Concentration.

Sunday, March 30, 2025 5:56 PM

We have seen Chebyshev's inequality:

$$\mathbb{P}[|X-\mathbb{E}X|>\beta] \le \frac{\mathrm{Var}X}{\beta^2}$$

- Today: more concentration inequalities.
  - Chernoff - Cramér bound $\Rightarrow$ large-deviation
  - Bernstein inequality / Hoeffding inequality (sum of independent)
  - Azuma-Hoeffding inequality / bounded difference estimates (Martingale)

Applications:
- norm bound of random matrices
- first-passage percolation
- chromatic number of graph.

- Chernoff-Cramér: $\mathbb{P}[X \ge \beta] \le e^{-s\beta}\underline{\mathbb{E}[e^{sX}]}$, $\forall s$

  moment generating function (MGF) $= \sum_{k=0}^{\infty} \frac{s^k \mathbb{E}[X^k]}{k!}$

  proof: Markov inequality.

- Useful for sum of independent:

  MGF factors: $\mathbb{E}[e^{s\sum_{i=1}^m X_i}] = \prod_{i=1}^m \mathbb{E}[e^{sX_i}]$

- Example: Large deviation

  $\Lambda(s) = \log \mathbb{E}_\mu[e^{sX}]$ cumulant-generating function. $\left[\begin{array}{l}\Lambda \text{ convex} \\ \Lambda_* \text{ convex}\end{array}\right]$

  Legendre transform: $\Lambda_*(x) = \sup_s sx - \Lambda(s)$

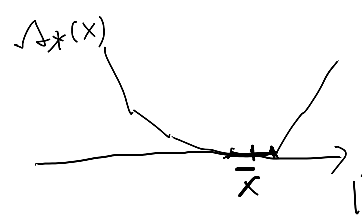
  $\hookrightarrow \ge 0$ since $\Lambda(0)=0$ $(\bar{X} = \mathbb{E}X)$

  $\Lambda_*(\bar{X}) = 0$, since $\Lambda(s) = \log \mathbb{E}_\mu[e^{sX}] \ge \mathbb{E}_\mu[\log e^{sX}] = s\bar{X}$

  $A_m = \frac{1}{m}(X_1 + \cdots + X_m)$, i.i.d. $\sim \mu$

  (Cramér) $-\frac{1}{m}\log(\mathbb{P}[A_m \in I]) \to \inf_{x\in I} \Lambda_*(x)$ as $m \to \infty$

<u>proof</u>. Upper bound: $\mathbb{P}[A_m \in I] \le \exp(-m \inf_{x\in I} \Lambda_*(x))$

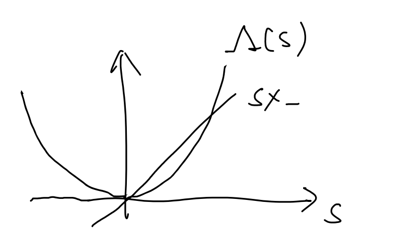


If $\bar{x} \in I$, obvious

Assume $\inf I = x_- > \bar{x}$

$\mathbb{P}[A_m \ge x_-] \le \exp(-smx_- + m\Lambda(s))$, $\forall s \ge 0$

$\le \exp(-m\Lambda_*(x_-))$ [Here $sx_- - \Lambda(s)$ maximized when $s \ge 0$, since $\Lambda'(0) = \mathbb{E}X = \bar{x} < x_-$]



Lower bound: $\mathbb{P}[A_m \in I] \gtrsim \exp(-m\Lambda_*(x))$ for any $x \in I$.

Change of measure: $\frac{\tilde{\mu}}{\mu}(x) = \exp(sx - \Lambda(s))$

$$\mathbb{P}[A_m \in (x-\varepsilon, x+\varepsilon)] \approx \mathbb{E}[\exp(\underset{\sim}{-m\tilde{A}_m s + m\Lambda(s)})\mathbb{1}[\tilde{A}_m \in (x-\varepsilon, x+\varepsilon)]]$$

$$\approx \exp(-m(sx - \Lambda(s)))\,\mathbb{P}[\tilde{A}_m \in (x-\varepsilon, x+\varepsilon)]$$

Choose s such that $\mathbb{P}[\tilde{A}_m \in (x-\varepsilon, x+\varepsilon)] \to 1$ $(\mathbb{E}_{\tilde{\mu}}X = x)$

- Some concentration inequalities

  Take $X_1, \dots, X_n$ independent, with $\mu_i = \mathbb{E}X_i$, $\sigma_i^2 = \mathrm{Var}(X_i)$ and $|X_i - \mu_i| \le c_i$ $\quad c = \max c_i$

  For $T = X_1 + \cdots + X_n$. $\Rightarrow \mathbb{P}[T - \mathbb{E}T > \beta] < \begin{cases} \exp(-\beta^2/4\sum_{i=1}^n \sigma_i^2) & 0 \le \beta \le \frac{\sum \sigma_i^2}{c} \\ \exp(-\beta/4c) & \beta > \frac{\sum\sigma_i^2}{c}\end{cases}$ $\Big/ \exp\left(-\frac{\beta^2}{2\sum_{i=1}^n \sigma_i^2 + 2c\beta/3}\right)$

  (Bernstein)

  $$\mathbb{P}[T - \mathbb{E}T > \beta] < \exp\left(-2\beta^2 / \sum_{i=1}^n c_i^2\right)$$

  (Hoeffding) (Both for sub-Gaussian as well)

<u>proof</u> Using Chernoff-Cramér; need: compute MGF

(Bernstein) $\mathbb{E}\exp(s(X_i - \mu_i))$

$$= \sum_{k=0}^{\infty} \frac{s^k \mathbb{E}(X_i-\mu_i)^k}{k!} \le 1 + \sum_{k=2}^{\infty} \frac{s^k}{k!}\sigma_i^2 \cdot c^{k-2} \le 1 + \frac{s^2\sigma_i^2}{2} + \frac{s^2\sigma_i^2}{6}\sum_{k=3}^{\infty}(sc)^{k-2} = 1 + \frac{s^2\sigma_i^2}{2}\left(1 + \frac{1}{3}\frac{sc}{1-sc}\right) \le 1 + s^2\sigma_i^2 \le \exp(s^2\sigma_i^2)$$

$(\mathbb{E}(X_i - \mu_i)^k \le \mathbb{E}(X_i-\mu_i)^2 \cdot c^{k-2})$ $\quad\downarrow$ when $s \le \frac{1}{2c}$

(Hoeffding) $\mathbb{E}\exp(s(X_i - \mu_i)) \le \frac{1}{2}(e^{c_i s} + e^{-c_i s}) \le e^{2c_i^2 s^2}$

- Operator norm for random matrices

$X = (X_{ij})$ $m \times n$ random matrix,

$$\|X\|_2 = \sup_{\substack{\|u\|_2 = \|v\|_2 = 1 \\ u\in\mathbb{R}^m,\, v\in\mathbb{R}^n}} u^T X v$$

Suppose X_{ij} i.i.d. rademacher $(\mathbb{P}[X_{ij}=1] = \mathbb{P}[X_{ij}=-1] = \frac{1}{2})$

<u>Claim</u>. $\mathbb{P}[\|X\|_2 \ge C(\sqrt{m} + \sqrt{n} + t)] < e^{-t^2}$

($C > 0$ is some constant)

<u>proof</u>. ① Bound u^TXv for fixed u, v

$u^TXv = \sum_{i=1}^m\sum_{j=1}^n u_i v_j X_{ij}$

$$\mathbb{P}[u^TXv \ge C(\sqrt{m}+\sqrt{n}+t)] < \exp\left(-\frac{2C^2(\sqrt{m}+\sqrt{n}+t)^2}{\sum_{i=1}^m\sum_{j=1}^n u_i^2 v_j^2}\right) = \exp(-2C^2(\sqrt{m}+\sqrt{n}+t)^2) \le \exp(-2C^2(m+n+t^2))$$

(Hoeffding)

② For S^{m-1} $(\frac{1}{4}\text{-net})$

(i.e. $\{u \in \mathbb{R}^m : \|u\|_2 = 1\}$)

$\exists\ P_m \subseteq S^{m-1}$, $|P_m| = 12^m$, $\bigcup_{x\in P_m} B_{\frac{1}{4}}(x) \supseteq S^{m-1}$

$\Rightarrow \mathbb{P}[\max_{u\in P_m, v\in P_n} u^TXv \ge C(\sqrt{m}+\sqrt{n}+t)] < e^{-t^2}$ (union bound)

③ $\forall\ u \in P_m$, $v \in P_n$, $\tilde{u} \in S^{m-1}$, $\tilde{v} \in S^{n-1}$, $\|u - \tilde{u}\|_2 \le \frac{1}{4}$, $\|v - \tilde{v}\|_2 \le \frac{1}{4}$,

$$|u^TXv - \tilde{u}^TX\tilde{v}| \le |(u^T - \tilde{u}^T)Xv| + |\tilde{u}^TX(v - \tilde{v})|$$

$$\le \frac{1}{4}\|X\|_2 + \frac{1}{4}\|X\|_2 = \frac{1}{2}\|X\|_2$$

$\Rightarrow \tilde{u}^TX\tilde{v} \le \frac{1}{2}\|X\|_2 + C(\sqrt{m}+\sqrt{n}+t)$, with prob $> 1 - e^{-t^2}$

$\Rightarrow \|X\|_2 \le 2C(\sqrt{m}+\sqrt{n}+t)$, with prob $> 1 - e^{-t^2}$

Martingale concentration

Recall: Martingale $Z_1, Z_2, \dots, Z_n, \dots$

s.t. $\mathbb{E}|Z_n| < \infty$

$\mathbb{E}[Z_{n+1} | Z_1, \dots, Z_n] = Z_n$

(A_i, B_i can depend on $X_1, \dots, X_{i-1}$) (but c_i is constant)

Azuma/Azuma-Hoeffding ineq: $A_i \le Z_i - Z_{i-1} \le B_i$, $B_i - A_i \le c_i$

$$\mathbb{P}[\max_{1\le i\le n} Z_i - Z_0 \ge \beta] \le \exp\left(-\frac{2\beta^2}{\sum c_i^2}\right)$$

<u>proof</u> $\mathbb{P}[\max_{1\le i\le n} Z_i - Z_0 \ge \beta]$

$\le \mathbb{P}[\max_{0\le i\le n} \exp(s(Z_i - Z_0)) \ge e^{s\beta}]$

$\le \mathbb{E}[\exp(s(Z_\tau - Z_0))]e^{-s\beta}$

$\tau := \min\{t : Z_t - Z_0 \ge \beta\}$ or n (stopping time)

$\le \mathbb{E}[\exp(s(Z_n - Z_0))]e^{-s\beta} = \mathbb{E}[\exp(s\sum_{i=1}^n (Z_i - Z_{i-1}))]e^{-s\beta}$

Then Chernoff-Cramér.

- Bounded difference estimates.

  $X_1, X_2, \dots, X_n$ independent

  each $X_i \in \Omega_i$

  $f: \Omega_1 \times \cdots \times \Omega_n \to \mathbb{R}$

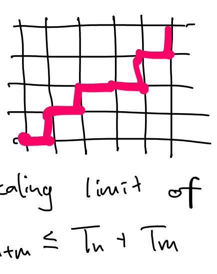
  $D_i = \max_{\substack{x_1\in\Omega_1,\dots,x_n\in\Omega_n \\ \tilde{x}_i \in \Omega_i}} |f(x_1,\dots,x_i,\dots,x_n) - f(x_1,\dots,\tilde{x}_i,\dots,x_n)|$

  Let $Z_i = \mathbb{E}[f(X_1,\dots,X_n) | X_1,\dots,X_i]$ (McDiarmid's ineq)

  $\Rightarrow A_i \le Z_i - Z_{i-1} \le B_i$, $B_i - A_i \le D_i$

  $\Rightarrow \mathbb{P}[f(X_1,\dots,X_n) - \mathbb{E}f(X_1,\dots,X_n) \ge \beta] \le \exp\left(-\frac{2\beta^2}{D_1^2 + \cdots + D_n^2}\right)$

- Example: First-Passage Percolation (Last-)



$w_e \sim \mathrm{Unif}[0,1]$ $\forall e \in E$ (edge weights)

$T_n = \min_\gamma \sum_{e\in\gamma} w_e$

γ: up-right path from $(0,0)$ to (n,n)

Scaling limit of T_n as $n \to \infty$?

$T_{n+m} \le T_n + T_m$ (sub-additivity)

$\Rightarrow \mathbb{E}T_{n+m} \le \mathbb{E}T_n + \mathbb{E}T_m$, $\frac{1}{n}\mathbb{E}T_n$ converges as $n \to \infty$

(Exercise: $\lim_{n\to\infty} \frac{1}{n}\mathbb{E}T_n > 0$, by large deviation)

By bounded difference estimates,

(X_i: w_e for e between $x+y=i-1$ and $x+y=i$)

$\mathbb{P}[T_n - \mathbb{E}T_n \ge \beta] \le \exp\left(-\frac{\beta^2}{4n}\right)$ $\quad(D_1 = D_2 = \cdots = D_{2n} = 1)$

(In particular, $\mathrm{Var}(T_n)$ is $O(n)$)

(KPZ universality conjecture: $\mathrm{Var}(T_n)$ is of order $O(n^{2/3})$)

(state of art: $c\log(n) \le \mathrm{Var}(T_n) \le \frac{cn}{\log(n)}$) (Newman-Piza) (Benjamini-Kalai-Schramm)

Example. • Pattern matching

$X_1, X_2, \dots, X_n$, i.i.d., each uniform from $\{1, \dots, s\}$

For $a = (a_1, \dots, a_k) \in \{1, \dots, s\}^k$

N_n: # i such that $(X_i, \dots, X_{i+k-1}) = (a_1, \dots, a_k)$

$\mathbb{E}N_n = (n-k+1)s^{-k}$

$\mathbb{P}[|N_n - (n-k+1)s^{-k}| \ge bk\sqrt{n}] \le 2\exp\left(-\frac{2b^2k^2n}{k^2n}\right) \le 2e^{-2b^2}$

$\downarrow$

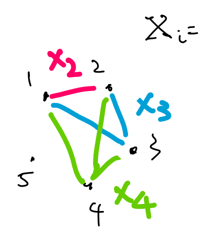
$D_i \le k$

- Chromatic number

  Take Erdős-Rényi graph $G(n,p)$

  χ: minimum number of colors to properly color $G(n,p)$

  $X_i = \{\mathbb{1}(i,j) \in G(n,p) : 1 \le j \le i\}$



Claim: change X_i alters χ by ≤ 1

$(D_i \le 1)$

$\Rightarrow \mathbb{P}[|\chi - \mathbb{E}\chi| \ge b\sqrt{n-1}] \le 2\exp\left(-\frac{2b^2(n-1)}{n-1}\right) = 2e^{-2b^2}$