will apply Hoeffding-Azuma to a random version of the Traveling Salesman Problem.

1.1 Problem description

We first describe the deterministic version of TSP:

Traveling Salesman Problem (TSP): Find an optimal circuit traversing n points $p_1, \dots, p_n$ in the unit square. More explicitly,

- Input: $p_1, \dots, p_n \in [0,1]^2$
- Goal: Find a permutation $\pi \in S_n$ which minimizes the sum

$$\sum_{i=1}^{n} \left\| p_{\pi(i+1)} - p_{\pi(i)} \right\| ,$$

  where we identify p_{n+1} with p_1, and $\|\cdot\|$ denotes the Euclidean norm.

Unlike our previous applications (e.g. finding the chromatic number of a graph), TSP has a good polynomial-time approximation scheme. However, it is NP-hard to nail down the precise optimum path. We consider the following stochastic version:

Random TSP:

- The points p_i are IID uniform in $[0,1]^2$.
- If OPT denotes the length of an optimal path, what is $\mathbb{E}[\mathrm{OPT}]$?

1.2 Bounds on the expected length of the optimal path

Fact: $\frac{1}{\sqrt{n}}\,\mathbb{E}[\mathrm{OPT}] \to c^*$ as $n \to \infty$ (where c^* is a constant).

We will prove the weaker statement that $\mathbb{E}[\mathrm{OPT}] \asymp \sqrt{n}$, where the notation $f(n) \asymp g(n)$ means that there are constants c_1 and c_2 such that $c_1 g(n) \leq f(n) \leq c_2 g(n)$ for all large enough n (the notation $f(n) = \Theta(g(n))$ means the same thing).

We first prove a deterministic bound on the maximum length of the optimal path.

Proposition 1.1 (Upper bound for OPT)**.** *Any set of n points in $[0,1]^2$ admits a tour (circuit) of total length $2\sqrt{n}+3$.*

Proof. The idea is to partition the square into strips of height h, then traverse the points in each strip left to right or right to left. It costs about $h \cdot (\#\text{ points in strip})$ to cover the points in each strip, and we have to do this $1/h$ times. The optimal h for this strategy (if we want it to work in general) will be of order $1/\sqrt{n}$. In more detail:

Partition the square into $\sqrt{n}$ horizontal strips of height $h = 1/\sqrt{n}$. (More precisely, we could make $\lfloor\sqrt{n}\rfloor$ strips of height $1/\sqrt{n}$ and one strip of height $\{\sqrt{n}\}/\sqrt{n}$ or something, but we'll ignore this detail…) Now we add $\sqrt{n}$ points $q_1, \ldots, q_{\sqrt{n}}$ alternately to the bottom right or bottom left corner of each strip. That is, q_1 goes at the bottom right corner of the top strip (strip 1), q_2 goes at the bottom left corner of the second strip, and so on down to the bottom strip, which gets the point $q_{\sqrt{n}}$ added to one of its bottom corners. For the upper bound, we are free to add the points q_i to our tour by the triangle inequality.

Now we construct the tour as follows:

- Start with the leftmost point in strip 1, connecting the points in order from left to right and ending at the "new" point q_1 in the bottom right corner. Successive points are connected with straight lines.
- Similarly, connect the points in strip 2 from right to left, starting with q_1 and ending with q_2.
- Continue in this manner, connecting the points in each successive strip alternately from left to right or right to left, ending with the $\sqrt{n}$-th strip at the bottom.
- Connect the final point $q_{\sqrt{n}}$ to the first point in strip 1 to complete the circuit, again with a straight line path. This connection costs at most $\sqrt{2}$, the length of the diagonal of the square.

By the triangle inequality, we can bound the length of the tour by the sum of the horizontal and vertical distances between consecutive points in the tour:

- For each strip, the sum of the horizontal distances between points in the strip is at most 1. Therefore, the sum of the horizontal distances (except for that of the final segment) is at most $1 \cdot (\#\text{ of strips}) = \sqrt{n}$.

- For each point (except for the last point $q_{\sqrt{n}}$), the vertical distance to the next point is at most $h = 1/\sqrt{n}$. Therefore, the sum of the vertical distances (except for that of the final segment) is at most $h \cdot (\#\text{ of points}) = \frac{1}{\sqrt{n}}(n + \sqrt{n}) = \sqrt{n} + 1$.

Adding in the final segment, which has length at most $\sqrt{2}$, we have

$$\begin{aligned}|\text{tour}| &\leq 1 \cdot (\#\text{ of strips}) + h \cdot (\#\text{ of points}) + \sqrt{2} \\ &= \sqrt{n} + (\sqrt{n} + 1) + \sqrt{2} \\ &< 2\sqrt{n} + 3.\end{aligned}$$

$\square$

We now obtain the following lower bound on the expected length of the optimal path.

Proposition 1.2 (Lower bound for $\mathbb{E}[\mathrm{OPT}]$)**.**

$$\mathbb{E}[\mathrm{OPT}] \geq \left(\frac{1}{2} - o(1)\right)\sqrt{n}.$$

Proof. Define $X_i = \operatorname{dist}\big(p_i, \{p_1, \ldots, p_{i-1}, p_{i+1}, \ldots, p_n\}\big)$. Then for any permutation $\pi \in S_n$ we have

$$\sum_{i=1}^{n} \left\|p_{\pi(i+1)} - p_{\pi(i)}\right\| \geq \sum_{i=1}^{n} X_i$$

because the segment beginning at p_i in the tour corresponding to π must have length at least X_i. Note that this lower bound does not depend on the permutation π. In particular, it holds for whatever permutation corresponds to OPT, so we have

$$\mathbb{E}[\mathrm{OPT}] \geq \sum_{i=1}^{n} \mathbb{E}\, X_i = n\, \mathbb{E}\, X_1,$$

where the last step follows because the X_i are identically distributed since the p_i are IID.

<u>Claim:</u> $\mathbb{P}(X_1 \geq x) \geq (1 - \pi x^2)^{n-1}$. Why? The event $\{X_1 \geq x\}$ means that all the points $p_2, \ldots, p_n$ lie outside a ball of radius x around p_1. Conditioning on the location of p_1, we have

$$\begin{aligned}\mathbb{P}(X_1 \geq x \mid p_1) &= \big(1 - \left|B_{p_1}(x) \cap [0,1]^2\right|\big)^{n-1} \\ &\geq (1 - \pi x^2)^{n-1},\end{aligned}$$

where the inequality follows by considering the worst case, when $B_{p_1}(x)$ lies entirely within the square. Since this bound holds independently of p_1, the same bound holds when we take the average over p_1 to get $\mathbb{P}(X_1 \geq x)$.

Now fix $\varepsilon < \frac{1}{2}$. Then there is some ε' such that $\frac{1}{1-\pi x^2} < 1 + \varepsilon'$ for all $x < \varepsilon$, and $\varepsilon' \to 0$ as $\varepsilon \to 0$. Using the inequality $1 - y \geq e^{-\frac{y}{1-y}}$, we have

$$\mathbb{P}(X_1 \geq x) \geq (1 - \pi x^2)^{n-1} \geq \exp\left(-\frac{\pi x^2(n-1)}{1 - \pi x^2}\right) \geq e^{-(1+\varepsilon')\pi x^2 n}.$$

Therefore

$$\begin{aligned}
\mathbb{E}\, X_1 &= \int_0^\infty \mathbb{P}(X_1 \geq x)\, dx \\
&\geq \int_0^\varepsilon e^{-(1+\varepsilon')\pi x^2 n}\, dx \\
&= \frac{1}{\sqrt{2\pi n(1+\varepsilon')}} \int_0^{\varepsilon\sqrt{2\pi n(1+\varepsilon')}} e^{-t^2/2}\, dt,
\end{aligned}$$

so

$$\begin{aligned}
\lim_{n\to\infty} \sqrt{n} \cdot \mathbb{E}\, X_1 &\geq \lim_{n\to\infty} \frac{1}{\sqrt{2\pi(1+\varepsilon')}} \int_0^{\varepsilon\sqrt{2\pi n(1+\varepsilon')}} e^{-t^2/2}\, dt \\
&= \frac{1}{\sqrt{2\pi(1+\varepsilon')}} \cdot \frac{\sqrt{2\pi}}{2} \\
&= \frac{1}{2\sqrt{1+\varepsilon'}}.
\end{aligned}$$

Letting $\varepsilon \to 0$ (hence $\varepsilon' \to 0$) we have $\lim_{n\to\infty} \sqrt{n} \cdot \mathbb{E}\, X_1 \geq \frac{1}{2}$. In particular,

$$\mathbb{E}[\mathrm{OPT}] \geq n\, \mathbb{E}\, X_1 \geq \left(\frac{1}{2} - o(1)\right) \sqrt{n}.$$

□

Combining Propositions 1.1 and 1.2 we see that

$$\left(\frac{1}{2} - o(1)\right) \sqrt{n} \leq \mathbb{E}[\mathrm{OPT}] \leq \Big(2 + o(1)\Big) \sqrt{n}.$$

Therefore $\mathbb{E}[\mathrm{OPT}] \asymp \sqrt{n}$, though it is not clear that $\frac{1}{\sqrt{n}}\, \mathbb{E}[\mathrm{OPT}] \to c^*$ since the constants 1/2 and 2 don't match.

1.3 Concentration about $\mathbb{E}[\mathrm{OPT}]$

To get a basic concentration bound for OPT, we can apply Hoeffding-Azuma to Doob's martingale: Expose $p_1, p_2, \ldots, p_n$ in succession, and let $y_i = \mathbb{E}[\mathrm{OPT} \mid p_1, \ldots, p_i]$. Then $|y_{i+1} - y_i| \leq 2\sqrt{2}$, so by Hoeffding-Azuma, $\mathbb{P}(|\mathrm{OPT} - \mathbb{E}[\mathrm{OPT}]| \geq a) \leq 2\exp\left(-\frac{a^2}{2n\cdot 8}\right)$. Taking $a = \varepsilon\sqrt{n}$, we get

$$\mathbb{P}\left(|\mathrm{OPT} - \mathbb{E}[\mathrm{OPT}]| \geq \varepsilon\sqrt{n}\right) \leq 2e^{-\varepsilon^2/16}.$$

But since $\mathbb{E}[\mathrm{OPT}] \asymp \sqrt{n}$, this doesn't tell us very much. The problem is that we used the constant bound of $2\sqrt{2}$ for the all the increments $|y_{i+1} - y_i|$. We need to be a bit more careful and find a better bound that varies with i. The situation is analogous to the proof of Proposition 1.1 above, where we needed to find a path with many short segments and only used the wasteful bound of $\sqrt{2}$ for the last step.

Proposition 1.3. *There is a constant A such that for any $\varepsilon > 0$,*

$$\mathbb{P}\left(|\mathrm{OPT} - \mathbb{E}[\mathrm{OPT}]| \geq \varepsilon\sqrt{n}\right) \leq 2e^{-A\varepsilon^2 \frac{n}{\log n}}.$$

Proof. For notational convenience, let L_n be the optimal path length OPT for a tour of the points $p_1, \ldots, p_n$. As before, we will consider Doob's martingale $y_i = \mathbb{E}[L_n \mid p_1, \ldots, p_i]$. For each i let $\widehat{L}_n(i)$ be the optimal path length for the TSP on the $n-1$ points $p_1, \ldots, p_{i-1}, p_{i+1}, \ldots, p_n$. We claim that

$$\widehat{L}_n(i) \leq L_n \leq \widehat{L}_n(i) + 2Z_i, \tag{1.1}$$

where $Z_i := \mathrm{dist}\big(p_i, \{p_{i+1}, \ldots, p_n\}\big)$. The first inequality in (1.1) holds because the optimal path length increases when we add the point p_i to the tour.

To prove the second inequality in (1.1), suppose p_j attains the minimal distance of p_i from $\{p_{i+1}, \ldots, p_n\}$, so $j \in \{i+1, \ldots, n\}$ and $|\overline{p_j p_i}| = Z_i$. Then we can use the following tour of the n points: Suppose π is the permutation of $\{1, 2 \ldots, i-1, i+1, \ldots, n\}$ corresponding to $\widehat{L}_n(i)$, so the tour for $\widehat{L}_n(i)$ is

$$p_{\pi(1)} \to p_{\pi(2)} \to \ldots \to p_j \to p_s \to \ldots \to p_{\pi(n)} \to p_{\pi(1)}.$$

Then we can add p_i to the tour between the points p_j and p_s:

$$p_{\pi(1)} \to p_{\pi(2)} \to \ldots \to p_j \to p_i \to p_s \to \ldots \to p_{\pi(n)} \to p_{\pi(1)}.$$

The cost of this new path is $\widehat{L}_n(i)$ for the original tour of $n-1$ points, plus $|\overline{p_j p_i}| + |\overline{p_i p_s}|$ for the two segments we added, minus $|\overline{p_j p_s}|$ for the segment we deleted. By the reverse triangle inequality, we have $|\overline{p_i p_s}| - |\overline{p_j p_s}| \leq |\overline{p_j p_i}|$. Thus, since the optimal path on all n points can be no longer than the path we constructed, we have

$$\begin{aligned}
\widehat{L}_n &\leq \widehat{L}_n(i) + |\overline{p_j p_i}| + |\overline{p_i p_s}| - |\overline{p_j p_s}| \\
&\leq \widehat{L}_n(i) + |\overline{p_j p_i}| + |\overline{p_j p_i}| \\
&= \widehat{L}_n(i) + 2Z_i,
\end{aligned}$$

proving the upper bound in (1.1).

Now observe that $\widehat{L}_n(i)$ does not depend on p_i. Therefore, if $\mathcal{F}_t = \sigma(p_1, \ldots, p_t)$, we have

$$\mathbb{E}\left[\widehat{L}_n(i) \mid \mathcal{F}_i\right] = \mathbb{E}\left[\widehat{L}_n(i) \mid \mathcal{F}_{i-1}\right]. \tag{1.2}$$

Taking conditional expectations of (1.1) we have

$$\mathbb{E}\left[\widehat{L}_n(i) \mid \mathcal{F}_{i-1}\right] \leq y_{i-1} \leq \mathbb{E}\left[\widehat{L}_n(i) \mid \mathcal{F}_{i-1}\right] + 2\,\mathbb{E}\Big[Z_i \mid \mathcal{F}_{i-1}\Big]$$

and

$$\mathbb{E}\left[\widehat{L}_n(i) \mid \mathcal{F}_i\right] \le y_i \le \mathbb{E}\left[\widehat{L}_n(i) \mid \mathcal{F}_i\right] + 2\,\mathbb{E}\big[Z_i \mid \mathcal{F}_i\big].$$

Subtracting these inequalities, (1.2) implies

$$y_i - y_{i-1} \le 2\,\mathbb{E}\big[Z_i \mid \mathcal{F}_i\big] \quad \text{and} \quad y_{i-1} - y_i \le 2\,\mathbb{E}\big[Z_i \mid \mathcal{F}_{i-1}\big].$$

Therefore

$$|y_i - y_{i-1}| \le 2 \max\left\{\mathbb{E}\big[Z_i \mid \mathcal{F}_i\big],\, \mathbb{E}\big[Z_i \mid \mathcal{F}_{i-1}\big]\right\}. \tag{1.3}$$

Notice that we defined Z_i starting at the index $i+1$ so that it doesn't depend on the past, which we are conditioning on.

Now we use an argument similar to that in the proof of Proposition 1.2 above to get a bound on the expectation of Z_i. For a point $Q \in [0,1]^2$, let $Z_i(Q) = \operatorname{dist}\big(Q, \{p_{i+1}, \ldots, p_n\}\big)$. Then there is some constant c such that for any Q we have

$$\mathbb{P}(Z_i(Q) \ge x) \le (1 - cx^2)^{n-i}$$

for $0 \le x \le \sqrt{2}$, and $\mathbb{P}(Z_i(Q) \ge x) = 0$ for $x \ge \sqrt{2}$. This follows by considering the worst possible location for Q, which is a corner of the square. Using the inequality $1 - y \le e^{-y}$, we have, for any Q,

$$\mathbb{E}[Z_i(Q)] \le \int_0^{\sqrt{2}} (1 - cx^2)^{n-i}\, dx \le \int_0^{\sqrt{2}} e^{-cx^2(n-i)}\, dx \le \frac{c'}{\sqrt{n-i}}$$

for some constant c' when $i \ne n$, and $\mathbb{E}[Z_n(Q)] \le \sqrt{2}$. Combining this with (1.3), we have

$$|y_i - y_{i-1}| \le \frac{c'}{\sqrt{n-i}} \quad \text{for } i \ne n, \quad \text{and } |y_n - y_{n-1}| \le 2\sqrt{2}.$$

Now, applying Hoeffding-Azuma to the martingale y_i using these bounds on the increments, we get

$$\mathbb{P}\left(|L_n - \mathbb{E}\,L_n| \ge a\right) \le 2\exp\left(-\frac{a^2}{2\left[8 + \sum_{i=1}^{n-1} \frac{4c'^2}{n-i}\right]}\right).$$

The sum in the denominator is of order $\log n$, so setting $a = \varepsilon\sqrt{n}$ we get

$$\mathbb{P}\left(|L_n - \mathbb{E}\,L_n| \ge \varepsilon\sqrt{n}\right) \le 2\exp\left(-A\varepsilon^2 \frac{n}{\log n}\right)$$

for some constant A. $\square$