Day 9 2/16/2015

Activity:

Let $X_i \overset{iid}{\sim} N(0, \sigma^2)$ $\qquad p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-x^2/2\sigma^2}$

What roughly is $\max_{i \in [n]} |X_i|$? $\sim \sigma\sqrt{\log n}$

Write a high probability upper bound for $\max_{i \in [n]} |X_i|$

$$P(|X_i| > t\sigma) \leq \frac{1}{\sqrt{2\pi}} e^{-t^2/2} \quad \text{for } t\sigma > 1$$

$$P(\max_i |X_i| > t\sigma) \leq \frac{1}{\sqrt{2\pi}} n e^{-t^2/2}$$

$$P(\max |X_i| > c\sqrt{2\log n}\,\sigma) \leq \frac{1}{\sqrt{2\pi}} n e^{-\frac{c^2 2\log n}{2}}$$

$$\leq \frac{1}{\sqrt{2\pi}} n\, n^{-c^2}$$

$$\leq \frac{1}{\sqrt{2\pi}} \frac{1}{n^{c^2-1}}$$

Fat-tailed distributions

- Coordinate random vectors:

  Let $\{e_i\}_{i=1}^n$ be canonical basis of $\mathbb{R}^n$.

  Choose one $\sqrt{n}\, e_i$ at random.

  $\mathbb{E}[XX^t] = I_n.$ $\quad\quad \|X\|_{\psi_2} = \sqrt{n}. \gg 1.$

  Bad subgaussian bound. Treat as fat tailed.

- Random Fourier modes

  Let $W_{w,t} = e^{-2\pi i w t/n}$ $\quad\quad w, t \in (0, \cdots, n-1)$

  $\quad$ DFT$_{n \times n}$ matrix

  Let X be random row of W.

  $\mathbb{E}[XX^t] = I_n$

  Bad subgaussian bound. Treat as fat tailed

Coupon Collectors Problem

Urn contains n unique coupons.
Pick uniformly w/ replacement.
How many picks until you get all coupons.

Answer : $\sim n \log n$

Analysis of expectation:

Let T - time to collect all coupons

t_i - after i coupons collected, how many picks until new coupon

$$E[T] = E[t_0] + \cdots + E[t_{n-1}]$$

$$E[t_i] = \frac{n}{n-i}$$

$$E[T] = \frac{n}{n} + \frac{n}{n-1} + \cdots + \frac{n}{1} = n\Big[\underbrace{\frac{1}{n} + \frac{1}{n-1} + \cdots + \frac{1}{1}}_{\text{harmonic series}}\Big]$$

$$\approx n \log n$$

$$E[T] = n H_n \quad \text{w/ } H_n = \sum_{i=1}^{n} \frac{1}{i} \approx \log n + \gamma + \cdots$$

$$P(T \geq c n H_n) \leq \frac{1}{c} \quad \text{by Markov.}$$

Concentration of singular Values of tall matrices with heavy-tailed rows

Thm:

Let A be $N \times n$ w/ rows A_i indep. isot. rv in $\mathbb{R}^n$

~~Let~~ $\cancel{\|A_i\|_2 \le c\sqrt{n}}$

Let $\|A_i\|_2 \le \sqrt{m}$ almost surely $\forall i$. $\forall t \geq 0$

$$\sqrt{N} - t\sqrt{m} \le \sigma_{min}(A) \le \sigma_{max}(A) \le \sqrt{N} + t\sqrt{m} \quad \text{w/ prob } 1-2ne^{-ct^2}$$

Typically, take $m = O(n)$.

$$\sqrt{N} - t\sqrt{n} \le \sigma_{min}(A) \le \sigma_{max}(A) \le \sqrt{N} + t\sqrt{n} \quad \text{w/ prob } 1-2ne^{-c't^2}$$

Compare to subgaussian Thm

Thm: A is $N \times n$ w/ rows A_i indep isot. subgaussian (w norm K) rv in $\mathbb{R}^n$

$$\sqrt{N} - C_K\sqrt{n} - \tilde{t} \le \sigma_{min}(A) \le \sigma_{max}(A) \le \sqrt{N} + C_K\sqrt{n} + \hat{t}$$

w/ prob $1-2e^{-c_K\hat{t}^2}$.

Differences: Boundedness assumption $\|A_i\|_2 \le \sqrt{m}$ almost surely

worse probability bound — $t \approx \hat{t}/\sqrt{n}$ $\quad \hat{t} = \sqrt{n}\, t$