Lecture Notes 11: Probability Inequalities(3)

Professor: Zhihua Zhang *Scribe:*

11 Probability Inequalities(3)

Theorem 11.1. *If $X \in G(v)$, $\mathbb{E}X = 0$, then for any $p > 0$,*

$$\mathbb{E}|X|^p \leq p(2v)^{\frac{p}{2}}\Gamma(\frac{p}{2}).$$

Thus for any integer $q \geq 1$,

$$\mathbb{E}(X^{2q}) \leq 2q!(2v)^q \leq q!(4v)^q$$

Conversely, if for any constant C,

$$\mathbb{E}(X^{2q}) \leq q!C^q,$$

then $X \in G(4C)$.

Let's prove the converse part.

Proof. Let $\hat{X}$ be a copy of X, which means has the same distribution like X and be independent.

$$\begin{aligned}\mathbb{E}e^{\lambda X}\mathbb{E}e^{-\lambda\hat{X}} &= \mathbb{E}e^{\lambda X-\hat{X}} \\ &\quad (\mathbb{E}(X-\hat{X})^{2q+1} = 0 \text{ and} \\ &\quad \text{lebesgue dominated convergence theorem }) \\ &\leq \sum_{q=0}^{\infty}\frac{\lambda^{2q}\mathbb{E}((X-\hat{X})^{2q})}{(2q)!}\end{aligned}$$

By Jensen inequality, $(\frac{1}{2}X+\frac{1}{2}(-\hat{X}))^{2q} \leq \frac{1}{2}X^{2q}+\frac{1}{2}(\hat{X})^{2q}$. So $\mathbb{E}((X-\hat{X})^{2q}) \leq 2^{2q-1}[\mathbb{E}(X^{2q})+\mathbb{E}(\hat{X}^{2q})] = 2^{2q}\mathbb{E}(X^2q)$. Thus

$$\mathbb{E}e^{\lambda X}\mathbb{E}e^{-\lambda\hat{X}} \leq \sum_{q=0}^{\infty}\frac{\lambda^{2q}2^{2q}C^q q!}{(2q)!}$$

And $\frac{(2q)!}{q!} = \Pi_{j=1}^{q}(q+j) \geq 2^q q!$, by Jensen inequality $\mathbb{E}e^{-\lambda\hat{X}} \geq e^{-\lambda\mathbb{E}\hat{X}} = 1$. Thus

$$\begin{aligned}\mathbb{E}e^{\lambda X} &\leq \sum_{q=0}^{\infty}\frac{\lambda^{2q}2^q C^q}{q!} \\ &= e^{2\lambda^2 C} \\ &= e^{\frac{4\lambda^2 C}{2}}\end{aligned}$$

□

Theorem 11.2 (Bernstein's inequality)**.** *Let $X_1, ..., X_n$ be independent real-valued random variables. Assume that there exist positive number ν and c, such that $\sum_{i=1}^n \mathbb{E}[X_i^2] \le \nu$ and $\sum_{i=1}^n \mathbb{E}[(X_i)_+^q] \le \frac{q!\nu c^{q-2}}{2}$ for all integers $q \ge 3$. If $S = (X_i - \mathbb{E}X_i)$, then for all $\lambda \in (0, \frac{1}{c})$ and $t > 0$, $\psi_S(\lambda) \le \frac{\nu\lambda^2}{2(1-c\lambda)}$ and $\psi_S^*(t) \ge \frac{\nu}{c^2} h_1(\frac{ct}{\nu})$, where $h_1(u) = 1 + u - \sqrt{1+2u}$ for $u > 0$. $\mathbb{P}(\{S \ge \sqrt{2\nu t} + ct\}) \le e^{-t}$.*

Proof. Let $\phi(u) = e^u - u - 1$, for $u \le 0$, $\phi(u) \le \frac{u^2}{2}$.

By Taylor's expansion, for $\lambda > 0$, $\phi(\lambda X_i) \le \frac{\lambda^2 X_i^2}{2} + \sum_{q=3}^{\infty} \frac{\lambda^q (X_i)_+^q}{q!}$. Thus

$$\mathbb{E}\phi(\lambda X_i) \le \mathbb{E}(\frac{\lambda^2 X_i^2}{2}) + \sum_{q=3}^{\infty} \frac{\lambda^q \mathbb{E}((X_i)_+^q)}{q!}$$

$$\begin{aligned}
\sum_{i=1}^n \mathbb{E}\phi(\lambda X_i) &\le \sum_{i=1}^n \mathbb{E}(\frac{\lambda^2 X_i^2}{2}) + \sum_{q=3}^{\infty} \frac{\lambda^q}{q!} \sum_{i=1}^n \mathbb{E}((X_i)_+^q) \\
&\le \frac{\nu}{2} \sum_{q=2}^{\infty} \lambda^q c^{q-2} \\
&= \frac{\nu\lambda^2}{2} \sum_{q=0}^{\infty} (\lambda c)^q
\end{aligned}$$

$\lambda \in (0, \frac{1}{c})$, so $\sum_{q=0}^{\infty} (\lambda c)^q = \frac{1}{\lambda c}$

$$\begin{aligned}
\psi_S(\lambda) &= \log \mathbb{E} e^{\lambda S} \\
&= \log \mathbb{E} e^{\lambda \sum_{i=1}^n (X_i - \mathbb{E}X_i)} \\
&= \log \Pi_{i=1}^n \mathbb{E} e^{\lambda (X_i - \mathbb{E}X_i)} \\
&= \sum_{i=1}^n (\log \mathbb{E} e^{\lambda X_i} - \lambda \mathbb{E}X_i) \\
&= \sum_{i=1}^n (\mathbb{E} e^{\lambda X_i} - 1 - \lambda \mathbb{E}X_i) \\
&= \sum_{i=1}^n \mathbb{E}\phi(\lambda X_i) \\
&\le \frac{\nu}{2} \frac{\lambda^2}{1 - c\lambda}
\end{aligned}$$

$$\begin{aligned}
\psi_S^*(t) &= \sup_{\lambda \in (0, \frac{1}{c})} (t\lambda - \psi_S(\lambda)) \\
&\ge \sup_{\lambda \in (0, \frac{1}{c})} (t\lambda - \frac{\lambda^2 \nu}{2(1 - c\lambda)}) \\
&= \frac{v}{c^2} h_1(\frac{ct}{\nu})
\end{aligned}$$

h_1 is an increasing funciton from $(0,\infty)$ to $(0,\infty)$. $h^{-1}(u) = u + \sqrt{2u}$. $\psi^*(t) = \frac{\nu}{c^2} h_1(\frac{ct}{\nu})$ and $\psi^{*-1}(t) = ct + \sqrt{2\nu t}$. Thus

$$\mathbb{P}(S \geq t) \leq \exp(-\frac{\nu}{c^2} h_1(\frac{ct}{\nu}))$$

$$\mathbb{P}(S \geq \psi^{*-1}(t)) \geq \exp(-\psi^*(\psi^{*-1}(t))) = \exp(-t)$$

Then $\mathbb{P}(S \geq ct + \sqrt{2\nu t}) \leq \exp(-t)$. $\square$

Corollary 11.1. *Let $X_1, X_2, ..., X_n$ be independent real-valued random variables, satisfying the condition of theorem, and $S = \sum_{i=1}^n (X_i - \mathbb{E}X_i)$. Then for all $t > 0$.*

$$\mathbb{P}(S \geq t) \leq \exp(-\frac{t^2}{2(v + ct)})$$

11.1 Random projection and Johnson-Lindenstrauss lemma

Let $U = [u_1, ..., u_p]^T \in \mathbb{R}^p$, $R\mathbb{R}^{p\times d}$, $R^T R = I_d$. Let $V = \sqrt{\frac{p}{d}} R^T U$. Then $\mathbb{E}(||V||_2^2) = ||U||_2^2$.

Proof.

$$R = \begin{bmatrix} -R_{(1)}^T- \\ \vdots \\ -R_{(p)}^T- \end{bmatrix} = \begin{bmatrix} | & & | \\ R_1 & ... & R_d \\ | & & | \end{bmatrix}$$

$\sum_{i=1}^p R_{(i)}^T R_{(i)} = \mathbb{E}[tr(RR^T)] = \mathbb{E}[tr(R^T R)] = d$, then $R_{(i)}^T R_{(i)} = \frac{d}{p}$. And $R_i^T R_j = 0 \implies R_{(i)}^T R_{(j)} = 0$. So

$$\mathbb{R}(V^T V) = U^T \mathbb{E}(RR^T) U = \frac{d}{p} U^T U = \frac{d}{p} ||U||_2^2$$

$\square$

Definition 11.1. *$f : U \in \mathbb{R}^p \to V \in \mathbb{R}^d$, if $(1-\epsilon)||x-y||^2 \leq ||f(x) - f(y)||^2 \leq (1+\epsilon)||x-y||^2$. It is called $\epsilon-$isometric.*

Lemma 11.1. *Each entry of a $p \times d$ matrix R be chosen independently from $N(0,1)$. Let $V = \frac{1}{\sqrt{d}} R^T U$ for $U \in \mathbb{R}^p$. Then for any $\epsilon > 0$,*

1. $\mathbb{E}(||V||^2) = ||U||^2$
2. $\mathbb{P}(|||V||^2 - ||U||^2| \geq \epsilon ||U||^2) \leq 2\exp(-(\epsilon^2 - \epsilon^3)\frac{d}{4})$

Proof. $\mathbb{E}(||V||^2) = \frac{1}{d}\mathbb{E}(U^T RR^T U) = \frac{1}{d} U^T \mathbb{E}(RR^T) U$. $R_{(i)}^T R_{(i)} \sim \chi^2(d)$, then $\mathbb{E}(R_{(i)}^T R_{(i)}) = d$. $\mathbb{E}(R_{(i)}^T R_{(j)}) = Cov(R_{(i)}^T, R_{(j)}^T) = 0$. Then $\mathbb{E}(||V||^2) = ||U||^2$.

Let $X = \frac{d}{||U||^2}||V||^2 = \frac{||R^TU||^2}{||U||^2} = \frac{\sum_{j=1}^d (R_jU)^2}{||U||^2} = \sum_{j=1}^d X_j^2$, where $X_j = \frac{R_j^TU}{||U||}$. Since $R_j \sim N(0, I_p)$, so $X_i \sim N(0,1)$.

$$
\begin{aligned}
\mathbb{P}(||V||^2 \geq (1+\epsilon)||U||^2) &= \mathbb{P}(X \geq (1+\epsilon)d) \\
&= \mathbb{P}(e^{\lambda X} \geq e^{\lambda(1+\epsilon)d}) \\
&\leq \frac{\mathbb{E}e^{\lambda X}}{e^{\lambda(1+\epsilon)d}} \\
&= \frac{\mathbb{E}e^{\lambda \sum_{i=1}^d X_i^2}}{e^{\lambda(1+\epsilon)d}} \\
&= \frac{\Pi_{i=1}^d \mathbb{E}e^{\lambda X_i^2}}{e^{\lambda(1+\epsilon)d}} \\
&= \left(\frac{\mathbb{E}e^{\lambda X_i^2}}{e^{\lambda(1+\epsilon)}}\right)^d
\end{aligned}
$$

Similarly, $\mathbb{P}(||V||^2 \leq (1-\epsilon)||U||^2) \leq \left(\frac{\mathbb{E}e^{-\lambda X_i^2}}{e^{-(1-\epsilon)\lambda}}\right)^d$

$\mathbb{E}e^{\lambda X_i^2} = \int e^{\lambda x_i^2} \frac{1}{\sqrt{2\pi}} e^{-\frac{x_i^2}{2}} dx_i = \frac{1}{\sqrt{1-2\lambda}}$. Thus

$$
\mathbb{P}(X \geq (1+\epsilon)d) \leq \left(\frac{e^{-2(1+\epsilon)\lambda}}{1-2\lambda}\right)^{\frac{d}{2}}
$$

where $\lambda \in (0, \frac{1}{2})$. Let $\lambda = \frac{\epsilon}{2(1+\epsilon)}$, then $\mathbb{P}(X \geq (1+\epsilon)d) \leq ((1+\epsilon)e^{-\epsilon})^{\frac{d}{2}}$. And $1+\epsilon < e^{\epsilon - \frac{\epsilon^2-\epsilon^3}{2}}$. So $\mathbb{P}(X \geq (1+\epsilon)d) \leq e^{-(\epsilon^2-\epsilon^3)\frac{d}{4}}$. Thus $\mathbb{P}(|||V||^2 - ||U||^2| \geq \epsilon||U||^2) \leq 2\exp(-(\epsilon^2-\epsilon^3)\frac{d}{4})$ $\square$

Theorem 11.3 (John-Lidenstrauss lemma)**.** *$R : \mathbb{R}^p \to \mathbb{R}^d$. Let A be finite subset of $\mathbb{R}^p$ with $|A| = n$. Assume that for some $\nu \geq 1$, $R_{ij} \in G(\nu)$, and $\mathbb{E}R_{ij} = 0$, $Var(R_{ij}) = 1$. and let $\epsilon, \delta \in (0,1)$. If $d \geq 100\nu^2\epsilon^{-2}\log(\frac{n}{\sqrt{\delta}})$. There with probability at least $1-\delta$,*

$$
(1-\epsilon)||X-Y||^2 \leq ||R^TX - R^TY||^2 \leq (1+\epsilon)||X-Y||^2
$$

for $X, Y \in A$. It is calle ϵ-isometric on A.

Proof. Let S be the unit sphere of $\mathbb{R}^p$ and let T be the subset of S defined by $T = \{\frac{X-Y}{||X-Y||} : X, Y \in A, X \neq Y\}$. Let $|A| = n$, then $|T| = N \leq \frac{n(n-1)}{2}$.

$$
\mathbb{P}\left(||R^T\frac{(X-Y)}{||X-Y||}||_2^2 - 1 \leq \epsilon\right) \leq \mathbb{P}\left(\sup_{\alpha \in T}|R^T\alpha - 1| \leq \epsilon\right)
$$

Let $Z_i(\alpha) = \sum_{j=1}^p \alpha_j R_{ji}$, so $\mathbb{E}(Z_i(\alpha)^2) = Var(Z_i(\alpha)) = \sum_{j=1}^p \alpha_j^2 Var(X_{ji}) = 1$.

$$\begin{aligned}
\mathbb{E}\exp(\lambda Z_i(\alpha)) &= \mathbb{E}\exp(\lambda \sum_{j=1}^p \alpha_j X_{ji}) \\
&= \Pi_{j=1}^p \mathbb{E}\exp(\lambda \alpha_j X_{ji}) \\
&\le \exp(\sum_{j=1}^p \frac{\lambda^2 \alpha_j^2 \nu}{2}) \\
&= \exp(\frac{\lambda^2 \nu}{2})
\end{aligned}$$

Since $\mathbb{E}(Z_i(\alpha)^{2q}) \le \frac{q!}{2} 4(2\nu)^q \le \frac{q!}{2}(4\nu)^q$, so $\sum_{j=1}^d \mathbb{E}(Z_i(\alpha)^{2q}) \le \frac{q! d (4\nu)^q}{2}$.
So according to Bernstein's inequality, $\nu \leftarrow (4\nu)^2 d$, $c \leftarrow 4\nu$, thus

$$\mathbb{P}(|\sum_{i=1}^d (Z_i(\alpha)^2 - 1)| \ge 4\nu\sqrt{2dt} + 4\nu t) \le 2e^{-t}$$

Then

$$\mathbb{P}\left(\sup_{\alpha \in T} \sum_{i=1}^d (Z_i(\alpha)^2 - 1) \ge 4\nu\sqrt{2dt} + 4\nu t\right) \le |T| 2e^{-t} \le n^2 e^{-t}$$

Let $t = \log \frac{n^2}{\delta}$, so $4\nu\sqrt{2dt} + 4\nu t \le \epsilon$. Thus

$$\mathbb{P}\left(\sup_{\alpha \in T} \sum_{i=1}^d (Z_i(\alpha)^2 - 1) \ge \epsilon\right) \le \delta$$

$\square$

Homework

1. Prove the following one-sided improvement of Chebyshev's inequality: for any real-valued random variable Y and $t > 0$, $\mathbb{P}(Y - \mathbb{E}Y \ge t) \le \frac{Var(Y)}{Var(Y)+t^2}$
2. Show that if Y is nonnegative random variable then for any $\alpha \in (0,1)$, $\mathbb{P}(Y \ge \alpha \mathbb{E}Y) \ge (1-\alpha)^2 \frac{(\mathbb{E}Y)^2}{\mathbb{E}(Y^2)}$
3. Prove that if Z has a centered normal random variable with variance σ^2 then $\sup_{t>0} \mathbb{P}(\{Z \ge t\})$