IEMS 402 Statistical Learning - 2025 Winter
HW4

Tuorui Peng[1]

Exercise 1 Le Cam One-step estimators

For convenience, we label $l_n(\theta) = \dfrac{1}{n}L_n(\theta)$.

Solving the first order equation is equivalent to solving the following equation: (I guess there is a typo in the question setting, should be $\nabla l_n(\hat{\theta}_n) + \nabla^2 l_n(\hat{\theta}_n)\delta_n = 0$)

$$\delta_n = -\left(\nabla^2 l_n(\hat{\theta}_n)\right)^{-1}\nabla l_n(\hat{\theta}_n)$$

in which note that $\hat{\theta}_n - \theta_0 = O(n^{-1/2})$.

Then we have

- $$\nabla l_n(\theta_0) \xrightarrow{\mathrm{d}} \mathcal{N}(0, var(\nabla\ell(\theta_0, X))/n) = \mathcal{N}(0, I(\theta_0)/n)$$

- $$\nabla^2 l_n(\theta_0) = \nabla^2\mathbb{E}\left[\ell(\theta_0, X)\right] + O(n^{-1/2})$$

- $$\begin{aligned}\nabla^2 l_n(\hat{\theta}_n) - \nabla^2 l_n(\theta_0) =& \frac{1}{n}\sum_{i=1}^n \nabla^2\ell(\hat{\theta}_n, X_i) - \frac{1}{n}\sum_{i=1}^n \nabla^2\ell(\theta_0, X_i)\\ \leq& \frac{1}{n}\sum_{i=1}^n M(X_i)\left|\hat{\theta}_n - \theta_0\right|\\ =& \left|\hat{\theta}_n - \theta_0\right|(\mathbb{E}\left[M(X)\right] + O(n^{-1/2}))\\ =& O(n^{-1/2})\end{aligned}$$

- $$\begin{aligned}\nabla l_n(\hat{\theta}_n) - \nabla l_n(\theta_0) =& \frac{1}{n}\sum_{i=1}^n \nabla\ell(\hat{\theta}_n, X_i) - \frac{1}{n}\sum_{i=1}^n \nabla\ell(\theta_0, X_i)\\ =& \frac{1}{n}\sum_{i=1}^n \nabla^2\ell(\theta_0, X_i)(\hat{\theta}_n - \theta_0) + o(\hat{\theta}_n - \theta_0)\\ =& (\nabla^2 l_n(\theta_0) + O(n^{-1/2}))(\hat{\theta}_n - \theta_0) + o(n^{-1/2})\\ =& \nabla^2 l_n(\theta_0)(\hat{\theta}_n - \theta_0) + o(n^{-1/2})\\ =& \nabla^2\mathbb{E}\left[\ell(\theta_0, X)\right](\hat{\theta}_n - \theta_0) + o(n^{-1/2})\end{aligned}$$

[1] TuoruiPeng2028@u.northwestern.edu

Now we have

$$\begin{aligned}\delta_n =& -(\nabla^2 l_n(\hat\theta_n))^{-1}\nabla l_n(\hat\theta_n)\\ =& -(\nabla^2\mathbb{E}\left[\ell(\theta_0,X)\right]+O(n^{-1/2}))^{-1}(\mathcal{N}(0,I(\theta_0)/n)+\nabla^2\mathbb{E}\left[\ell(\theta_0,X)\right](\hat\theta_n-\theta_0)+o(n^{-1/2}))\\ =& -\left(\nabla^2\mathbb{E}\left[\ell(\theta_0,X)\right]^{-1}+O(n^{-1/2})\right)(\mathcal{N}(0,I(\theta_0)/n)+\nabla^2\mathbb{E}\left[\ell(\theta_0,X)\right](\hat\theta_n-\theta_0)+o(n^{-1/2}))\\ \xrightarrow{\mathrm{d}}&\theta_0-\hat\theta_n+\nabla^2\mathbb{E}\left[\ell(\theta_0,X)\right]\mathcal{N}(0,I(\theta_0)/n)+o(n^{-1/2})\end{aligned}$$

then

$$\sqrt{n}(\bar\theta_n-\theta_0)=\sqrt{n}(\hat\theta_n+\delta_n)\xrightarrow{\mathrm{d}}\mathcal{N}(\theta_0,(\nabla^2\mathbb{E}\left[\ell(\theta_0,X)\right])^{-1}I(\theta_0)(\nabla^2\mathbb{E}\left[\ell(\theta_0,X)\right])^{-1})$$

Exercise 2 Sub-Gaussianity of bounded R.V.s

2.(a)

Consider $W:=Y-(a+b)/2$, for which we notice that $|W|\le(b-a)/2$ and $var(W)=var(Y)$. Now we have

$$\begin{aligned}var(Y)=&var(W)\\ =&\mathbb{E}\left[W^2\right]-\mathbb{E}\left[W\right]^2\\ \le&\mathbb{E}\left[W^2\right]\\ \le&\frac{(b-a)^2}{4}\end{aligned}$$

2.(b)

We have

$$\phi'(\lambda)=\frac{\mathbb{E}_P\left[Xe^{\lambda X}\right]}{\mathbb{E}_P\left[e^{\lambda X}\right]}=\mathbb{E}_P\left[X\frac{e^{\lambda X}}{\mathbb{E}_P\left[e^{\lambda X}\right]}\right]=\mathbb{E}_{Q_\lambda}\left[X\right]$$

and

$$\begin{aligned}\phi''(\lambda)=&\frac{\mathbb{E}_P\left[X^2e^{\lambda X}\right]\mathbb{E}_P\left[e^{\lambda X}\right]-\mathbb{E}_P\left[Xe^{\lambda X}\right]^2}{\mathbb{E}_P\left[e^{\lambda X}\right]^2}\\ =&\mathbb{E}_P\left[X^2\frac{e^{\lambda X}}{\mathbb{E}_P\left[e^{\lambda X}\right]}\right]-\left(\mathbb{E}_P\left[X\frac{e^{\lambda X}}{\mathbb{E}_P\left[e^{\lambda X}\right]}\right]^2\right)\\ =&\mathbb{E}_{Q_\lambda}\left[X^2\right]-(\mathbb{E}_{Q_\lambda}\left[X\right])^2\\ =&var_{Q_\lambda}(X)\end{aligned}$$

2.(c)

In this part we consider relabel $\phi(\lambda)=\log\mathbb{E}\left[e^{\lambda(X-\mathbb{E}[X])}\right]$.

Note that if $X\in[a,b]$, then we have

$$\phi''(\lambda)=var_{Q_\lambda}(X)\le\frac{(b-a)^2}{4}$$

on the other hand we notice that $\phi(0) = 0$ and $\phi'(0) = 0$

$$\begin{aligned}\phi'(\lambda) =&\phi'(0) + \int_0^\lambda \phi''(t)dt \leq \frac{(b-a)^2}{4}\lambda\\ \phi(\lambda) =&\phi(0) + \int_0^\lambda \phi'(t)dt \leq \frac{(b-a)^2}{8}\lambda^2\end{aligned}$$

Together we have

$$\phi(\lambda) = \log \mathbb{E}\left[e^{\lambda(X-\mathbb{E}[X])}\right] \leq \frac{(b-a)^2}{8}\lambda^2 \Rightarrow \mathbb{E}\left[e^{\lambda(X-\mathbb{E}[X])}\right] \leq e^{(b-a)^2\lambda^2/8}$$

i.e. X is sub-Gaussian with parameter $(b-a)^2/4$.

Exercise 3 Concentration inequalities

Note that by

3.(a)

By taylor expansion we have

$$\begin{aligned}\mathbb{E}\left[e^{\lambda X}\right] =&\mathbb{E}\left[1 + \lambda X + \frac{\lambda^2 X^2}{2} + \sum_{k=3}^{\infty}\frac{\lambda^k X^k}{k!}\right]\\ \leq& 1 + 0 + \mathbb{E}\left[\frac{X^2}{c^2}\sum_{k=2}^{\infty}\frac{\lambda^k c^k}{k!}\right]\\ =&1 + \frac{\sigma^2}{c^2}(e^{\lambda c} - 1 - \lambda c)\\ \leq& \exp\left[\frac{\sigma^2}{c^2}(e^{\lambda c} - 1 - \lambda c)\right]\end{aligned}$$

3.(b) Bennett's inequality

Using the result from the previous part, we have

$$\begin{aligned}\mathbb{P}\left(\sum_{i=1}^n X_i \geq t\right) =&\mathbb{P}\left(e^{\lambda\sum_{i=1}^n X_i} \geq e^{\lambda t}\right)\\ \leq&\frac{\mathbb{E}\left[e^{\lambda\sum_{i=1}^n X_i}\right]}{e^{\lambda t}}\\ \leq& \exp\left[\frac{\sum_{i=1}^n \sigma_i^2}{c^2}(e^{\lambda c} - 1 - \lambda c) - \lambda t\right]\end{aligned}$$

Optimizing the right hand side with respect to λ, we have optimal $\lambda_* = \dfrac{1}{c}\log\dfrac{ct}{n\sigma^2}$. Substituting this back we have

$$\mathbb{P}\left(\sum_{i=1}^n X_i \geq t\right) \leq \exp\left[\frac{n\sigma^2}{c^2}(\frac{ct}{n\sigma^2} - 1 - \log\frac{ct}{n\sigma^2}) - \frac{t}{c}\log\frac{ct}{n\sigma^2}\right] = \exp\left[-\frac{n\sigma^2}{c^2}h(\frac{ct}{n\sigma^2})\right]$$

3.(c) Bernstein's inequality

It suffices to prove for one sided part (cuz we can apply the whole part to $-X_i$).

$$h(u) = (1+u)\log(1+u) - u$$

We have

$$\mathbb{P}\left(\sum_{i=1}^{n} X_i \geq nt\right) \leq \exp\left[-\frac{n\sigma^2}{c^2}h(\frac{ct}{\sigma^2})\right]$$

so it suffice to prove the following inequality:

$$h(u) \geq \frac{u^2}{2+2u/3}, \quad u \geq 0$$

which is trivial by taking derivative.

$$\begin{aligned}
g(u) :=& (2+2u/3)h(u) - u^2, \quad g(0) = 0 \\
g'(u) =& \frac{4}{3}\log(u+1) + \frac{2}{3}(2(u+1)\log(u+1) - u) - 2u, \quad g'(0) = 0 \\
g''(u) =& \frac{4}{3(u+1)} + \frac{4}{3}\log(u+1) - \frac{4}{3} \geq 0
\end{aligned}$$

With the above, substitute $u = \dfrac{ct}{\sigma^2}$ and we can obtain the desired result.

3.(d)

Berstein' inequality is stronger than Hoeffding's inequality for small t, to be specific, when

$$ct \lesssim \sigma^2, \quad i.e. \quad t \lesssim \frac{\sigma^2}{c}$$

Exercise 4 Application of Concentration Inequalities

It suffices to lower bound the covering number of $\{0,1\}^n$.

WLOG assume $n/4$ is an integer, otherwise we apply the following argument to $\lfloor n/4 \rfloor$.

Assume we have a $n/4$ minimum covering set $\mathcal{S} = \{z_1, \ldots, z_S\}$. For each $i \in [S]$, there are $\sum_{i=0}^{n/4}\binom{n}{n/4}$ points that are in $\mathcal{B}_H(z_i, n/4)$. And this covering sould cover all 2^n points, so we have

$$S\sum_{i=0}^{n/4}\binom{n}{n/4} \geq 2^n \Rightarrow \frac{1}{S} \leq \frac{2^n}{\sum_{i=0}^{n/4}\binom{n}{n/4}} \asymp \mathbb{P}\left(\mathrm{Binom}(n, 1/2) < n/4\right) \leq \exp\left[-\frac{n}{8}\right]$$

As a result, we have

$$S \geq \exp\left[\frac{n}{8}\right]$$

Next note that we have relation between covering number and packing number, thus gives that

$$M(\{0,1\}^n, n/4) \geq N(\{0,1\}^n, n/4) = S \geq \exp\left[\frac{n}{8}\right]$$

this packing set is the set A desired that satisfies $|A| \leq e^{n/8}$, while any $x, y \in A$ satisfies $\|x-y\|_{\text{Hamming}} \geq n/4$.