Lecture Notes 10: Probability Inequalities(2)

Professor: Zhihua Zhang *Scribe:*

10 Probability Inequalities(2)

10.1 Hoeffding's Inequality

If X_1, ... , X_n are indenpendent random variables with a finite mean value such that for some non-empty interval I, $\mathbb{E}e^{\lambda X_i}$ is finite. Then we define

$$S = \sum_{i=1}^{n}(X_i - \mathbb{E}X_i)$$

. Assume X_i takes its values in a bounded interval $[a_i, b_i]$. Then

$$\mathbb{P}(S \geq t) \leq \exp\left(-\frac{2t^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right)$$

for all $t > 0$.

Lemma 10.1. ***(Hoeffding's Lemma)*** *Let Y be a random variable with $\mathbb{E}Y = 0$, taking values in a bounded interval $[a, b]$ and let $\psi_Y(\lambda) = \log \mathbb{E}e^{\lambda Y}$. Then $\psi''_Y(\lambda) \leq \frac{(b-a)^2}{4}$ and $\psi_Y(\lambda) \leq \frac{\lambda^2}{2}\frac{(b-a)^2}{4}$.*

We first show that if Hoeffding's lemma is true, then Hoeffding's inequality is also true.

Proof. (Hoeffding's Lemma $\rightarrow$ Hoeffding's Inequality)

Let $Y_i = X_i - \mathbb{E}X_i$. So $\mathbb{E}Y_i = 0$,$S = \sum_{i=1}^{n} Y_i$ and $Y_i \in [a_i - \mathbb{E}Y_i, b_i - \mathbb{E}Y_i]$.

$$\begin{aligned}
\Pr(S \geq t) &\leq \frac{\mathbb{E}e^{\lambda S}}{e^{\lambda t}} \\
&= \exp(-\lambda t + \log \mathbb{E}e^{\lambda S}) \\
&(Y_i' s \text{ are independent}) \\
&= \exp(-\lambda t + \log \Pi_{i=1}^{n} \mathbb{E}e^{\lambda Y_i}) \\
&= \exp(-\lambda t + \sum_{i=1}^{n} \log \mathbb{E}e^{\lambda Y_i}) \\
&\leq \exp(-\lambda t + \frac{\lambda^2}{2}\sum_{i=1}^{n}\frac{(b_i - a_i)^2}{4}) \\
&\leq \exp(-\frac{2t^2}{\sum_{i=1}^{n}}(b_i - a_i)^2)
\end{aligned}$$

$\square$

Next, we will give two proofs for Hoeffding's Lemma.

Proof. Let $g(y) = \exp(-\psi_Y(\lambda))e^{\lambda y} f(y)$, then we have

$$\int g(y)dy = \frac{\int e^{\lambda y} f(y)dy}{\exp(\psi_Y(\lambda))} = 1$$

Because $Y \in [a, b]$, we can get $|Y - \frac{a+b}{2}| \leq \frac{b-a}{a}$. So $Var(Y) = \mathbb{E}((Y - \frac{a+b}{2})^2) \leq \frac{(b-a)^2}{4}$. And we have

$$\begin{aligned}
\mathbb{E}(Y - \mathbb{E}_g Y)^2 &= \int (y^2 - 2y\mathbb{E}_g Y + \mathbb{E}_g^2(Y))g(y)dy \\
&= \exp(-\psi_Y(\lambda)) \int y^2 e^{\lambda y} f(y)dy - 2\exp(-2\psi_Y(\lambda)) \int y e^{\lambda y} f(y) \int y e^{\lambda y} f(y)dydy + \mathbb{E}_g^2(Y) \\
&= \exp(-\psi_Y(\lambda)) \int y^2 e^{\lambda y} f(y)dy - 2[\exp(-\psi_Y(\lambda))]^2 \left(\int y e^{\lambda y} f(y)dy \right)^2 + \\
&\quad [\exp(-\psi_Y(\lambda))]^2 \left(\int y e^{\lambda y} f(y)dy \right)^2 \\
&= \exp(-\psi_Y(\lambda)) \int y^2 e^{\lambda y} f(y)dy - [\exp(-\psi_Y(\lambda))]^2 \left(\int y e^{\lambda y} f(y)dy \right)^2 \\
&= \exp(-\psi_Y(\lambda))\mathbb{E}_f(Y^2 e^{\lambda Y}) - \exp(-2\psi_Y(\lambda))(\mathbb{E}_f(Y e^{\lambda Y}))^2 \\
&\leq \frac{(b-a)^2}{4}
\end{aligned}$$

By Taylor's expansion, we have

$$\psi_Y(\lambda) = \psi_Y(0) + \lambda\psi_Y'(0) + \frac{\lambda^2}{2}\psi_Y''(\theta)$$

where $\theta \in (0, \lambda)$. And $\psi_Y(0) = \psi_Y'(0) = 0$ and $\psi_Y''(\lambda) = \frac{\mathbb{E}_f(Y^2 e^{\lambda Y})}{\mathbb{E}_f e^{\lambda Y}} - \frac{[\mathbb{E}_f(Y e^{\lambda Y})]^2}{\mathbb{E}_f e^{2\lambda Y}} \leq \frac{(b-a)^2}{4}$. So,

$$\begin{aligned}
\psi_Y(\lambda) &= \frac{\lambda^2}{2}\psi_Y''(\theta) \\
&\leq \frac{\lambda^2}{2}\psi_Y''(\lambda) \\
&\leq \frac{\lambda^2}{2}\frac{(b-a)^2}{4}
\end{aligned}$$

□

Let X be any real-valued random variable with expected value $\mathbb{E}X = 0$ and such that $a \leq X \leq b$ almost surely. Then, for all $\lambda \in \mathbb{R}$,

$$\mathbb{E}[e^{\lambda X}] \leq \exp(\frac{\lambda^2(b-a)^2}{8})$$

Proof. Since $e^{\lambda x}$ is a convex function, we have

$$e^{\lambda x} \leq \frac{b-x}{b-a} e^{\lambda a} + \frac{x-a}{b-a} e^{\lambda b}, \forall a \leq x \leq b$$

So,

$$\mathbb{E}[e^{\lambda X}] \leq \frac{b - \mathbb{E}X}{b-a} e^{\lambda a} + \frac{\mathbb{E}X - a}{b-a} e^{\lambda b}.$$

Let $\alpha = \frac{-a}{b-a} \in [0,1]$, $u = \lambda(b-a)$, and $L(u) = -\alpha u + \ln(1 - \alpha + \alpha e^u)$

Then $\frac{b-\mathbb{E}X}{b-a} e^{\lambda a} + \frac{\mathbb{E}X-a}{b-a} e^{\lambda b} = e^{L(u)}$ since $\mathbb{E}X = 0$. Taking derivative of $L(u)$,

$$L(0) = L'(0) = 0 \text{ and } L''(h) \leq \frac{1}{4}$$

By Taylor's expansion,

$$L(u) \leq \frac{1}{8} u^2 = \frac{1}{8} \lambda^2 (b-a)^2$$

Hence, $\mathbb{E}[e^{\lambda X}] \leq e^{\frac{1}{8}\lambda^2(b-a)^2}$ □

10.2 Bennett's Inquality

Let X_1, ..., X_n be independent random variables with finite variance such that $X_i \leq b$ for some $b > 0$ almost surely for $i \leq n$. And let $v = \sum_{i=1}^n \mathbb{E}(X_i^2)$. Assume $\psi(u) = e^u - u - 1$. For $u \in \mathbb{R}$, and $\forall \lambda > 0$.

$$\psi_S(\lambda) = \log \mathbb{E} e^{\lambda S} \leq n \log(1 + \frac{v}{nb^2} \psi(b\lambda)) \leq \frac{v}{b^2} \psi(b\lambda).$$

And for any $t > 0$,

$$\mathbb{P}(S \geq t) \leq \exp(-\frac{v}{b^2} h(\frac{bt}{v}))$$

where $h(u) = (1+u)\log(1+u) - u$ for $u > 0$.

Theorem 10.1. ***(Bernstein's Inequality)*** *Let $h(x) = (1+x)\log(1+x) - x$ for $x \geq 0$. And $h(x) \geq \frac{x^2}{2(1+\frac{x}{3})} = g(x)$. So,*

$$\mathbb{P}(S \geq t) \leq \exp\left(-\frac{t^2}{2(v + \frac{bt}{3})}\right)$$

Proof. $h'(x) = \log(1+x)$ and $h''(x) = \frac{1}{x+1}$

$g'(x) = \frac{x}{1+\frac{x}{3}} - \frac{x^2}{6(1+\frac{x}{3})^2}$ and $g''(x) = \frac{27}{(x+3)^3}$

It's easy to see that $h^{(n)}(0) \geq g^{(n)}(0)$. □

Lemma 10.2. ***(Johnson-Lindenstrauss Lemma)*** *Given $0 < \epsilon < 1$, a set X of m points in $\mathbb{R}^N$, and a number $n > 8\ln(m)/\epsilon^2$, there is a linear map $f : \mathbb{R}^N \to \mathbb{R}^n$ such that*

$$(1-\epsilon)||u-v||^2 \leq ||f(u) - f(v)||^2 \leq (1+\epsilon) \leq (1+\epsilon)||u-v||^2$$

for all $u, v \in X$.

Definition 10.1. ***(Sub-Gaussian Random Variables)*** *A real-valued random variable X is said to be subgaussian if it has the property that there is some $b > 0$ such that for every $\lambda \in \mathbb{R}$ one has*

$$\mathbb{E}e^{\lambda X} \leq e^{\lambda^2 t^2/2}$$

We denote that $X \in G(t^2)$.

Proposition 10.1. *If $X \in G(t^2)$, then*

1. $\mathbb{E}X = 0$
2. $Var(X) \leq t^2$

Proof. Using Taylor's expansion for the exponential function

$$\mathbb{E}\sum_{n=0}^{\infty} \frac{\lambda^n x^n}{n!} = \mathbb{E}e^{\lambda X}$$

and Lebesgue's Dominated Convergence Theorem, for any $\lambda \in \mathbb{R}$,

$$\sum_{n=0}^{\infty} \frac{\lambda^n}{n!}\mathbb{E}(X^n) = \mathbb{E}e^{\lambda X} \leq e^{\lambda^2 t^2/2} = \sum_{n=0}^{\infty} \frac{\lambda^{2n} t^{2n}}{2^n n!}$$

Thus

$$\lambda \mathbb{E}X + \frac{\lambda^2}{2}\mathbb{E}X^2 \leq \frac{\lambda^2 t^2}{2} + o(\lambda^2) \qquad \text{as } \lambda \to 0$$

Dividing through by $\lambda > 0$ ans letting $\lambda \to 0$ we get $\mathbb{E}(X) \leq 0$. Dividing through by $\lambda < 0$ and letting $\lambda \to 0$ we get $\mathbb{E}(X) \geq 0$. Thus $\mathbb{E}(X) = 0$. Now that this is established, we divided through by λ^2 and let $\lambda \to 0$, thus getting $Var(X) \leq t^2$. □

Example 10.1. *Let X be a random variable with the Rademacher distribution, meaning that the law of X if $\mathbb{P}X = \frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_1$[here δ_t is the point mass at x]. Then for any $t \in \mathbb{R}$,*

$$BEe^{tX} = \frac{1}{2}e^{-t} + \frac{1}{2}e^t = cosht \leq e^{t^2/2}$$

. So $X \in G(1)$.