Prof. Yiping Lu
IEMS 402 Statistical Learning
November 3, 2024

Homework 5: Asymptotic Theory and Concentration Inequality

Question 1. (Le Cam One-step estimators) Let $\{P_\theta\}_{\theta\in\Theta}$ be a family of models where $\Theta \subset \mathbb{R}^d$ is open and let $X_i \overset{\text{iid}}{\sim} P_{\theta_0}$, where P_θ has density p_θ with respect to the measure μ as usual. Assume that $\ell_\theta(x) = \log p_\theta(x)$ is twice continuously differentiable in θ and $\nabla^2 \ell_\theta(x)$ is $M(x)$-Lipschitz, where $\mathbb{E}_{\theta_0}[M^2(X)] < \infty$ for all $\theta \in \Theta$. You may assume that the order of differentiation and expectation can be exchanged.

Suppose that $\hat{\theta}_n$ is a $\sqrt{n}$-consistent estimator, that is,

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = O_{P_{\theta_0}}(1).$$

Let $L_n(\theta) = \sum_{i=1}^n \log p_\theta(X_i)$, where $X_i \overset{\text{iid}}{\sim} P_{\theta_0}$. Consider the one-step estimator δ_n that solves the first-order approximation to $\nabla L_\theta(\theta) = 0$ given by

$$\nabla L_n(\hat{\theta}_n) + \nabla^2 L_n(\hat{\theta}_n)(\delta_n - \hat{\theta}_n) = 0.$$

What is the asymptotic distribution of δ_n?

(*hint*: $\sqrt{n}(\hat{\theta}_n - \theta^*) \xrightarrow{d} \mathcal{N}\left(0, (H^*)^{-1}\Sigma^*(H^*)^{-1}\right)$where matrices $H^* := \nabla^2 F(\theta)$ and $\Sigma^* := \mathbb{E}[\nabla f_\theta(X)\nabla f_\theta(X)^\top]$ both exist and are non-singular.)

Question 2. (Sub-Gaussianity of bounded R.V.s) Let X be a random variable taking values in $[a, b]$ with probability distribution P. You may assume w.l.o.g. that $\mathbb{E}[X] = 0$. Define the cumulant generating function $\varphi(\lambda) := \log \mathbb{E}_P[e^{\lambda X}]$, and let Q_λ be the distribution on X defined by

$$dQ_\lambda(x) := \frac{e^{\lambda x}}{\mathbb{E}_P[e^{\lambda X}]} dP(x).$$

You may assume that differentiation and computation of expectations may be exchanged (this is valid for bounded random variables).

(a) Show that $\mathrm{Var}(Y) \le \frac{(b-a)^2}{4}$ for any random variable Y taking values in $[a, b]$.
(b) Show that $\varphi'(\lambda) = \mathbb{E}_{Q_\lambda}[X]$ and $\varphi''(\lambda) = \mathrm{Var}_{Q_\lambda}(X)$.
(c) Show that $\varphi(\lambda) \le \frac{\lambda^2(b-a)^2}{8}$ for all $\lambda \in \mathbb{R}$.

With these three parts, you have shown that if $X \in [a, b]$, then X is $\frac{(b-a)^2}{4}$-sub-Gaussian.

Question 3. (Concentration inequalities) Let X_i be independent random variables with $|X_i| \le c$ and $\mathbb{E}[X_i] = 0$.

(a) Let $\sigma_i^2 = \mathrm{Var}(X_i)$. Prove that

$$\mathbb{E}[e^{\lambda X}] \le \exp\left(\frac{\sigma_i^2}{c^2}(e^{\lambda c} - 1 - \lambda c)\right).$$

(b) Let $h(u) = (1+u)\log(1+u) - u$ and let $\sigma^2 = \frac{1}{n}\sum_{i=1}^n \sigma_i^2$. Prove *Bennett's inequality*, that is, for any $t \ge 0$ we have

$$\mathbb{P}\left(\sum_{i=1}^n X_i \ge t\right) \le \exp\left(-\frac{n\sigma^2}{c^2} h\left(\frac{ct}{n\sigma^2}\right)\right).$$

(c) Under the notation of part (b), prove *Bernstein's inequality*, that is, that for any $t \ge 0$

$$\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^n X_i \ge t\right) \vee \mathbb{P}\left(\frac{1}{n}\sum_{i=1}^n X_i \le -t\right) \le \exp\left(-\frac{nt^2}{2\sigma^2 + 2ct/3}\right),$$

where $a \vee b = \max\{a, b\}$.

(d) When is Bernstein's inequality tighter than Hoeffding's inequality for bounded random variables? Recall that Hoeffding's inequality states (under the above conditions on X_i) that

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

Further Reading: https://arxiv.org/pdf/1910.02884 and https://terrytao.wordpress.com/2010/01/03/254a-notes-1-concentration-of-measure/

Question 4. (Application of Concentration Inequalities) For any integer $n > 0$, define the Hamming distance on the hypercube $\{0,1\}^n$ by

$$d_H(x,y) := \sum_{i=1}^{n} \mathbf{1}_{x_i \neq y_i}, \quad \text{for any } x, y \in \{0,1\}^n.$$

Show that there exists a universal constant $c > 0$, such that for any $n > 0$, there exists a subset $A \subseteq \{0,1\}^n$ with $|A| \geq e^{cn}$, satisfying

$$d_H(x,y) \geq \frac{n}{4}, \quad \text{for any pair } x, y \in A.$$

[Hint: consider a subset formed by i.i.d. uniform random samples from the hypercube.]

Northwestern University