CLT, Hoeffding's ≠

Note Title 2015-09-08

**Prof:** Yaniv Plan
("Yuh-neev Plan")

**Class website:**

www.yanivplan.com/math-608d

**Theory:** Understanding "probabilistic objects" in high dimensions, e.g., random variables, random matrices, functions with random input, shapes as seen through "prob. lens".

**Applications:** Compressed sensing, statistical estimation, dimension reduction, etc.

# Concentration of measure

## 1. Deviation inequalities for sums of indep r.v.'s

### 1.1 CLT: asymptotic and nonasymptotic

Normal r.v.: $g \sim N(0,1)$

density $f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$

Ubiquitous in science. Why?

Thm: (CLT, Lindeberg-Levy)

Let $X_1, X_2, \dots$ be iid r.v.'s with $\mathbb{E} X_i = 0$, $Var(X_i) = 1$. Then

$$\lim_{n\to\infty} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i \longrightarrow N(0,1)$$

uniformly in distribution.

Remarks:

① Role of $\frac{1}{\sqrt{n}}$: $Var\left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i\right) = Var(g)$

$= 1.$

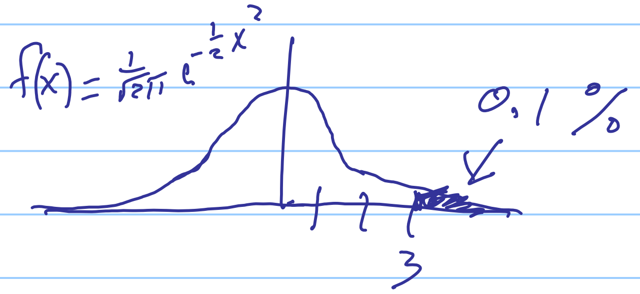
(2) "Uniformly in distribution"

$$\lim_{n\to\infty} \sup_{t\in\mathbb{R}} \left| P\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i < t\right) - P(g<t) \right| \to 0$$

↑ c.d.f. of $\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i$ ↑ normal c.d.f.

Why is $N(0,1)$ so useful?

- Fast tail decay




Roughly: $P(g>t) \approx \frac{1}{t}\frac{e^{-\frac{1}{2}t^2}}{\sqrt{2\pi}} = \frac{1}{t} f(t)$

See: [Durrett, Probability: Theory & Examples Thm 1.4]

Precisely: $\left(\frac{1}{t}-\frac{1}{t^3}\right) f(t) \le P(g>t) \le \frac{1}{t} f(t)$

- $P(g>t) \le \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2}, \quad t \ge 1$

↑ super-exponential tail decay

**Question** Does $\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i$ have the same fast tail decay?

CLT suggests "yes", but this does not follow from CLT. Depends on convergence rate of CLT.

**Thm** (CLT: Berry-Esseen)

$$\sup_{t\in\mathbb{R}} \left| P\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i < t\right) - P(g<t)\right| \le \frac{C\rho}{\sqrt{n}}$$

where $\rho := \mathbb{E}|X_1|^3$, $C$ = abs. const

**HW** Show by example that the rate $\sim \frac{\rho}{\sqrt{n}}$ cannot be improved, i.e., find a distribution for $X_i$, and an abs. const $c$, so that

$\sup_t \left| P\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i < t\right) - P(g<t)\right| \ge c\frac{\rho}{\sqrt{n}}$ for all $n \ge 1$.

**Remark:** Non-asymptotic version of CLT.

This gives a non-asymptotic tail bound:

$$P\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i > t\right) \le P(g>t) + \frac{C\rho}{\sqrt{n}} \lesssim e^{-t^2} + \frac{\rho}{\sqrt{n}}$$

($\lesssim$ hides abs. const.)

- Unfortunately, $\frac{\rho}{\sqrt{n}}$ does not depend on $t$, and can ruin exponential decay.

- Further, one cannot hope for exponential decay with only the condition $|X_1|^3 < \infty$. (why?)

**Soln:** Deviation inequalities: Bound tail of $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i$ directly, without comparing to $N(0,1)$.

## 1.2: Hoeffding's inequality for Rademacher r.v.'s

**Def:** $X \sim$ Rademacher iff $P(X=1) = P(X=-1) = 0.5$

Consider $\sum_{i=1}^{n} a_i X_i$, $X_i \overset{iid}{\sim}$ Rademacher

Normalization: Require variance $= 1$.

$$Var\left(\sum a_i X_i\right) = \sum a_i^2 Var(X_i) = \sum a_i^2.$$

Req: $\boxed{\sum_{i=1}^{n} a_i^2 = 1}$

**First attempt: Bound tail with moment bound.**

$$P\left(\left|\sum a_i X_i\right| > t\right) = P\left(\left(\sum a_i X_i\right)^2 > t^2\right)$$

$$\leq \frac{\mathbb{E}\left(\sum a_i X_i\right)^2}{t^2} = \frac{1}{t^2}$$

Chebychev
$\uparrow$

$\frac{1}{t^2}$ is not exponential!

2nd attempt:

$$P\left(\sum_i a_i X_i > t\right) = P\left(\exp\left(\lambda \sum a_i X_i\right) > e^{\lambda t}\right)$$

($\lambda$ will be optimized later)

"expon. Chebychev ≠" ↗

$$\leq e^{-\lambda t}\, \mathbb{E} \exp\left(\lambda \sum_i a_i X_i\right)$$

$$= e^{-\lambda t} \prod_{i=1}^{n} \mathbb{E}\, e^{(\lambda a_i X_i)} \quad \text{(by indep.)}$$

$$= (*)$$

$$\mathbb{E}\, e^{\lambda a_i X_i} = \frac{1}{2} e^{\lambda a_i} + \frac{1}{2} e^{-\lambda a_i} = \cosh(\lambda a_i)$$

Note: $\cosh(x) = 1 + \frac{x^2}{2!} + \frac{x^4}{4!} + \ldots$

$\exp\left(\frac{x^2}{2}\right) = 1 + \frac{x^2}{2} + \frac{x^4}{4 \cdot 2!} + \ldots$

Exercize: $\cosh(x) \leq \exp\left(\frac{x^2}{2}\right)$

$$\Rightarrow (*) \leq e^{-\lambda t} \prod e^{\frac{(\lambda a_i)^2}{2}} = \exp\left(-\lambda t + \frac{\lambda^2}{2} \sum a_i^2\right)$$

$$= \exp\left(-\lambda t + \frac{\lambda^2}{2}\right)$$

Optimize $\lambda$: $\lambda = t$

$$\Rightarrow (*) \leq \exp\left(\frac{-t^2}{2}\right)$$

We have proven:

Thm Hoeffding ≠

Let $X_i$ be indep. Rademacher.

$\sum_{i=1}^{n} a_i^2 = 1$. Then

$$P\left(\sum_{i=1}^{n} a_i X_i > t\right) \le \exp\left(-\frac{t^2}{2}\right), \quad t > 0.$$

Remarks:

① ~ matches normal tail $P(g > t) \lesssim e^{-\frac{t^2}{2}}$

② Non-asymptotic

③ Method is very flexible.
(Due to S. Bernstein.)
Generalizes to other distributions.
(Even to random matrices.)

Literature: [G. Lugosi, concentration of measure inequalities]

Next lecture: Largest class of r.v.'s s.t. result of above flavor occurs.