Tail Inequalities

Markov, Chebyshev, Chernoff, SubGaussian, Hoeffding, Maximal Inequalities

Sun Woo Lim

Dec 2023

Outline

Markov and Chebyshev inequalities

Markov's Inequality

For a nonnegative random variable X with finite mean μ, $\mathbb{P}(X \geq t) \leq \frac{\mathbb{E}(X)}{t}$, $\forall t \geq 0$.

pf) Let $f(\cdot)$ denote the density function of X.

$\mathbb{E}X = \int_0^t xf(x)dx + \int_t^\infty xf(x)dx \geq \int_t^\infty xf(x)dx \geq \int_t^\infty tf(x)dx = t\mathbb{P}(X \geq t)$.

Chebyshev's Inequality

For a random variable X with finite variance, $\mathbb{P}(\mid X - \mu \mid \geq t) \leq \frac{Var(X)}{t^2}, \forall t \geq 0$.

pf) $\mathbb{P}(\mid X - \mu \mid \geq t) = \mathbb{P}\left((X-\mu)^2 \geq t^2\right) \leq \frac{\mathbb{E}(X-\mu)^2}{t^2} = \frac{Var(X)}{t^2}$ by Markov's inequality.

Polynomial Markov

Whenever X has a central moment of order k, $\mathbb{P}(\mid X - \mu \mid \geq t) \leq \frac{\mathbb{E}|X-\mu|^k}{t^k}, \forall t \geq 0$.

If central moment of X exists for all $k \in \mathbb{N}$, $\mathbb{P}(\mid X - \mu \mid \geq t) \leq \inf_{k \in 0,1,\dots} \frac{\mathbb{E}|X-\mu|^k}{t^k}, \forall t \geq 0$.

Implications

Once we have information of existence of the central moments, we can upper bound the tail bounds.

However, the optimization problem in **polynomial Markov** is not practical because it is an integer optimization.

Chernoff Bound

Suppose X has a MGF in a neighborhood with some radius b around zero.
In other words, $\exists b > 0$ s.t. $\mathbb{E}e^{\lambda(X-\mu)} < \infty, \forall |\lambda| \geq b$.
Then, for all $\lambda \in [0, b)$, apply Markov Inequality for $Y = e^{\lambda(X-\mu)}$:

$$\mathbb{P}\left((X-\mu) \geq t\right) = \mathbb{P}\left(\lambda(X-\mu) \geq \lambda t\right) = \mathbb{P}\left(e^{\lambda(X-\mu)} \geq e^{\lambda t}\right) \leq \frac{\mathbb{E}e^{\lambda(X-\mu)}}{e^{\lambda t}}. \tag{1}$$

Since (1) holds for all $\lambda \in [0, b)$, we can optimize (minimization) λ to obtain the following Chernoff bound

$$\log \mathbb{P}\left((X-\mu) \geq t\right) \leq \inf_{\lambda \in [0,b)} \left(\log \mathbb{E}_X e^{\lambda(X-\mu)} - e^{\lambda t}\right). \tag{2}$$

Remarks

1. For $X \geq 0$, for any $\delta > 0$ and $k \in 0, 1, \ldots$, the following inequality implies that if attainable, the optimized bound from Polynomial Markov is more useful than that from Chernoff.

$$\inf_{k=0,1,\ldots} \frac{\mathbb{E}X^k}{\delta^k} \leq \inf_{\lambda>0} \frac{\mathbb{E}e^{\lambda X}}{e^{\lambda\delta}}$$

2. Chernoff bound characterized by MGF yields useful tail bounds on probabilities of the original variable (see next slides)

Outline

SubGaussian Random Variables

Motivation: Tail of Gaussian

Suppose $X \sim \mathsf{N}(\mu, \sigma^2)$. It is known that $\mathbb{E}e^{\lambda(X-\mu)} = e^{\sigma^2\lambda^2/2}, \forall \lambda \in \mathbb{R}$. By Chernoff argument,

$$\log \mathbb{P}(X-\mu \geq t) \leq \inf_{\lambda \geq 0} \left(\log \mathbb{E}(e^{\lambda(X-\mu)}) - \lambda t \right) = \inf_{\lambda \geq 0} \left(\frac{\sigma^2\lambda^2}{2} - \lambda t \right) = -t^2/2\sigma^2, \tag{3}$$

which gives

$$\mathbb{P}(X-\mu \geq t) \leq \exp\left(-\frac{-t^2}{2\sigma^2} \right), \quad \forall t \geq 0. \tag{4}$$

Similarly, one can obtain $\mathbb{P}(X-\mu \leq -t) = \mathbb{P}(-X+\mu \geq t) \leq \exp\left(-\frac{-t^2}{2\sigma^2} \right)$, using $-X+\mu \sim \mathsf{N}(0, \sigma^2)$.
By union bound (: $\mathbb{P}(A \text{ or } B) \leq \mathbb{P}(A) + \mathbb{P}(B)$),

$$\mathbb{P}(\mid X-\mu \mid \geq t) \leq 2\exp\left(-\frac{t^2}{2\sigma^2} \right), \quad \forall t \geq 0. \tag{5}$$

Note) The bound in (5) also holds for random variable X satisfying $\mathbb{E}e^{\lambda(X-\mu)} \leq e^{\sigma^2\lambda^2/2}, \forall \lambda \in \mathbb{R}$, motivating SubGaussian random variables.

SubGaussian Random Variables

Def) A random variable X is subGaussian if $\exists \sigma > 0$ such that $\mathbb{E}e^{\lambda(X-\mu)} \leq e^{\sigma^2\lambda^2/2}, \forall \lambda \in \mathbb{R}$.
Note that $-X$ is subGaussian iff X is subGaussian, implying the tail bound (5).

Equivalent definition of subGaussian

The following statements are equivalent.

1. $\exists \sigma > 0$ such that $\mathbb{E}e^{\lambda(X-\mu)} \leq e^{\sigma^2\lambda^2/2}, \forall \lambda \in \mathbb{R}$.
2. $\exists c > 0$ and $Z \sim \mathrm{N}(0, t^2)$ such that $\mathbb{P}(|X| \geq s) \leq c\mathbb{P}(|Z| \geq s), \quad \forall s \geq 0$.
3. $\exists \theta \geq 0$ such that $\mathbb{E}X^{2k} \leq \frac{(2k)!}{2^k k!}\theta^{2k}, \forall k \in \mathbb{N}$.
4. $\exists \sigma > 0$ such that $\mathbb{E}e^{\frac{\lambda X^2}{2\sigma^2}} \leq \frac{1}{\sqrt{1-\lambda}}$.

Examples of non-Gaussian SubGaussian random variables

1. Rademaker random variable $\in \mathsf{SG}(\sigma^2 = 1)$.

A Rademaker random variable ϵ takes values $\{1, -1\}$ with probability 0.5 each.

Note that $\mathbb{E}e^{\lambda\epsilon} = \frac{1}{2}e^{\lambda} + \frac{1}{2}e^{-\lambda} = \frac{1}{2}\left(\sum_{k=0}^{\infty}\frac{(-\lambda)^k}{k!} + \sum_{k=0}^{\infty}\frac{\lambda^k}{k!}\right) = \sum_{k=0}^{\infty}\frac{\lambda^{2k}}{(2k)!} \leq 1 + \sum_{k=0}^{\infty}\frac{\lambda^{2k}}{2^k k!} = e^{\frac{\lambda^2}{2}}$

2. Bounded random variable

Let X be a zero-mean RV with support $[a, b]$ almost surely. Then, $X \in \mathsf{SG}((b-a)^2)$.

pf) Let X' be an iid copy of X. Note that

$$\mathbb{E}_X e^{\lambda(X-0)} = \mathbb{E}_X e^{\lambda(X - \mathbb{E}_{X'}(X'))} \leq \mathbb{E}_{X,X'} e^{\lambda(X-X')}$$

by Jensen's inequality (exponential is convex). Also, observe that $X - X' \stackrel{d}{=} \epsilon(X - X')$, since X is symmetric around zero. Thus,

$$\begin{aligned}
\mathbb{E}_X e^{\lambda X} \leq \mathbb{E}_{X,X'} e^{\lambda(X-X')} &= \mathbb{E}_{X,X'}\mathbb{E}_{\epsilon} e^{\epsilon\lambda(X-X')|X,X'} \\
&\leq \mathbb{E}_{X,X'} e^{\frac{\lambda^2(X-X')^2}{2}} \\
&\leq e^{\frac{\lambda^2(b-a)^2}{2}}
\end{aligned}$$

Note) One can advance $X \in \mathsf{SG}\left(\left(\frac{b-a}{2}\right)^2\right)$: Hoeffding's lemma.

Properties of SubGaussian random variables

1. If $X \in \mathsf{SG}(\sigma^2), \mathsf{Var}(X) \leq \sigma^2$.

   Pf) Since $\mathbb{E}e^{\lambda(X-\mu)} \leq e^{\frac{\lambda^2\sigma^2}{2}}, \forall\lambda \in \mathbb{R}$, by Taylor series expansion of second order,

$$1 + \lambda E(X-\mu) + \frac{\lambda^2}{2}\mathbb{E}(X-\mu)^2 + o(\lambda^2) \leq 1 + \frac{\lambda^2\sigma^2}{2} + o(\lambda^2).$$

   Note that $E(X-\mu) = 0$. Divide both sides by λ^2 and take $\lambda \to 0$.

2. If $a \leq X - \mu \leq b$ almost surely, $X \in \mathsf{SG}\left(\left(\frac{b-a}{2}\right)^2\right)$: Hoeffding's lemma.

3. If $X \in \mathsf{SG}(\sigma^2)$ and $Y \in \mathsf{SG}(\tau^2)$,
   1. $\alpha X \in \mathsf{SG}(\alpha^2\sigma^2), \quad \alpha \in \mathbb{R}$.
   2. $X + Y \in \mathsf{SG}((\sigma+\tau)^2)$.

      Pf) WLOG, assume that $\mathbb{E}X = \mathbb{E}Y = 0$. Then,

$$\begin{aligned}\mathbb{E}e^{\lambda(X+Y)} &\leq \left(\mathbb{E}e^{\lambda pX}\right)^{1/p}\left(\mathbb{E}e^{\lambda qY}\right)^{1/q} \\ &\leq e^{\frac{\lambda^2p^2\sigma^2}{2}\times\frac{1}{p}+\frac{\lambda^2q^2\tau^2}{2}\times\frac{1}{q}} = e^{\frac{\lambda^2}{2}(p\sigma^2+q\tau^2)} = e^{\frac{\lambda^2}{2}(\sigma+\tau)^2}\end{aligned}$$

      Take $p = \tau/\sigma + 1$ and $q = \sigma/\tau + 1$, to apply Holder's inequality.
   3. If X and Y are independent, $X + Y \in \mathsf{SG}(\sigma^2 + \tau^2)$.

SubGaussian random vector

Def)
A random vector $\mathbf{X} \in \mathbb{R}^d$ is SubGaussian random vector with variance proxy σ^2
if it is centered and for any $u \in \mathbb{R}^d$ such that $\|u\|_2 = 1$, $u^T\mathbf{X} \in \mathsf{SG}(\sigma^2)$.

Ex)
Let $X_1, \ldots, X_d$ be independent $\mathsf{SG}(\sigma^2)$ random variables. Let $\mathbf{X} = (X_1, \ldots, X_d)$ is subGaussian random vector with variance proxy σ^2.
Pf) $\forall u \in \mathbb{R}^d$ with $\|u\|_2 = 1$, and $\lambda \in \mathbb{R}$,

$$\mathbb{E}e^{\lambda u^T \mathbb{X}} = \prod_{i=1}^{d} \mathbb{E}e^{\lambda u_i X_i} \leq \mathbb{E}e^{\sigma^2\lambda^2 u_i^2/2} = e^{\sigma^2\lambda^2 \sum_{i=1}^{d} u_i^2/2} = e^{\sigma^2\lambda^2/2}.$$

Outline

Hoeffding's inequality

For $X_i, i = 1, \dots, n$ are independent and each $X_i \in \mathsf{SG}(\sigma_i^2)$. Then, $\forall t \geq 0$,

$$\mathbb{P}\left(\mid \frac{1}{n} \sum_{i=1}^{n} (X_i - \mu_i) \mid \geq t \right) \leq 2 \exp\left(- \frac{n^2 t^2}{2 \sum_{i=1}^{n} \sigma_i^2} \right)$$

Pf) It suffices to show that $\frac{1}{n} \sum_{i=1}^{n} (X_i - \mu_i) \sim \mathsf{SG}\left(\frac{1}{n^2} \sum_{i=1}^{n} \sigma_i^2 \right)$ and then apply (5).
Use SG properties of independent sum and scalar multiples.

Ex) Bernoulli inequalities
Let $X_1, \dots, X_n$ be independent Ber(p_i) random variables. Then, $X_i \in \mathsf{SG}(1/4)$ by Hoeffding's inequality. Thus,

$$\mathbb{P}\left(\mid \frac{1}{n} \sum_{i=1}^{n} (X_i - p_i) \mid \geq t \right) \leq 2 \exp(-2nt^2), \quad \forall t \geq 0.$$

Maximal Inequalities

Suppose $X_1, \ldots, X_n$ be zero mean SG(σ^2) random variables. They are not necessarily independent. Then,

$$\begin{aligned} \mathbb{E} \max_{i=1,\ldots,n} X_i &\leq \sigma\sqrt{2\log n}, \\ \mathbb{P}(\max_{i=1,\ldots,n} X_i \geq t) &\leq ne^{-\frac{t^2}{2\sigma^2}}, \forall t \geq 0. \end{aligned} \tag{6}$$

Also,

$$\begin{aligned} \mathbb{E} \max_{i=1,\ldots,n} |X_i| &\leq \sigma\sqrt{2\log(2n)}, \\ \mathbb{P}(\max_{i=1,\ldots,n} |X_i| \geq t) &\leq 2ne^{-\frac{t^2}{2\sigma^2}}, \forall t \geq 0. \end{aligned} \tag{7}$$