

Concentration.

Sunday, March 30, 2025 5:56 PM

We have seen Chebychev's inequality:

$$\Pr[|X - \mathbb{E}X| > \beta] \leq \frac{\text{Var}(X)}{\beta^2}$$

- 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: $\Pr[X \geq \mu] \leq e^{-\frac{\beta}{2}} \mathbb{E}[e^{sX}]$, $\forall s$

proof: Markov inequality.

moment generating function (MGF)

- Useful for sum of independent:

$$= \sum_{k=0}^{\infty} \frac{s^k \mathbb{E}[X^k]}{k!}$$

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

- Example: Large deviation

$\Delta(s) = \log \mathbb{E}[e^{sX}]$ cumulant-generating function.

Legendre transform: $\Delta^*(x) = \sup_s sX - \Delta(s)$

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

(Cramér) $-\frac{1}{m} \log(\Pr[A_m \in I]) \rightarrow \inf_{x \in I} \Delta^*(x)$ as $m \rightarrow \infty$

$\Delta(s) \geq \log \mathbb{E}[e^{sX}]$

$\Delta^*(x) \geq x$ since $\Delta^*(0) = 0$, since

$\Delta(s) = \log \mathbb{E}[e^{sX}] \geq \mathbb{E}[\log e^{sX}] = s\bar{X}$

Proof. Upper bound: $\Pr[A_m \in I] \leq \exp(-m \inf_{x \in I} \Delta^*(x))$

$\Delta^*(x)$ If $\bar{X} \in I$, obvious

Assume $\inf I = X_- > \bar{X}$

$\Pr[A_m \in I] \leq \exp(-s m X_- + m \Delta^*(s))$, $\forall s \geq 0$

$\leq \exp(-m \Delta^*(X_-))$

Lower bound: $\Pr[A_m \in I] \geq \exp(-m \Delta^*(s))$ for any $s \in I$.

Change of measure: $\frac{1}{m} (x) = \exp(sX - \Delta(s))$

$\Pr[A_m \in (x-\epsilon, x+\epsilon)] \approx \mathbb{E}[\exp(-\frac{1}{m} s(x-\epsilon) + m \Delta(s))] \Pr[\tilde{A}_m \in (x-\epsilon, x+\epsilon)]$

$\approx \exp(-m(sX - \Delta(s))) \Pr[\tilde{A}_m \in (x-\epsilon, x+\epsilon)]$

choose s such that $\Pr[\tilde{A}_m \in (x-\epsilon, x+\epsilon)] \rightarrow 1$ ($\mathbb{E}_{\tilde{A}_m} X = 0$)



- Some concentration inequalities

Take X_1, \dots, X_n independent, with $M_i = \mathbb{E}X_i$, $\sigma_i^2 = \text{Var}(X_i)$ and $|X_i - M_i| \leq c_i$, $C = \max c_i$

For $T = X_1 + \dots + X_n$. $\Rightarrow \Pr[T - \mathbb{E}T > \beta] < \begin{cases} \exp(-\beta/\sqrt{\sum_i \sigma_i^2}) & 0 \leq \beta \leq \frac{\sqrt{\sum_i \sigma_i^2}}{C} \\ \exp(-\beta/4c) & \beta > \frac{\sqrt{\sum_i \sigma_i^2}}{C} \end{cases}$

$$\Pr[T - \mathbb{E}T > \beta] < \exp(-2\beta^2/\sum_i c_i^2)$$

(Hoeffding)

(Both for sub-Gaussian as well)

proof Using Chernoff-Cramér; need: compute MGF

$$(Bernstein) \quad \mathbb{E} \exp(s(X_i - M_i)) = \sum_{k=0}^{\infty} \frac{s^k \mathbb{E}(X_i - M_i)^k}{k!} \leq (1 + \sum_{k=2}^{\infty} \frac{s^k}{k!} \sigma_i^2)^{C-1} \leq 1 + \frac{s^2 \sigma_i^2}{2} + \frac{s^2 \sigma_i^2}{6} \sum_{k=3}^{\infty} (sc)^{k-2} = 1 + \frac{s^2 \sigma_i^2}{2} \left(1 + \frac{sc}{3} \right) \leq 1 + s^2 \sigma_i^2 \leq \exp(s^2 \sigma_i^2)$$

$(\mathbb{E}(X_i - M_i)^k \leq (\mathbb{E}(X_i - M_i)^2)^{k/2})$

when $s \leq \frac{1}{2c}$

$$(Hoeffding) \quad \mathbb{E} \exp(s(X_i - M_i)) \leq \frac{1}{2} (e^{cs} + e^{-cs}) \leq e^{2cs^2}$$

- Operator norm for random matrices

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

$$\|X\|_2 = \sup_{\|u\|_2=1, \|v\|_2=1} u^T X v$$

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

Claim. $\Pr[\|X\|_2 \geq C(\sqrt{m} + \sqrt{n} + t)] < e^{-t^2}$

($C > 0$ is some constant)

proof. ① Bound $u^T X v$ for fixed u, v

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

$$\Pr[u^T X v \geq c(\sqrt{m} + \sqrt{n} + t)] < \exp\left(-\frac{2c^2(\sqrt{m} + \sqrt{n} + t)^2}{\sum_{i=1}^m \sum_{j=1}^n u_i^2 v_j^2}\right) = \exp(-2c^2(\sqrt{m} + \sqrt{n} + t)^2) \leq \exp(-2c^2(m+n+t)^2)$$

② For $\sum_{i=1}^m u_i^2 = 1$

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

$$\exists P_m \subseteq \mathbb{S}^{m-1}, |P_m| = 2^m, \cup_{x \in P_m} B_{\frac{1}{4}}(x) \geq \mathbb{S}^{m-1}$$

$\Rightarrow \Pr[\max_{u \in P_m} u^T X v \geq c(\sqrt{m} + \sqrt{n} + t)] < e^{-t^2}$ (union bound)

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

$$|u^T X v - \tilde{u}^T X \tilde{v}| \leq |(\tilde{u} - u)^T X v| + |\tilde{u}^T X (v - \tilde{v})|$$

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

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

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

Martingale concentration

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

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

$$\mathbb{E}[Z_1 | Z_2, \dots, Z_n] = Z_n$$

(A_1, A_2, \dots, A_n can depend on Z_1, \dots, Z_{n-1})

Azuma/Hoeffding inequality: $A_i \in \mathbb{R}, B_i \in \mathbb{R}$, $B_i - A_i \leq c_i$ (but c_i is constant)

$$\Pr[\max_{0 \leq i \leq n} Z_i - Z_0 \geq \beta] \leq \exp\left(-\frac{2\beta^2}{\sum_i c_i^2}\right)$$

proof $\Pr[\max_{0 \leq i \leq n} Z_i - Z_0 \geq \beta]$

$$\leq \mathbb{E}[\exp(s(Z_n - Z_0))] e^{-\beta s}$$

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

$$\leq \mathbb{E}[\exp(sZ_\tau)] e^{-\beta s} = \mathbb{E}[\exp(s \sum_{i=1}^{\tau} (Z_i - Z_{i-1}))] e^{-\beta s}$$

Then Chernoff-Cramér.

- Bounded difference estimates:

X_1, X_2, \dots, X_n independent

each $X_i \in \Omega_i$

$f : \Omega_1 \times \dots \times \Omega_n \rightarrow \mathbb{R}$

$$D_i = \max_{x_i \in \Omega_i, x_{i+1}, \dots, x_n \in \Omega_n} |f(x_1, \dots, x_i, \dots, x_n) - f(x_1, \dots, \tilde{x}_i, \dots, x_n)|$$

$x_i \in \Omega_i$

Let $Z_n = \mathbb{E}[f(x_1, \dots, x_n) | x_1, \dots, x_{n-1}]$ (Markov's inequality)

$\Rightarrow \Pr[f(x_1, \dots, x_n) - \mathbb{E}f(x_1, \dots, x_n) \geq \beta] \leq \exp(-\frac{2\beta^2}{\sum_i D_i^2})$

- Example: First Passage Percolation

Last

$w_e \sim \text{Uniform}$, $\forall e \in E$ (edge weights)

$$T_n = \min_{\pi} \sum_e w_e$$

π : up-right path from $(1,1)$ to (n,n)

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

$T_{nm} \leq T_n + T_m$ (sub-additivity)

$\Rightarrow \mathbb{E} T_{nm} \leq \mathbb{E} T_n + \mathbb{E} T_m, \frac{1}{n} \mathbb{E} T_n \text{ converges as } n \rightarrow \infty$

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

By bounded difference estimates,

(X_i : we flip between x_{i-1} and x_i)

$$\Pr[T_n - \mathbb{E} T_n \geq \beta] \leq \exp\left(-\frac{2\beta^2}{4n}\right) \quad (D_i = D_1 = \dots = D_{n-1})$$

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

(KPP universality conjecture: $\text{Var}(T_n)$ is of order $O(n^{\frac{1}{3}})$)

(state of art: $\sqrt{\log n} \leq \text{Var}(T_n) \leq \frac{C}{\sqrt{\log n}}$ (Benjamini-Kalai-Schramm))

Example. Pattern matching

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

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

$N_n = \#\{i : (X_i, \dots, X_{i+k-1}) = a\}$

$$\mathbb{E} N_n = (n-k+1)s^k \leq b \sqrt{n} \leq 2 \exp\left(-\frac{2b^2}{kn}\right) \leq 2e^{-2b^2}$$

$D_i \leq k$

• Chromatic number

Take Erdős-Rényi graph $G(n,p)$

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

$X_i = \#\{(i, j) \in G(n,p)\}$: $1 \leq i \leq \frac{n}{2}$

Claim: change X_i alters X by ≤ 1

($D_i \leq 1$)

$$\Rightarrow \Pr[|X - \mathbb{E}X| \geq b\sqrt{n}] \leq 2 \exp\left(-\frac{2b^2 n}{n-1}\right) = 2e$$