Basic tail and concentration bounds

Why concentration inequality?

The concentration inequality quantify how a random X deviates around its mean, i.e.

$$\mathbb{P}(|X-\mu|>t),$$

where $\mu=\mathbb{E}X$.

Chernoff Inequality

$$\mathbb{P}(X-\mu>t)\le\frac{\mathbb{E}e^{\lambda(X-\mu)}}{e^{\lambda t}}$$

$$\Rightarrow \log\mathbb{P}(X-\mu>t)\le \log\mathbb{E}e^{\lambda(X-\mu)}-\lambda t$$

$$(\Rightarrow)\le\inf_{\lambda}\left(\log\mathbb{E}e^{\lambda(X-\mu)}-\lambda t\right),$$

Example

If $X \sim N(\mu, \sigma^2)$, then

$$\mathbb{E} e^{\lambda(X-\mu)} = e^{\sigma^2\lambda^2/2}$$

$$\Rightarrow \log \mathbb{P}(X \geq \mu+t) \leq \inf_\lambda \left(\frac{\sigma^2\lambda^2}{2} - \lambda t\right)$$

$$\Rightarrow \mathbb{P}(X \geq \mu+t) \leq e^{-\frac{t^2}{2\sigma^2}}.$$

Hoeffding Inequality

If $\mathbb{P}(X=1) = \mathbb{P}(X=-1) = \frac{1}{2}$, we call X the Rademacher variable.

Theorem $\alpha \in \mathbb{R}^n$, then

$$\mathbb{P}\left(\sum_{i=1}^n \alpha_i X_i > t\right) \leq \exp\left(-\frac{t^2}{2\|\alpha\|_2^2}\right),$$

where X_i is the iid Rademacher variable.

Pf: $\mathbb{P}\left(\sum_{i=1}^{n} a_i X_i > t\right)$

$$\leq \frac{\mathbb{E} e^{\lambda \sum_{i=1}^{n} a_i X_1}}{e^{\lambda t}}$$

$$= \frac{\prod_{i=1}^{n} \mathbb{E} e^{\lambda a_i X_1}}{e^{\lambda t}}, \quad \lambda > 0$$

where $\mathbb{E} e^{\lambda a_i X_1}$

$$= \left(e^{\lambda a_i} + e^{-\lambda a_i}\right) / 2.$$

and $e^{\lambda a_j} = \sum_{j=0}^{\infty} \frac{(\lambda a_i)^j}{j!}$

$\Rightarrow$

$$\left(e^{\lambda a_i} + e^{-\lambda a_i}\right)/2$$

$$= \sum_{j=0}^{\infty} \frac{(\lambda a_i)^{2j}}{(2j)!} \leq \sum_{j=0}^{\infty} \left(\frac{(\lambda a_i)^2}{2}\right)^j / j!$$

$$= e^{\frac{(\lambda a_i)^2}{2}}.$$

$$\Rightarrow \prod_{i=1}^{n} \mathbb{E} e^{\lambda \alpha_i X_i} \le e^{\frac{\lambda^2 \|\alpha\|_2^2}{2}}$$

$$\Rightarrow \mathbb{P}\left(\sum_{i=1}^{n} \alpha_i X_i > t\right) \le \exp\left(\frac{\lambda^2 \|\alpha\|_2^2}{2} - \lambda t\right)$$

$$(\Rightarrow) \le \exp\left(-\frac{t^2}{2\|\alpha\|_2^2}\right). \quad \square$$

Remark:

① $\mathbb{P}\left(\left|\sum_{i=1}^{n} \alpha_i X_i\right| > t\right) \le 2\exp\left(-\frac{t}{2\|\alpha\|_2^2}\right)$

② If $|Y_i| < M_i$, $\mathbb{E} Y_i = 0$, then

$$\mathbb{E} e^{\lambda |Y_i|} = \mathbb{E} e^{\lambda |Y_i - \mathbb{E} Y_i'|}$$

$Y_i =_d Y_i'$

$Y_i \perp Y_i'$

$$\le \mathbb{E}_{Y_i} \mathbb{E}_{Y_i'} e^{\lambda |Y_i - Y_i'|}$$

$$= \mathbb{E}_{Y_i} \mathbb{E}_{Y_i'} \mathbb{E}_{X_i} e^{\lambda X_i |Y_i - Y_i'|},$$

$X_i \perp (Y_i, Y_i')$

$$\le \mathbb{E}_{Y_i}\mathbb{E}_{Y_i'} e^{\lambda^2 (Y_i - Y_i')^2/2}$$

$$\le e^{\lambda^2 (2M_i)^2/2}$$

$$\Rightarrow \mathbb{P}\left(\left|\sum_{i=1}^{n} Y_i\right| > t\right)$$

$$\le e^{-\frac{t^2}{8\sum_{i=1}^{n} M_i^2}}$$

③ If $X \in [a,b]$, then

$$p_{Y_\lambda}(y) = e^{\lambda y} p_X(y) / \mathbb{E}e^{\lambda X}$$

$$\Rightarrow \quad Y_\lambda \in [a,b] \text{ a.s.}$$

$$\mathrm{Var}(Y_\lambda) = \mathbb{E}(Y_\lambda - \mathbb{E}Y_\lambda)^2$$

$$\le \mathbb{E}\left(Y_\lambda - \frac{a+b}{2}\right)^2 \le \frac{(b-a)^2}{4}$$

Let $\psi(\lambda) = \log \mathbb{E}e^{\lambda X}$, then

$$\psi(\lambda) \le \lambda \mathbb{E}X + \frac{\lambda^2}{2}\frac{(b-a)^2}{4}.$$

(We have $\psi^{(2)}(\lambda) = \mathrm{Var}(Y_\lambda)$)

(Azuma-Hoefding)

Let $(D_k, \mathcal{F}_k)_{k=1}^{\infty}$ be a martingale difference sequence such that

$$D_k \in [a_k, b_k], \ a.s.$$

then

$$\mathbb{P}\left(\left|\sum_{k=1}^{n} D_k\right| > t\right) \le 2e^{-\frac{2t^2}{\sum_{i=1}^{n}(b_k-a_k)^2}}$$

Pf:

$$\mathbb{E}\, e^{\lambda\left(\sum_{k=1}^{n} D_k\right)}$$

$$= \mathbb{E}\, e^{\lambda\left(\sum_{k=1}^{n-1} D_k\right)} \mathbb{E}\, e^{\lambda D_n} \mid \mathcal{F}_{n-1}$$

And

$$\mathbb{E}\, e^{\lambda D_n} \mid \mathcal{F}_{n-1} \le \frac{\lambda^2}{2}\left(\frac{b_k-a_k}{2}\right)^2$$

$\square$

Now, let f be the bounded difference function, i.e.

$$|f(\dots, \underset{\substack{\uparrow \\ k^{th}}}{x}, \dots) - f(\dots, \underset{\substack{\uparrow \\ k^{th}}}{y}, \dots)| \le L_k$$

and $X_1, \dots, X_n$ is independent, then

$$f(X^n) - \mathbb{E} f(X^n)$$

$$= \sum_{i=1}^{n} \mathbb{E} f(X) | X^i - \mathbb{E} f(X) | X^{i-1}$$

where $X^i = (X_1, \dots, X_i)$,

Let $D_i := \mathbb{E} f(X^n) | X^i - \mathbb{E} f(X^n) | X^{i-1}$,

$$= \mathbb{E}_{X_{(i+1):n}} f(X^i, X_{(i+1):n}) - \mathbb{E}_{X_{i:n}} f(X^{i-1}, X_{i:n})$$

Let

$$D_i(x) := \mathbb{E}_{X_{(i+1):n}} f(X^{i-1}, x, X_{(i+1):n})$$

$$- \mathbb{E}_{X_{(i+1):n}} f(X^{i-1}, X_i, X_{(i+1):n})$$

then

$$\mathbb{E}\,|D_i(x) - D_i(y)|\,|\,\mathcal{F}_{i-1} \le L_i$$

$\Rightarrow$ D_i lies ~~wit~~ in an interval of length L_i a.s.

$$\Rightarrow \mathbb{P}\left(\left|\sum_{i=1}^{n} D_i\right| \ge t\right) \le 2e^{-\frac{2t^2}{\sum_{i=1}^{n} L_i^2}}$$

$$\Rightarrow \mathbb{P}\left(|f(X^n) - \mathbb{E}f(X^n)| \ge t\right) \le 2e^{-\frac{2t^2}{\sum_{i=1}^{n} L_i^2}}$$

Example 1

If $X_i \in [a,b]$, $\mathbb{E} X_i = \mu_i$

then

$$\mathbb{P}\left(\left|\sum_{i=1}^{n}(X_i-\mu_i)\right| \geq t\right) \leq 2e^{-\frac{2t^2}{n(b-a)^2}}$$

Example 2 U-statistics

Let $|g(x,y)| \leq b$ and

$$\hat{U} := U(X^{(n)}) := \frac{1}{C_n^2}\sum_{j<k} g(X_j, X_k),$$

Note that

$$|U(\cdots,x,\cdots) - U(\cdots,y,\cdots)| \leq \frac{1}{C_n^2}(n-1)2b = \frac{4b}{n}$$

$$\Rightarrow \mathbb{P}(|\hat{U} - \mathbb{E}\hat{U}| \geq t) \leq 2e^{-\frac{nt^2}{8b^2}}$$

Sub-gaussian variable

A random variable X is called the sub-gaussian variable iff $\exists\, c$ s.t.

$$\mathbb{P}(|X|>t)\le 2e^{-ct^2}.$$

Mills ratio

If $Z\sim N(0,\sigma^2)$, then

$$\phi(t)\left(\frac{1}{t}-\frac{1}{t^3}\right)\le \mathbb{P}(Z\ge t)\le \phi(t)/t,$$

where $\phi(t)=\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{t^2}{2\sigma^2}}$

Pf: $\mathbb{P}(Z\ge t)=\int_t^{+\infty}\phi(x)\,dx$

$$\le \frac{1}{2t}\int_t^{+\infty}\phi(x)\,dx^2$$

$= \sigma^2 \phi(t) / t,$

Since $\phi'(t) + \frac{t}{\sigma^2}\phi(t) = 0$

$$\mathbb{P}(Z > t) = \int_t^{+\infty} \phi(x)\,dx$$

$$= -\sigma^2 \int_t^{+\infty} \frac{\phi'(x)}{x}\,dx$$

$$= -\frac{\sigma^2 \phi(x)}{x}\Big|_t^{+\infty} - \sigma^2 \int_t^{+\infty} \frac{\phi(x)}{x^2}\,dx$$

$$\geq \sigma^2 \phi(t) \left(\frac{1}{t} - \frac{1}{t^3}\right)$$

Remark:

If $X \sim N(0, \sigma^2)$, then we have:

$$\mathbb{P}(X > t) \simeq e^{-\frac{t^2}{2\sigma^2}}$$

If X is sub-Gaussian,

then $\exists\ Z \sim N(0, \sigma^2)$, s.t.

then $\exists\ Z \sim N(0, \sigma^2)$ s.t.

$$\mathbb{P}(|X| > t) \leq \mathbb{P}(|Z| > t),$$

Theorem

(1) $\mathbb{P}(|X| > t) \leq 2e^{-\frac{t^2}{2C_1^2}}$

$\Leftrightarrow$

(2) $\mathbb{E}X^{2k} \leq \frac{(2k)!}{2^k k!} C_2^{2k}$

$\Leftrightarrow$

(3) $\mathbb{E}e^{X^2/C_3^2} \leq 2$

$\Leftrightarrow$

(4) $\mathbb{E}e^{\lambda X} \leq e^{C_4^2 \lambda^2}$, if $\mathbb{E}X = 0$, for $\forall \lambda \in \mathbb{R}$.

Pf: (1) $\Rightarrow$ (2):

$$\mathbb{E}X^{2k} \leq 2\mathbb{E}Z^{2k} = \frac{(2k)!}{2^k k!}\sigma^2$$

where $Z \sim N(0, \sigma^2)$, $\sigma^2 = C_1$.

$(2) \Rightarrow (3)$:

$$\mathbb{E}\, e^{\lambda^2 X^2} = \sum_{i=0}^{\infty} \mathbb{E} \frac{\lambda^{2k} X^{2k}}{k!}$$

$$\leq \sum_{i=0}^{\infty} \frac{\lambda^{2k}}{k!} \frac{(2k)!}{2^k k!} C_2^{2k}$$

$$\left(k! \simeq \sqrt{2\pi k} \left(\frac{k}{e}\right)^k \right)$$

$$\leq \sum_{i=0}^{\infty} (2\lambda^2 C_2^2)^k$$

$$= \frac{1}{1 - 2\lambda^2 C_2^2}$$

$(2) \Rightarrow (3)$:

$$\mathbb{E}\, e^{\lambda X^2} = \mathbb{E} \sum_{k=0}^{\infty} \frac{\lambda^{2k} X^{2k}}{k!}$$

$$\leq \sum_{k=0}^{\infty} \frac{\lambda^{2k} (2k)^k}{k!} C^{2k}$$

$$\left(k! \simeq \sqrt{2\pi k} \left(\frac{k}{e}\right)^k \right)$$

$$\leq \sum_{k=0}^{\infty} \lambda^{2k} C_1^{2k} C_2^{2k}$$

$$= \frac{1}{1-c\lambda^2} \qquad \square$$

(2) $\Rightarrow$ (4):

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

(If $\mathbb{E} X^{2k+1} = 0$, then)

$$\le \sum_{k=0}^{\infty} \frac{\lambda^{2k}}{(2k)!} \frac{(2k)!}{2^k k!} C_2^{2k}$$

$$= e^{\lambda^2 C_2^2 / 2}$$

For general X,

$$|\lambda|^{2k+1} \mathbb{E}|X|^{2k+1} \le \sqrt{\lambda^{2k} \mathbb{E} X^{2k} \lambda^{2k+2} \mathbb{E} X^{2k+2}}$$

$$\le \frac{1}{2} \left(\lambda^{2k} \mathbb{E} X^{2k} + \lambda^{2k+2} \mathbb{E} X^{2k+2} \right)$$

$$\le \frac{1}{2}\left(\lambda \ \mathbb{E}X + \lambda \ \mathbb{E}X\right)$$

$$\Rightarrow \sum_{k=0}^{\infty} \frac{\lambda^k \mathbb{E}X^k}{k!} \le 1 + \sum_{k=1}^{\infty} \frac{\lambda^{2k}}{(2k)!} \mathbb{E}X^{2k}$$

$$+ \frac{1}{2} \sum_{k=1}^{\infty} \frac{\lambda^{2k}}{(2k+1)!} \mathbb{E}X^{2k}$$

$$+ \frac{1}{2} \sum_{k=1}^{\infty} \frac{\lambda^{2k+2}}{(2k+1)!} \mathbb{E}X^{2k+2}$$

$$= \left[\sum_{k=2}^{\infty} \left(\frac{1}{2k!} + \frac{1}{2}\frac{1}{(2k+1)!} + \frac{1}{(2k-1)!}\right) \cdot \lambda^{2k} \mathbb{E}X^{2k}\right]$$

$$+ 1 + \left(\frac{1}{2} + \frac{1}{2 \cdot 3!}\right) \lambda^2 \mathbb{E}X^2$$

$$\le \sum_{k=0}^{\infty} \frac{2^k \lambda^{2k} \mathbb{E}X^{2k}}{(2k)!}$$

$$\le e^{(\sqrt{2}\lambda c_?)^2/2}$$

$\leq e$

(3) $\Rightarrow$ (1):

$$\mathbb{P}(|X| > t) \leq \frac{\mathbb{E}\, e^{X^2/C_3^2}}{e^{X^2/C_3^2}}$$

$$\leq 2e^{-t^2/C_3}$$

(4) $\Rightarrow$ (1)

$$\mathbb{P}(X > t) \leq \frac{\mathbb{E}e^{\lambda X}}{e^{\lambda t}}$$

$$\leq e^{(C_4\lambda)^2 - \lambda t}$$

$$(\Rightarrow) \quad \leq e^{-\frac{t^2}{4C_4^2}}$$

Similarly,

$$\mathbb{P}(-X > t) \leq e^{-\frac{t^2}{4C_4^2}}.$$

$\square$

Always, we define a norm for a sub-gaussian variable:

$$\|X\|_{\psi_2} = \inf\{t>0,\ \mathbb{E}e^{\frac{X^2}{t^2}} \le 2\}$$

Theorem

$$\mathbb{E}\, e^{X^2/\|X\|_{\psi_2}^2} \le 2$$

$\Leftrightarrow$

$$\mathbb{P}(|X| \ge t) \le 2e^{-\frac{t^2}{C_1\|X\|_{\psi_2}}}$$

$\Leftrightarrow$

$$\mathbb{E}|X|^{2k} \le \frac{(2k)!}{2^k k!}(C_2\|X\|_{\psi_2})^{2k}$$

$\Leftrightarrow$ $\mathbb{E}e^{\lambda X} \le e^{C_3\lambda^2\|X\|_{\psi_2}^2}$, if $\mathbb{E}X = 0$.

Remark: $\|X_1 + X_2\|_{2k}$

$\le \|X_1\|_{2k} + \|X_2\|_{2k}$

$$\le C_2 \left(\frac{\cdots}{2^k k!}\right)^{2k} (\|X_1\|_{\psi_2} + \|X_2\|_{\psi_2})$$

$$\Rightarrow \|X_1\|_{\psi_2} + \|X_2\|_{\psi_2} \ge \|X_1 + X_2\|_{\psi_2}$$

$\|\cdot\|_{\psi_2}$ is a norm.

General Hoefding Inequality

Let $X_1, \cdots, X_n$ be independent, mean zero, sub-gaussian variable, then

$$\mathbb{P}\left(\left|\sum_{i=1}^n X_i\right| \ge t\right) \le 2\exp\left(-\frac{ct^2}{\sum_{i=1}^n \|X\|_{\psi_2}^2}\right)$$

Pf: Note that

$$\mathbb{E} e^{\lambda \sum_{i=1}^n X_i} = \prod_{i=1}^n \mathbb{E} e^{\lambda X_i}$$

$$\le \mathbb{E} e^{\lambda^2 C_3 \sum_{i=1}^n \|X_i\|_{\psi_2}^2}$$

$$\Rightarrow \left\|\sum_{i=1}^n X_i\right\|_{\psi_2}^2 \le \sum_{i=1}^n \|X_i\|_{\psi_2}^2$$

$$\Rightarrow \left\| \sum_{i=1}^{n} X_i \right\|_{\psi_2} \lesssim \sum_{i=1}^{n} \| X_i \|_{\psi_2} \quad \square$$