STA3000F: Homework 2

November 1, 2023

Each question is worth 33.33% points, and if you solve all of them correctly, you will get a total of 133.33% points (i.e. bonus points). We do not distinguish regular and bonus questions.

1 Q1: Logistic regression

Consider d-dimensional random vectors $X_1, X_2, \cdots, X_n \overset{\text{i.i.d.}}{\sim} \mathbb{P}$, with $\mathbb{E}\big[\, \|X\|_2^2 \,\big] < +\infty$. Let the binary labels be generated as

$$Y_i | X_i \sim \text{Ber}(\pi_{\theta^*}(X_i)), \quad \text{for } i = 1, 2, \cdots, n,$$

where we define

$$\pi_\theta(x) := \frac{1}{1 + \exp(-x^\top \theta)}.$$

Consider the maximal likelihood estimator

$$\widehat{\theta}_n := \arg\max \frac{1}{n} \sum_{i=1}^n \Big\{ Y_i \log \pi_\theta(X_i) + (1 - Y_i) \log \big(1 - \pi_\theta(X_i)\big) \Big\}.$$

Let $n \to +\infty$ with everything else fixed. Assume that the Fisher information is non-singular, derive and prove the convergence rate and asymptotic distribution for $\widehat{\theta}_n$.

2 Q2: One-step Newton method

Let K be a compact subset of $\mathbb{R}^d$, and let θ^* lie in the interior of K. Suppose the function $\theta \mapsto f_\theta(x)$ is third order continuously differentiable for any x, satisfying

$$\|\nabla^3 f_\theta(x)\|_{\text{tsr}} \le L(x), \quad \text{for any } x$$

in a local neighborhood of θ^*, with $\mathbb{E}[L(X)] < +\infty$. Furthermore, assume that the population-level loss function $F(\theta) := \mathbb{E}[f_\theta(X)]$ is uniquely minimized at θ^*, and that the matrices $H^* := \nabla^2 F(\theta)$, $\Sigma^* := \mathbb{E}[\nabla f_\theta(X) \nabla f_\theta(X)^\top]$ both exist and are non-singular. (Basically the "classical conditions" in our lectures).

Let $\widetilde{\theta}_n$ be an estimator satisfying $\Big\| \widetilde{\theta}_n - \theta^* \Big\|_2 = O_p(n^{-1/2})$. Let $F_n(\theta) := \frac{1}{n} \sum_{i=1}^n f_\theta(X_i)$ be the empirical loss. Consider the one-step Newton estimator

$$\widehat{\theta}_n := \widetilde{\theta}_n - \nabla^2 F_n(\widetilde{\theta}_n)^{-1} \nabla F_n(\widetilde{\theta}_n).$$

Show that

$$\sqrt{n}\big(\widehat{\theta}_n - \theta^*\big) \xrightarrow{d} \mathcal{N}\big(0, (H^*)^{-1} \Sigma^* (H^*)^{-1}\big).$$

3 Q3: Empirical CDF in dimension one

In this question (and throughout the rest of this class), you will use the following refined empirical process lemma without having to prove it.

Lemma 1 (Lemma 19.34 of [VdV00]). *Given a function class $\mathcal{F}$, if $\mathbb{E}[f(X)^2] \le \eta^2$ for any $f \in \mathcal{F}$. Let F be a dominating function, such that $|f(x)| \le F(x)$ for any $f \in \mathcal{F}$ and any x. Define the scalar*

$$a(\eta) := \frac{\eta \|F\|_{L^2(\mathbb{P})}}{\sqrt{\log N_{[]}\big(\eta \|F\|_{L^2(\mathbb{P})}; \mathcal{F}, L^2(\mathbb{P})\big)}},$$

we have that

$$\mathbb{E}\Big[\sup_{f\in\mathcal{F}} |(P_n - P)f|\Big] \le \frac{c\|F\|_{L^2(P)}}{\sqrt{n}} \int_0^\eta \sqrt{\log N_{[]}\big(\delta \|F\|_{L^2(\mathbb{P})}; \mathcal{F}, L^2(\mathbb{P})\big)}\, d\delta + \mathbb{E}\Big[F(X) \cdot \mathbf{1}\big[|F(X)| \ge a(\eta)\sqrt{n}\big]\Big].$$

Taking this lemma as given, prove the following result.

Consider samples $X_1, X_2, \cdots, X_n \overset{\text{i.i.d.}}{\sim} \text{Uniform}([0,1])$. Define the empirical cumulative distribution function (CDF)

$$F_n(t) := \frac{1}{n}\sum_{i=1}^n \mathbf{1}[X_i \le t],$$

and the population-level CDF $F(t) = t$, for $t \in [0,1]$. Derive and prove the asymptotic distribution of the following stochastic process

$$\Big(\sqrt{n}\big(F_n(t) - F(t)\big) \ : \ t \in [0,1]\Big).$$

It is OK if you don't know the name of the limiting stochastic process. You will get the full credit as long as you figure out the covariance structure of the limiting Gaussian process, and prove the process convergence.

4 Q4: Covariance matrices

Suppose that the zero-mean d-dimensional random vector X satisfies

$$\mathbb{E}\Big[\exp(\lambda v^\top X)\Big] \le e^{\lambda^2/2}, \quad \text{for any } v \in \mathbb{R}^d \text{ such that } \|v\|_2 = 1.$$

Let $X_1, X_2, \cdots, X_n$ be i.i.d. copies of X. Define the empirical and population covariance matrices

$$\widehat{\Sigma}_n := \frac{1}{n}\sum_{i=1}^n X_i X_i^\top, \quad \text{and} \quad \Sigma := \mathbb{E}[XX^\top].$$

Prove the following claim: there exists universal constants $c_1, c_2 > 0$, such that whenever $n \ge c_1 d$, we have

$$|||\widehat{\Sigma}_n - \Sigma|||_{\text{op}} \le c_2\sqrt{\frac{d}{n}},$$

with probability at least 3/4. The constants (c_1, c_2) are independent of any problem parameters. In principle, we can find explicit constants (100, 1000, etc), but the numerical values are not important, and you will get full credit no matter what exact constants you get.

[Hint: re-write the operator norm into the variational form, i.e., as an optimization problem, and use tools of covering and packing.]

References

[VdV00] Aad W Van der Vaart. *Asymptotic statistics*, volume 3. Cambridge university press, 2000.