High Dimensional Statistics

Yanbo Tang

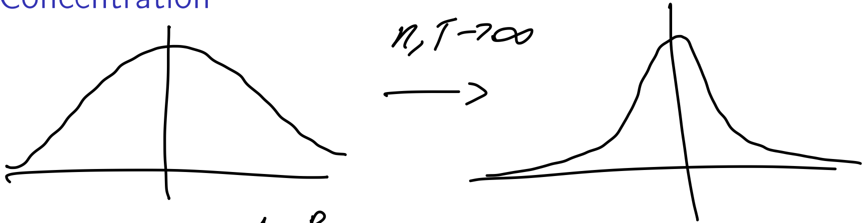
Imperial College London

Febuary 2025

Content of this week

- Concentration and its uses
- Sub-Gaussian random variables
- Sub-Exponential random variables

Concentration



$$\hat{\theta}_n \xrightarrow{P} \theta_0$$

Bayesian : Posterior Variance $\to 0$

Concentration cont.

Let $X_1, \ldots, X_n$ be a sequence of IID random variables. Take the familiar example of the weak law of large numbers,

$$\sum_{i=1}^{n} \frac{X_i}{n} \xrightarrow{p} \mu,$$

if $E|X_1| < \infty$, and where $\mu = E[X_1]$.

Concentration cont.

Let $X_1, \ldots, X_n$ be a sequence of IID random variables. Take the familiar example of the weak law of large numbers,

$$\sum_{i=1}^{n} \frac{X_i}{n} \xrightarrow{p} \mu,$$

if $E|X_1| < \infty$, and where $\mu = E[X_1]$.
Similarly we can think of the CLT, where for an IID sequence of random variables the second moment exists

$$\sqrt{n}\left(\sum_{i=1}^{n} \frac{X_i}{n} - \mu\right) \xrightarrow{p} Z,$$

where $Z \sim N(0, 1)$.

Example

But how fast is this happening? It turns out that convergence in probability/distribution alone can't answer that questions.

Example

Let $U \sim \text{uniform}(0,1)$, and $M_n \downarrow 0$ with $M_0 < 1$. Consider the sequence of random variable random variable $X_n = \mathbb{I}[u \in (0, M_n)]$, then $X_n \xrightarrow{p} 0$ and for all $0 < \epsilon < 1$

$$\mathbb{P}[|X_n| > \epsilon] \leq M_n.$$

let $M_n = (\log \log(n))^{-1}$ $\quad P(|X_n| > \frac{1}{2}) < \frac{1}{10} \leftarrow$ want

$\Rightarrow 10 \leq \log\log(n)$

$n \geq \exp(\exp(10)) \approx 2.68 \times 10^{43}$

High-dimensions

In what follows we generally thing of high-dimensional as being when the number of parameters is increasing with the number of parameters. In this scenario it is no longer clear if the CLT and WLLN hold:

$p \uparrow \infty$ as $\underset{\text{\# samples}}{n} \uparrow \infty$

High-dimensions

In what follows we generally thing of high-dimensional as being when the number of parameters is increasing with the number of parameters. In this scenario it is no longer clear if the CLT and WLLN hold:

Example

Consider a p-dimensional multivariate normal distribution $X_{n,p} \sim N(0, I_p)$, if the dimension p is fixed by the weak law of large numbers:

$$\frac{\sum_{i=1}^{n} X_{n,p}}{n} \xrightarrow{p} 0.$$

However, if p increase with n such that $p/n \to c > 0$ then

$$\left\| \frac{\sum_{i=1}^{n} X_{n,p}}{n} \right\|_2^2 \sim \frac{\chi_p^2}{n},$$

whose variance > 0, therefore this never converges in probability to 0.

Basic concentration inequality

Theorem

For a positive random variable X with $E[X] < \infty$:

$$\mathbb{P}(X \geq t) \leq \frac{\mathbb{E}[X]}{t}.$$

But what if we get more creative...

$$P(|X-\mu| \geq t) = P(|X-\mu|^k \geq t^k) \leq \frac{E[|X-\mu|^k]}{t^k}$$

well minimizing $\min_{k=1,2,3,\dots} \frac{E[|X-\mu|^k]}{t^k}$

$f(x): \mathbb{R} \to \mathbb{R}^+$ that is increasing

$$P(X-\mu \geq t) = P(f(X-\mu) \geq f(t)) \leq \frac{E[f(X-\mu)]}{f(t)}$$

$f(x) = \exp(\lambda x)$ $\lambda \geq 0$

$$P(X-\mu \geq t) \leq \inf_{\lambda \in [0,b]} E[\exp(\lambda(X-\mu) - \lambda t)]$$

$$= \exp\left\{ \inf_{\lambda \in [0,b]} \log\left(E(\exp(\lambda(X-\mu))\right) - \lambda t) \right\}$$

Chernoff approach for Gaussians

if $X \sim N(\mu, \sigma^2)$ $\quad X-\mu \sim N(0, \sigma^2)$

$$E[\exp(\lambda(X-\mu))] = \exp\left(\frac{\lambda^2\sigma^2}{2}\right) \quad \forall \lambda \in \mathbb{R}$$

$$\inf_{\lambda \in [0,\infty)} \frac{\lambda^2\sigma^2}{2} - \lambda t$$

minimized at $\lambda = t/\sigma^2$

$$P(X-\mu \geq t) \leq \exp\left(\frac{-t^2}{2\sigma^2}\right)$$

Sub-Gaussian

Definition

A random variable X is called sub-Gaussian if there exists a $\sigma > 0$:

$$\mathbb{E}[\exp(\lambda(X - \mu))] \leq \exp\left(\frac{\lambda^2\sigma^2}{2}\right)$$

for all $\lambda \in \mathbb{R}$.

Sub-Gaussian concentration

Proposition

A sub Gaussian random variable X with proxy variance σ satisfies:

$$1 \quad \mathbb{P}[X-\mu > t] \leq \exp\left(\frac{-t^2}{2\sigma^2}\right),$$

$$2 \quad \mathbb{P}[X-\mu < -t] \leq \exp\left(\frac{-t^2}{2\sigma^2}\right),$$

$$3 \quad \mathbb{P}[|X-\mu| > t] \leq 2\exp\left(\frac{-t^2}{2\sigma^2}\right).$$

All bounded variables are sub-Gaussian

Lemma

Hoeffding Lemma: *Let X be any random variable such that $a < X < b$ almost surely. Then for all $\lambda \in \mathbb{R}$:*

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

$$\sigma = \frac{(b-a)}{2}$$

Preservation of Sub-Gaussianity

Proposition

Suppose that X_1 and X_2 are 0 mean sub-Gaussian random variables with proxy variances of σ_1^2 and σ_2^2

- *If they are independent, then $X_1 + X_2$ is sub-Gaussian with proxy variance $\sigma_1^2 + \sigma_2^2$*
- *In general, $X_1 + X_2$ sub-Gaussian with proxy variance $(\sigma_1 + \sigma_2)^2$*
- *For $c \in \mathbb{R}$, cX_1 is subGaussian with proxy variance $c^2\sigma_1^2$.*

Hoeffding's bound

Theorem

Hoeffding bound for averages: *Let X_i for $i = 1, \dots, n$ be a sequence of IID random variables with proxy variances σ^2, then:*

$$\mathbb{P}\left(\left|\sum_{i=1}^{n} X_i/n - \mu\right| \geq t\right) \leq 2\exp\left(\frac{-nt^2}{2\sigma^2}\right)$$

Proof: $\frac{X_i}{n} \sim \text{subG}\left(0, \frac{\sigma^2}{n^2}\right)$

$\sum_{i=1}^{n} \frac{X_i}{n} \sim \text{subG}\left(0, \frac{\sigma^2}{n}\right)$

Application to Monte Carlo

$X_i \overset{IID}{\sim} \text{Uniform}(F')$

$$\frac{Vol(F)}{Vol(F')} \approx \sum_{i=1}^{n} \frac{I(X_i \in F)}{n}$$

$$Vol(F) \approx \sum_{i=1}^{n} \frac{Vol(F')\, I(X_i \in F)}{n}$$

$Vol(F')\, I(X_i \in F)$ is bounded $[0, Vol(F')]$

proxy $\sigma^2 = Vol(F')^2/4$

$$P\left(\left| Vol(F) - \sum_{i=1}^{n} \frac{Vol(F')\, I(X_i \in F)}{n} \right| > \epsilon \right) \le 2\exp\left(\frac{-\epsilon^2 \cdot 2n}{Vol(F')} \right)$$

$$P\left(\left| \frac{Vol(F) - \sum_{i=1}^{n} \frac{Vol(F')\, I(X_i \in F)}{n}}{Vol(F)} \right| > \epsilon \right)$$

Application to Monte Carlo cont.

Maxima of sub-Gaussians

Proposition

Let $X_1, \ldots, X_n$ be a sequence of sub-Gaussian random variables with common proxy variance σ^2 then

$$① \quad \mathbb{E}[\max_{1 \le n} X_i] \le \sigma\sqrt{2\log(n)},$$

$1 \le i \le n$

$$② \quad \mathbb{P}(\max_{1 \le n} X_i > t) \le N \exp\left(\frac{-t^2}{\sigma^2}\right).$$

$1 \le i \le n$

Note that independence is not needed.

Proof: By Jensen's inequality

$$\exp(\lambda \mathbb{E}[\max_{i=1,2,\ldots,n} Z_i]) \le \mathbb{E}[\exp(\lambda \max_{i=1,2,\ldots,n} Z_i)]$$

$$= \mathbb{E}[\max_{i=1,2,\ldots,N} \exp(\lambda Z_i)]$$

$$\max_{i=1,2,\ldots,n} a_i \le \sum_{i=1}^{n} a_i$$

if $a_i \ge 0$

Proof

$$\mathbb{E}\left[\max_{i=1,2,\dots,N} \exp(\lambda z_i)\right] \le \sum_{i=1}^{N} \mathbb{E}\left[\exp(\lambda z_i)\right] \le N \exp\left(\frac{\lambda^2\sigma^2}{2}\right)$$

$$\mathbb{E}\left[\max_{i=1,2,\dots N} z_i\right] \le \frac{\log(N)}{\lambda} + \frac{\lambda\sigma}{2}$$

minimized by $\lambda = \sqrt{2\log(N)/\sigma^2}$

$$P\left(\max_{1\le i\le N} z_i > t\right) = P\left(\bigcup_{1\le i\le N} \{z_i > t\}\right) \le \sum_{i=1}^{N} P(z_i > t)$$

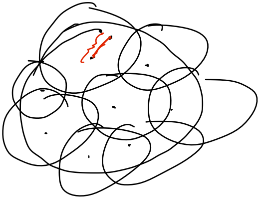
$$\le N \exp\left(\frac{-t^2}{2\sigma^2}\right)$$

Maximum over infinite sets

What if we wanted a maximum over an infinite set? For example, consider the unit ℓ_2 ball in $\mathbb{R}^d$ and we are interested in controlling for:

$$E[\sup_{\theta \in \mathcal{B}_2} \theta^\top X],$$

where X follows some sub-Gaussian distribution.



ϵ-cover

Definition

Fix $K \subset \mathbb{R}^d$ and $\varepsilon > 0$. A set $\mathcal{N}$ is called an ε-net of K with respect to a distance $d(\cdot, \cdot)$ on $\mathbb{R}^d$, if $\mathcal{N} \subset K$ and for any $z \in K$, there exists $x \in \mathcal{N}$ such that $d(x, z) \leq \varepsilon$.

Unit ball

Lemma

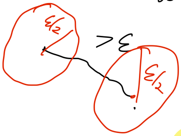
For any $\varepsilon \in (0,1)$, the unit Euclidean ball $\mathcal{B}_2$ has an ε-net $\mathcal{N}$ with respect to the Euclidean distance of cardinality $|\mathcal{N}| \le (3/\varepsilon)^d$.

take $x_1 = 0$ $\forall i \ge 2$ take x_i to be any $x \in B_2$ such that

$|x - x_j| > \varepsilon$ $\forall j < i$. If can't find such x_i then we are done.

How big is this collection? $\forall x, y \in N$ $|x - y| > \varepsilon$

$\therefore$ Balls centred at x_i of radius $\varepsilon/2$ are disjoint



$$\bigcup_{z \in N} \left\{ z + \tfrac{\varepsilon}{2} B_2 \right\} \subset \left(1 + \tfrac{\varepsilon}{2}\right) B_2$$

$$\mathrm{Vol}\left(\left(1+\tfrac{\varepsilon}{2}\right)B_2\right) \ge \mathrm{Vol}\left(\bigcup_{z \in N}\left\{z + \tfrac{\varepsilon}{2} B_2\right\}\right) = \sum_{z \in N} \mathrm{Vol}\left(z + \tfrac{\varepsilon}{2} B_2\right)$$

$$\left(1 + \tfrac{\varepsilon}{2}\right)^d \ge |N| \left(\tfrac{\varepsilon}{2}\right)^d$$

Proof

$|N| \leq (1+\frac{2}{\varepsilon})^d \leq (\frac{3}{\varepsilon})^d$

Supremum over L^2 ball

Theorem

Let $X \in \mathbb{R}^d$ be a sub-Gaussian random vector with variance proxy σ^2. Then

$$\mathbb{E}\big[\sup_{\theta \in B_2} \theta^T X\big] = \mathbb{E}\big[\sup_{\theta \in B_2} |\theta^T X|\big] \le 4\sigma\sqrt{d}.$$

Moreover, for any $\delta > 0$, with probability $1-\delta$, it holds

$$\sup_{\theta \in B_2} \theta^T X = \sup_{\theta \in B_2} |\theta^T X| \le 4\sigma\sqrt{d} + 2\sigma\sqrt{2\log(1/\delta)}.$$

Proof: take a $\frac{1}{2}$-Net then $|N| \le 6^d$

$\forall \theta \in B_2 \ \exists z \in N \ \exists x: \|x\|_2 < 1/2$ such that $\theta = z + x$

$$\max_{\theta \in B_2} \theta^T X \le \max_{z \in N} z^T X + \max_{x \in \frac{1}{2}B_2} x^T X = \max_{z \in N} z^T X + \frac{1}{2}\max_{x \in B_2} x^T X$$

$$\frac{1}{2}\max_{\theta \in B_2} \theta^T X = \max_{z \in N} z^T X$$

But $\max_{x \in \frac{1}{2}B_2} x^T X = \frac{1}{2} \max_{x \in B_2} x^T X$

$$E[\max_{\theta \in B_2} \theta^T X] \leq 2 E[\max_{z \in N} z^T X] \leq 2\sigma\sqrt{2\log(N) d}$$
$$\leq 4\sigma\sqrt{d}$$

Used $\to$ $N = 6^d$ $\quad \log(N) = 6\log(d)$

$$P(\max_{\theta \in B_2} \theta^T X > t) \leq P(2\max_{z \in N} z^T X > t)$$
$$\leq |N| \exp\left(\frac{-t^2}{8\sigma^2}\right)$$
$$\leq 6^d \exp\left(\frac{-t^2}{8\sigma^2}\right)$$

Solve for $6^d \exp(-t^2/8\sigma^2) = \delta$

$$\Rightarrow t \leq 4\sigma\sqrt{d} + 2\sigma\sqrt{2\log(\tfrac{1}{\delta})}$$

Sub-Exponential definition

Definition

A random variable with mean X with $\mu = E[X]$ is sub-exponential if there are non-negative parameters (ν, α) such that

$$E\left[\exp(\lambda(X-\mu))\right] \leq \exp\left(\frac{\nu^2\lambda^2}{2}\right) \text{ for all } |\lambda| < \frac{1}{\alpha}.$$

Sub-exponential concentration

Proposition

Suppose that X is a sub-exponential distribution with parameters (ν, α) then:

$$\mathbb{P}[X - \mu \geq t] \leq \begin{cases} e^{-\frac{t^2}{2\nu^2}} & \textit{if } 0 \leq t \leq \frac{\nu^2}{\alpha}, \\ e^{-\frac{t}{2\alpha}} & \textit{for } t > \frac{\nu^2}{\alpha}. \end{cases}$$

e.g. exponential distribution with $\theta = 1$ has mgf $\frac{1}{1-\lambda} = E[\exp(\lambda x)]$

$\lambda < 1$

Chernoff's approach

$$P(X \geq t) \leq e^{-\lambda t} E[e^{\lambda X}] \leq \exp(\underbrace{-\lambda t + \frac{\lambda^2 \nu^2}{2}}_{g(\lambda, t)})$$

for $\lambda \in [0, \frac{1}{\alpha}]$

Need $\forall t \geq 0$ $\quad g^*(t) = \inf_{\lambda \in [0, 1/\alpha]} g(\lambda, t)$ $\quad$ unconstrained minimum occurs at $\lambda^* = t/\nu^2$

① if $t < \nu^2/\alpha$ then we have that global is valid/feasible.

② if $t \geq \nu^2/\alpha$ minimum occurs at $1/\alpha$

$g(\lambda^*, t) = -t/\alpha + \frac{1}{2\alpha}\frac{\nu^2}{\alpha} \leq -t/2\alpha$.

Bernstein condition

Definition

Bernstein condition. Given a random variable X with mean μ and variance σ^2, we say that *Bernstein's condition* with parameter b holds if

$$\mathbb{E}[(X-\mu)^k] \leq \frac{1}{2} k! \sigma^2 b^{k-2} \quad \text{for all } k \in \mathbb{N}.$$

Bernsetin condition

Proposition

Another way to get sub-exp tails For any random variable satisfying the Bernstein condition with parameter b

$$\mathbb{E}[e^{\lambda(X-\mu)}] \leq \exp\left(\frac{\lambda\sigma^2}{2-2|\lambda|b}\right) \qquad \textit{for all } |\lambda| < \frac{1}{\nu},$$

and, moreover, the concentration inequality

$$\mathbb{P}[|X-\mu| \geq t] \leq 2\exp\left(\frac{-t^2}{2(\sigma^2+bt)}\right) \qquad \textit{for all } t \geq 0.$$

Preservation of sub-exponentials

Proposition

Preservation of sub-exponential property. *For a sequence of independent random variables X_i for $i = 1, \ldots, n$ which are sub-exponential (ν_i, α_i), the sum*

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

is sub-exponential with parameters (ν_*, α_*) where $\alpha_* = \max_{i=1,\ldots,n} \alpha_i$ and $\nu_\star = \sqrt{\sum_{i=1}^{n} \nu_i^2}$.

Chi-squared concentration

Concentration for $\chi_n^2 = \sum_{i=1}^{n} Z_i^2$ $\quad Z_i \sim N(0,1)$

$Z_i^2 \sim \text{subE}(\overset{\nu}{2}, \overset{\alpha}{4})$

$Y \sim \text{subE}(2\sqrt{n}, 4)$

$$P\left[\left|\frac{1}{n}\sum_{i=1}^{n} Z_i^2 - 1\right| \geq t\right] \leq 2e^{-nt^2/8} \quad \text{for } t \in (0,1)$$

Johnson-Lindenstrauss

$(u_1, \dots, u_N)$ $\quad u_i \in \mathbb{R}^d$

$F: \mathbb{R}^d \mapsto \mathbb{R}^m \quad m \ll d$

$$(1-\delta) \le \frac{\|F(u_i) - F(u_j)\|_2}{\|u_i - u_j\|_2} \le (1+\delta) \quad \delta > 0$$

Random Projection $\quad X \in \mathbb{R}^{m \times d}$ when $X_{ij} \overset{iid}{\sim} N(0,1)$

$$F: u \to \frac{Xu}{\sqrt{m}}$$

Note that $X_i \in \mathbb{R}^d$ $\quad X_i \sim N(0, I_{d \times d})$

then $\langle X_i, \frac{u}{\|u\|_2} \rangle \sim N(0,1)$

$$Y := \frac{\|Xu\|_2^2}{\|u\|_2^2} = \sum_{i=1}^m \langle X_i, \frac{u}{\|u\|_2} \rangle^2 \sim \chi^2_m$$

Johnson-Lindenstrauss Cont

$$P\left[\left|\frac{\|Xu\|_2^2}{m\|u\|_2^2}-1\right|>\delta\right]\leq 2\exp\left(\frac{-m\delta^2}{8}\right)$$

By definition of F $\forall u\in\mathbb{R}^d$

$$P\left[\frac{\|F(u)\|_2^2}{\|u\|_2^2}\notin[(1-\delta),(1+\delta)]\right]\leq 2\exp\left(\frac{-m\delta^2}{8}\right)$$

Let $u=v_i-v_j\in\mathbb{R}^d$

Use union bound over all i,j

$$\forall i,j \quad P\left[\frac{\|F(u_i)-F(u_j)\|_2^2}{\|u_i-u_j\|_2^2}\notin[(1-\delta),(1+\delta)]\right]$$

$$\leq 2\binom{N}{2}\exp\left(-\frac{m\delta^2}{8}\right)$$

w.p. $1-\delta$ $\quad m\geq\frac{16}{\delta^2}\log\left(\frac{N}{\varepsilon}\right)$