20200130 김유준
MAS 480 HW3
Concentration Inequalities

#1

$\psi_Y(\theta) = \log \mathbb{E}[e^{\theta Y}]$ by definition.

$$\Rightarrow e^{\psi_Y(\theta)} = \mathbb{E}[e^{\theta Y}] = \int f_Y(y) e^{\theta y} dy$$

$$\xRightarrow{\frac{\partial}{\partial\theta}} \psi_Y'(\theta) e^{\psi_Y(\theta)} = \int y e^{\theta y} f_Y(y) dy = \mathbb{E}[Y e^{\theta Y}]$$

$$\xRightarrow{\frac{\partial}{\partial\theta}} \left(\psi_Y''(\theta) + (\psi_Y'(\theta))^2\right) e^{\psi_Y(\theta)} = \int y^2 e^{\theta y} f_Y(y) dy$$

$$= \mathbb{E}[Y^2 e^{\theta Y}]$$

$$\Rightarrow \psi_Y''(\theta) = e^{-\psi_Y(\theta)} \mathbb{E}[Y^2 e^{\theta Y}] - \psi_Y'(\theta)^2$$

$$= \frac{\mathbb{E}[Y^2 e^{\theta Y}]}{\mathbb{E}[e^{\theta Y}]} - \frac{\mathbb{E}[Y e^{\theta Y}]}{\mathbb{E}[e^{\theta Y}]} \quad (\because e^{\psi_Y(\theta)} = \mathbb{E}[e^{\theta Y}]) \quad \cdots \text{①}$$

$$\mathrm{Var}(Y) \le \mathbb{E}\left[\left(Y - \tfrac{b+a}{2}\right)^2\right] \le \tfrac{1}{4}(b-a)^2. \quad (\because |Y - \tfrac{b+a}{2}| \le \tfrac{b-a}{2} \text{ always}) \cdots \text{②}$$

$\hookrightarrow \because \frac{\partial}{\partial\mu} \mathbb{E}[(Y-\mu)^2] = 2\mu - 2\mathbb{E}[Y] = 0$ gives $\mu = \mathbb{E}[Y]$.
Thus, $\mathbb{E}[(Y-\mu)^2]$ minimized at $\mu = \mathbb{E}[Y]$.

$$\frac{\mathbb{E}[Y^2 e^{\theta Y}]}{\mathbb{E}[e^{\theta Y}]} = \int y^2 e^{\theta y} \cdot \frac{1}{\mathbb{E}[e^{\theta Y}]} f_Y(y)\, dy$$

Let $L(y) = \frac{e^{\theta y}}{\mathbb{E}[e^{\theta Y}]} f_Y(y) dy$, $z = \int L(y) dy$.

Then, $\int dz = \int L(y) dy = \int \frac{e^{\theta y}}{\mathbb{E}[e^{\theta Y}]} f_Y(y) dy = \frac{\mathbb{E}[e^{\theta Y}]}{\mathbb{E}[e^{\theta Y}]} = 1$.

Thus, z is a probability measure.

$$\frac{\mathbb{E}[Y^2 e^{\theta Y}]}{\mathbb{E}[e^{\theta Y}]} = \int y^2 L(y) dy = \int y^2 dz = \mathbb{E}_z[Y^2] \cdots ③$$

$$\frac{\mathbb{E}[Y e^{\theta Y}]}{\mathbb{E}[e^{\theta Y}]} = \int y L(y) dy = \int y dz = \mathbb{E}_z[Y] \cdots ④$$

Thus, $\psi''_Y(\theta) \overset{①③④}{=} \mathbb{E}_z[Y^2] - \mathbb{E}_z[Y]^2$

$= \mathrm{Var}_z[Y] \overset{②}{\le} \frac{1}{4}(b-a)^2$.

Now,

$$\begin{cases} \psi'_Y(0) = \frac{\mathbb{E}[Y]}{\mathbb{E}[1]} = 0. \quad \psi_Y(0) = \log \mathbb{E}[1] = \log 1 = 0. \\ \psi_Y(\theta) = \int_0^\theta \psi'_Y(t) dt = \int_0^\theta \int_0^t \psi''_Y(u) du dt \end{cases}$$

$(\theta \ge 0) \Rightarrow \psi_Y(\theta) \le \int_0^\theta \int_0^t \frac{1}{4}(b-a)^2 du dt = \int_0^\theta \frac{1}{4}(b-a)^2 t dt = \frac{1}{8}(b-a)^2 \theta^2$

$(\theta < 0) \Rightarrow \psi_Y(\theta) = \int_0^\theta \int_0^t \psi''_Y(u) du dt = \int_0^\theta - \int_t^0 \psi''_Y(u) du dt$

$$= \int_\theta^0 \int_t^0 \psi_Y''(u)\,du\,dt \le \int_\theta^0 \int_t^0 \frac{1}{4}(b-a)^2\,du\,dt$$

$$= \int_\theta^0 \frac{1}{4}(b-a)^2(-t)\,dt = -\frac{1}{8}(b-a)^2 0^2 - \left(-\frac{1}{8}(b-a)^2\theta^2\right)$$

$$= \frac{1}{8}(b-a)^2\theta^2 .$$

Thus, $\psi_Y(\theta) \le \frac{1}{8}(b-a)^2\theta^2$ $\forall \theta \in \mathbb{R}$, proving Y is sub-Gaussian with parameter $\frac{b-a}{2}$.

#2 By Theorem 2,

$S_n = X_1 + \cdots + X_n$, $X_i \sim \text{Bernoulli}(p)$. $\Rightarrow 0 \le X_i \le 1$.

$$\mathbb{P}(|S_n - np| \ge 0.1np) \le e^{-\left(\frac{0.2np}{\sum (b_i - a_i)^2}\right)} = e^{-0.2p} \quad \cdots ①$$

By Theorem 3,

$$\mathbb{P}(S_n - np \ge 0.1np) = \mathbb{P}(S_n \ge (1+0.1)np)$$

$$\le \left(\frac{e^{0.1}}{1.1^{1.1}}\right)^{np} \cdots ② \le e^{-\frac{1}{3}(0.1)^2 np} \quad \cdots ④$$

$$\mathbb{P}(S_n - np \le -0.1np) = \mathbb{P}(S_n \le (1-0.1)np)$$

$$\le \left(\frac{e^{-0.1}}{0.9^{0.9}}\right)^{np} \cdots ③ \le e^{-\frac{1}{2}(0.1)^2 np}. \quad \cdots ⑤$$

(n.p)	① (theorem 2)	②+③ (theorem 3)	④+⑤ (theorem 3)	Program
(10.0.2)	0.96079	1.98007	1.98341	0.69801
(100, 0.2)	0.96079	1.80938	1.84034	0.62315
(1000, 0.2)	0.96079	0.73494	0.88130	0.11405
(10000, 0.2)	0.96079	0.00009	0.00132	0.00001

For $n = 10, 100$, theorem 2 gives better bound than theorem 3:

Program < ① < ②+③ < ④+⑤

For $n = 1000, 10000$, theorem 3 gives better bound than theorem 2:

Program < ②+③ < ④+⑤ < ①

this is because ① is constant wrt n while ②+③, ④+⑤ are n-dependent. Also, tail bound by software is of course smallest among all.