Chapter 2

Sub Gaussian random variables

Sources for this chapter, Philippe Rigollet and Jan-Christian Hütter lectures notes on high dimensional statistics (Chapter 1). Sudeep Kamath's course on concentration of measure at CIRM (2016). *Subgaussian random variables: An expository note* from Omar Rivasplata.

2.1 Introduction and characterization

2.1.1 Gaussian concentration

The centered Gaussian random variable X on $\mathbb{R}$ with variance $\sigma^2 > 0$ has density given by

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-x^2}{2\sigma^2}\right).$$

It plays a central role in statistics due to the central limit theorem. It also has a central position in statistical and signal processing estimation problems. Important properties of this distribution is closure under addition of iid replicates and concentration (Mill's inequality), if X is $\mathcal{N}\left(0, \sigma^2\right)$, we have, for any $t > 0$, $\mathbb{P}\left(|X| \geq t\right) \leq \frac{\sigma\sqrt{2}}{t\sqrt{\pi}} \exp\left(\frac{-t^2}{2\sigma^2}\right)$.

Proof.

$$\begin{aligned}
\mathbb{P}\left(|X| \geq t\right) &\leq 2\mathbb{P}\left(X \geq t\right) && \text{(symmetry and union bound)} \\
&= \frac{\sqrt{2}}{\sqrt{\pi\sigma^2}} \int_t^{+\infty} \exp\left(\frac{-x^2}{2\sigma^2}\right) dx \\
&\leq \frac{\sigma^2\sqrt{2}}{\sqrt{\pi\sigma^2}} \int_t^{+\infty} \frac{x}{\sigma^2 t} \exp\left(\frac{-x^2}{2\sigma^2}\right) dx \\
&= \frac{\sigma\sqrt{2}}{t\sqrt{\pi}} \int_t^{+\infty} -\frac{\partial}{\partial x} \exp\left(\frac{-x^2}{2\sigma^2}\right) dx \\
&= \frac{\sigma\sqrt{2}}{t\sqrt{\pi}} \exp\left(\frac{-t^2}{2\sigma^2}\right).
\end{aligned}$$

□

Sub Gaussian random variables are constrained to concentrate in a similar way which is sufficient for many purposes.

2.1.2 Equivalent definitions

The following provides equivalent definitions for sub Gaussianity with variance proxy $\sigma^2 > 0$ (up to multiplicative constants).

Theorem 2.1.1. *Let X be a centered random variable on $\mathbb{R}$, each statement bellow implies the next (we take $\sigma^2 > 0$ in the first definition as a variance proxy).*

- *Laplace transform: for any $s \in \mathbb{R}$, $\mathbb{E}\left[\exp(sX)\right] \leq \exp\left(\frac{\sigma^2 s^2}{2}\right)$.*
- *Concentration: for any $t > 0$, $\max\{\mathbb{P}\left(X \geq t\right), \mathbb{P}\left(X \leq -t\right)\} \leq \exp\left(\frac{-t^2}{2\sigma^2}\right)$.*
- *Moment condition: for any $q \in \mathbb{N}^*$, $\mathbb{E}\left[X^{2q}\right] \leq q!(4\sigma^2)^q$.*
- *Orlicz condition: $\mathbb{E}\left[\exp\left(\frac{X^2}{8\sigma^2}\right)\right] \leq 2$.*
- *Laplace transform: for any $t \in \mathbb{R}$, $\mathbb{E}\left[\exp(tX)\right] \leq \exp\left(\frac{24\sigma^2 t^2}{2}\right)$.*

Proof. The first implication is through Chernov's bound which is a consequence of Markov's inequality, for any $s > 0$, $t > 0$:

$$\begin{aligned}
\mathbb{P}(X > t) &= \mathbb{P}\left(\exp\left(sX\right) > \exp\left(st\right)\right) \\
&\leq \frac{\mathbb{E}\left[\exp\left(sX\right)\right]}{\exp\left(st\right)} \\
&\leq \exp\left(\frac{\sigma^2 s^2}{2} - st\right),
\end{aligned}$$

where the last inequalitie uses the Laplace transform condition. The result follows from the fact that $\min_{s>0} \frac{\sigma^2 s^2}{2} - st = \frac{-t^2}{2\sigma^2}$ attained for $s = t/\sigma^2$. For the second implication, we have, for any $q \in \mathbb{N}$,

$$\begin{aligned}
\mathbb{E}\left[X^{2q}\right] &= \int_0^{+\infty} \mathbb{P}\left(Z^{2q} > u\right) du \\
&= \int_0^{+\infty} \mathbb{P}\left(|Z| > u^{1/2q}\right) du \\
&\leq 2\int_0^{+\infty} \exp\left(\frac{-u^{1/q}}{2\sigma^2}\right) du \\
&= \left(2\sigma^2\right)^q 2q \int_0^{+\infty} \exp\left(-v\right) v^{q-1} dv && v = \frac{u^{1/q}}{2\sigma^2} \\
&= \left(2\sigma^2\right)^q 2qq! \\
&\leq \left(4\sigma^2\right)^q q! && 2q \leq 2^q
\end{aligned}$$

The next implication follows from the monotone convergence theorem. We obtain

$$\mathbb{E}\left[\exp\left(\frac{X^2}{8\sigma^2}\right)\right] = \mathbb{E}\left[\sum_{k=0}^{\infty} \frac{X^{2k}}{\left(4\sigma^2\right)^k k!} \frac{1}{2^k}\right] \leq \sum_{k=0}^{\infty} \frac{1}{2^k} = 2$$

Getting back to the first item is done as follows, for any $s \in \mathbb{R}$, using the fact that X is centered, for any $t \in \mathbb{R}$,

$$\begin{aligned}
\mathbb{E}\left[\exp(tX)\right] &= \mathbb{E}\left[\sum_{k=0}^{+\infty} \frac{(tX)^k}{k!}\right] \\
&= 1 + \mathbb{E}\left[\sum_{k=2}^{+\infty} \frac{(tX)^k}{k!}\right] && \mathbb{E}[X] = 0 \\
&\leq 1 + \frac{t^2}{2}\mathbb{E}\left[X^2 \exp(|tX|)\right] && \frac{(tX)^k}{k!} \leq \frac{t^2X^2}{2}\frac{|tX|^{k-2}}{(k-2)!},\ k \geq 2 \\
&\leq 1 + \frac{t^2}{2}\exp\left(4\sigma^2t^2\right)\mathbb{E}\left[X^2 \exp\left(\frac{X^2}{16\sigma^2}\right)\right] && \inf_a \left\{\frac{t^2}{2a} + \frac{aX^2}{2}\right\} = t|X|,\ a = \frac{1}{8\sigma^2} \\
&\leq 1 + 4\sigma^2t^2\exp\left(4\sigma^2t^2\right)\mathbb{E}\left[\exp\left(\frac{X^2}{8\sigma^2}\right)\right] && z \leq \exp\left(\frac{z}{2}\right) \\
&\leq \left(1 + 8\sigma^2t^2\right)\exp\left(4\sigma^2t^2\right) \\
&\leq \exp\left(\frac{24\sigma^2t^2}{2}\right) && (1+2z) \leq e^{2z}
\end{aligned}$$

□

2.1.3 Examples

Sub Gaussian random variables exist, for example the Gaussian random variable is subgaussian. Hoeffding's Lemma (1963) asserts that bounded random variables are also sub Gaussian.

Lemma 2.1.1. *Let X be a real centered random variable such that $X \in [a, b]$ almost surely. Then $\mathbb{E}[\exp(sX)] \leq \exp\left(s^2\frac{(b-a)^2}{8}\right)$ for any $s \in \mathbb{R}$, or X is sub Gaussian with variance proxy $\frac{(b-a)^2}{4}$.*

Proof. Consider the cumulent generating function $\psi\colon s \mapsto \log\left(\mathbb{E}\left[\exp(sX)\right]\right)$, we have

$$\psi'(s) = \frac{\mathbb{E}\left[X\exp(sX)\right]}{\mathbb{E}[\exp(sX)]} \qquad \psi''(s) = \frac{\mathbb{E}\left[X^2\exp(sX)\right]}{\mathbb{E}[\exp(sX)]} - \left(\frac{\mathbb{E}\left[X\exp(sX)\right]}{\mathbb{E}[\exp(sX)]}\right)^2$$

and ψ'' is the variance under the law of X reweighted by $\frac{\exp(sX)}{\mathbb{E}[\exp(sX)]}$. For any random variable Z in $[a, b]$, we have $\mathrm{var}\left[Z\right] = \mathrm{var}\left[Z - \frac{a+b}{2}\right] \leq \frac{(b-a)^2}{4}$. We can integrate two times using $\psi(0) = \log(1) = 0$ and $\psi'(0) = \mathbb{E}\left[X\right] = 0$. □

2.1.4 Sub Gaussian vectors

The definition extends similarly as for the Gaussian case.

Definition 2.1.1. *A random vector $X \in \mathbb{R}^d$ is said to be sub Gaussian with variance proxy σ^2 if it is centered and for any $u \in \mathbb{R}^d$ such that $\|u\| = 1$, the real random variable u^TX is subgaussian with variance proxy σ^2. We write $X \sim \mathrm{subG}(\sigma^2)$.*

There exists such random vectors, for example

Theorem 2.1.2. *Let $X_1, \ldots, X_p$ be independant $\mathrm{subG}(\sigma^2)$ real random variables then the random vector $X \in \mathbb{R}^p$ which i-th coordinates is X_i, is $\mathrm{subG}(\sigma^2)$.*

Proof. For any $u \in \mathbb{R}^p$ such that $\|u\| = 1$, we have for any $s \in \mathbb{R}$,

$$\mathbb{E}\left[\exp\left(su^TX\right)\right] = \prod_{i=1}^{p}\mathbb{E}\left[\exp\left(su_iX_i\right)\right] \leq \prod_{i=1}^{p}\exp\left(\frac{\sigma^2s^2u_i^2}{2}\right) = \exp\left(\frac{\sigma^2s^2\|u\|^2}{2}\right) = \exp\left(\frac{\sigma^2s^2}{2}\right)$$

□

This allows to obtain various concentration results for sub Gaussian random variables.

2.2 Maximal inequalities

We first provide tail bounds for maximum of a finite number of subgaussian random variables and then over polytopes and Euclidean ball.

Theorem 2.2.1. *Let $X_1, \dots, X_N$ be N real random variables with $X_i \sim \mathrm{subG}(\sigma^2)$, $i = 1, \dots, N$, not necessarily independant. Then*

$$\mathbb{E}\left[\max_{i=1,\dots,N} X_i\right] \le \sigma\sqrt{2\log(N)} \qquad \textit{and} \qquad \mathbb{E}\left[\max_{i=1,\dots,N} |X_i|\right] \le \sigma\sqrt{2\log(2N)}$$

and for any $t > 0$

$$\mathbb{P}\left[\max_{i=1,\dots,N} X_i > t\right] \le N\exp\left(\frac{-t^2}{2\sigma^2}\right) \qquad \textit{and} \qquad \mathbb{P}\left[\max_{i=1,\dots,N} |X_i| > t\right] \le 2N\exp\left(\frac{-t^2}{2\sigma^2}\right)$$

Proof. For any $s > 0$

$$\begin{aligned}
\mathbb{E}\left[\max_{i=1,\dots,N} X_i\right] &= \frac{1}{s}\mathbb{E}\left[\log\left(\exp\left(s\max_{i=1,\dots,N} X_i\right)\right)\right] \\
&\le \frac{1}{s}\log\left(\mathbb{E}\left[\exp\left(s\max_{i=1,\dots,N} X_i\right)\right]\right) && \text{(Jensen)} \\
&= \frac{1}{s}\log\left(\mathbb{E}\left[\max_{i=1,\dots,N} \exp(sX_i)\right]\right) \\
&\le \frac{1}{s}\log\left(\mathbb{E}\left[\sum_{i=1}^{N} \exp(sX_i)\right]\right) \\
&\le \frac{1}{s}\log\left(\sum_{i=1}^{N} \exp\left(\frac{s^2\sigma^2}{2}\right)\right) \\
&= \frac{\log(N)}{s} + \frac{s^2\sigma^2}{2}.
\end{aligned}$$

The result follows by taking $s = \sqrt{2\log(N)/\sigma^2}$. The result on the deviation probability is a simple union bound and the results on the absolute value follows by applying the two previous results to the $2N$ random variables $X_1, \dots, X_N, -X_1, \dots, -X_N$. $\square$

Remark 2.2.1. *For any $\delta > 0$, by taking $t = \sigma\sqrt{2\log(2N/\delta)}$, it holds with probability at least $1 - \delta$,*

$$\max_{i=1\dots N} |X_i| \le \sigma\sqrt{2\log(2N/\delta)}.$$

We will conclude this chapter by providing a bound for the maximum over an L_2 ball: if $X \in \mathbb{R}^p$ is $\mathrm{subG}(\sigma^2)$, can we control $\max_{\|c\|\le 1} c^T X$? We begin with a Lemma.

Lemma 2.2.1. *For any $\epsilon \in (0,1)$, it is possible to cover the Euclidean unit ball in $\mathbb{R}^p$ by at most $(3/\epsilon)^p$ Euclidean balls of radius ϵ.*

Proof. Build a covering iteratively, start with the unit ball of radius ϵ centered at 0, $\mathcal{S} = \{0\}$ and while there exists x, $\|x\| \le 1$ and $\mathrm{dist}(x, \mathcal{S}) > \epsilon$, add such an x to $\mathcal{S}$. After $\mathcal{N}$ iterations, call $x_1, \dots, x_{\mathcal{N}}$ the elements of S.

We clearly have that the balls of radius $\epsilon/2$ centered at the points in $\mathcal{S}$ are disjoint and contained in the euclidean ball of radius $1 + \epsilon/2$. Computing volumes, we obtain

$$\mathcal{N}\left(\frac{\epsilon}{2}\right)^p \le \left(1 + \frac{\epsilon}{2}\right)^p.$$

Hence the process must stop after at most $\left(\frac{2}{\epsilon} + 1\right)^p \le \left(\frac{3}{\epsilon}\right)^p$ iterations at which point we obtain a cover. $\square$

This allows to prove the following result

Theorem 2.2.2. *Let $X \sim \mathrm{subG}(\sigma^2)$ be a p dimensional random vector. Then*

$$\mathbb{E}\left[\max_{\|c\|\le 1} c^T X\right] = \mathbb{E}\left[\max_{\|c\|\le 1} |c^T X|\right] \le 4\sigma\sqrt{d}$$

and for any $t > 0$

$$\mathbb{P}\left[\max_{\|c\|\le 1} |c^T X| > t\right] = \mathbb{P}\left[\max_{\|c\|\le 1} c^T X > t\right] \le 6^d \exp\left(\frac{-t^2}{8\sigma^2}\right).$$

Proof. Consider a covering of the unit Euclidean ball with at most 6^d balls of radius $1/2$, denote by $x_1, \dots, x_{6^d}$ the centers of these balls. For any c such that $\|c\| \le 1$, there exists i such that $\|c - x_i\| \le \frac{1}{2}$. Hence we have

$$\max_{\|c\|\le 1} c^T X \le \max_{i=1,\dots,6^d} x_i^T X + \max_{\|c\|\le 1/2} c^T X = \max_{i=1,\dots,6^d} x_i^T X + \frac{1}{2}\max_{\|c\|\le 1} c^T X$$

and hence $\max_{\|c\|\le 1} c^T X \le \max_{i=1,\dots,6^d} 2x_i^T X$ and the result follows from Theorem 2.2.1 because $2x_i^T X \sim \mathrm{subG}(4\sigma^2)$ and $\log(6) \le 2$. □

Remark 2.2.2. *For any $\delta > 0$, taking $t = \sqrt{8\log(6)}\sigma\sqrt{d} + 2\sigma\sqrt{2\log(1/\delta)}$, we obtain that with probability $1-\delta$, it holds that*

$$\max_{\|c\|\le 1} c^T X = \max_{\|c\|\le 1} |c^T X| \le 4\sigma\sqrt{d} + 2\sigma\sqrt{2\log(1/\delta)} = 4\sigma\sqrt{d}\left(1 + \sqrt{\frac{\log(1/\delta)}{2d}}\right).$$

Theorem 2.2.3. *Let P be a polytope, the convex hull of N points, $v^{(1)}, \dots, v^{(N)}$ in $\mathbb{R}^d$. Let $X \in \mathbb{R}^d$ be a random variable such that for all $i = 1, \dots, n$, $[v^{(i)}]^T X \sim \mathrm{subG}(\sigma^2)$, then the conclusion of Theorem 2.2.1 holds*

$$\mathbb{E}\left[\max_{\theta\in P} \theta^T X\right] \le \sigma\sqrt{2\log(N)} \qquad \textit{and} \qquad \mathbb{E}\left[\max_{\theta\in P} |\theta^T X|\right] \le \sigma\sqrt{2\log(2N)}$$

and for any $t > 0$

$$\mathbb{P}\left[\max_{\theta\in P} \theta^T X > t\right] \le N\exp\left(\frac{-t^2}{2\sigma^2}\right) \qquad \textit{and} \qquad \mathbb{P}\left[\max_{\theta\in P} |\theta^T X| > t\right] \le 2N\exp\left(\frac{-t^2}{2\sigma^2}\right)$$

Exercises

Exercise 2.2.1. *Under the setting of Theorem 2.1.2, show that for any $t > 0$, we have*

$$\mathbb{P}\left[\frac{1}{p}\sum_{i=1}^{p} X_i \geq t\right] \leq \exp\left(\frac{-t^2 p}{2\sigma^2}\right).$$

Exercise 2.2.2. *Let Z be a real random variable with probability measure P_z on $\mathbb{R}$ such that $Z \geq 0$ almost surely. Show that*

$$\mathbb{E}[Z] = \int_0^{+\infty} \mathbb{P}(Z > u)du.$$

(Hint: use Funini's theorem. Beware: we did not assume that $\mathbb{E}[Z]$ is finite).

Exercise 2.2.3. *For $\mathbb{X} \in \mathbb{R}^{n \times d}$ and $Y \in \mathbb{R}^n$, the least squares estimator is written as*

$$\hat{\theta}^{LS} \in \arg\min_{\theta \in \mathbb{R}^d} \|\mathbb{X}\theta - Y\|_2^2. \tag{2.1}$$

We have $\mathbb{X}^T\mathbb{X}\hat{\theta}^{LS} = \mathbb{X}^T Y$ and one solution is given by $\hat{\theta}^{LS} = (\mathbb{X}^T\mathbb{X})^{\dagger}\mathbb{X}^T Y$, where $\dagger$ denotes the Moore-Penrose pseusdo inverse. (Hint: First assume that $\mathbb{X}^T\mathbb{X}$ is invertible, the pseudo inverse is then the usual matrix inverse. If you are familiar with convex analysis, the result can be deduced from convexity of the ojective, solving the first order conditions)

Recall that if D is diagonal, then its pseudo inverse is obtained by inverting the non zero diagonal elements (leaving the others unchainged). Pseudo inverse of real symmetric matrices are defined in the same way after diagonalization.

Exercise 2.2.4. *Let X be $\mathcal{N}(0, \sigma^2)$, prove that for any $t > 0$, $\mathbb{P}(|X| \geq t) \leq \frac{\sigma\sqrt{2}}{t\sqrt{\pi}} \exp\left(\frac{-t^2}{2\sigma^2}\right)$. This is called Mill's inequality.*

Exercise 2.2.5. *Let $v^{(1)}, \ldots, v^{(N)} \in \mathbb{R}^d$ and set*

$$P = \operatorname{conv}(v^{(1)}, \ldots, v^{(N)}) = \left\{\sum_{i=1}^{N} \lambda_i v^{(i)}, \quad \lambda_i \geq 0,\, i = 1, \ldots, N,\, \sum_{i=1}^{N} \lambda_i = 1\right\}$$

Show that for any $c \in \mathbb{R}^d$, the problem $\sup_{\theta \in P} c^T\theta$ is attained at $v^{(i)}$ for some $i \in \{1, \ldots, N\}$. Prove Theorem 2.2.3.

Exercise 2.2.6. *Let $X \sim \mathrm{subG}(\sigma^2)$ be a d-dimensional random vector, show that, for any $\delta > 0$, with probability $1 - \delta$,*

$$\sup_{\|\theta\|_1 \leq 1} |\theta^T X| \leq \sigma\sqrt{2\log(2d/\delta)}.$$

Exercise 2.2.7. *Let $A \in \mathbb{R}^{n \times m}$ be a random matrix which entries are iid subgaussian with variance proxy σ^2. The operator norm of A is given by $\|A\|_{op} = \sup_{x \in \mathbb{R}^m} \|Ax\|_2/\|x\|_2$. Show that $\mathbb{E}[\|A\|_{op}] \leq c\sigma\sqrt{m+n}$ for a constant c to be determined.*

Exercise 2.2.8. *Prove Jensen's inequality, if $D \subset \mathbb{R}$ is an interval and $\phi\colon D \mapsto \mathbb{R}$ is concave continuous, if X is a real random variable such that $X \in D$ with probability 1, then $\mathbb{E}[\phi(X)] \leq \phi(\mathbb{E}[X])$.*