Suppose that we observe data in pairs $(x, y) \in \mathbb{R}^d \times \{\pm 1\}$ where the data come from a logistic model with $X \sim P_0$ and

$$p_{y|x}(y|x) = \frac{1}{1 + \exp\{-yx^\top \theta\}}$$

with log-loss function $l_\theta(y|x) = \log(1 + \exp\{-yx^\top \theta\})$. Let $\hat{\theta}$ minimize the empirical logistic loss

$$L_n(\theta) := n^{-1} \sum_{i=1}^{n} l_\theta(Y_i|X_i) = n^{-1} \sum_{i=1}^{n} \log(1 + e^{-Y_iX_i^\top \theta})$$

for pairs $(X_i, Y_i)$ drawn from the logistic model with parameter $\theta_0$. Assume in addition that the data $X_i \in \mathbb{R}^d$ are i.i.d. and satisfy

$$E[X_iX_i^\top] = \Sigma > 0, \quad \text{and} \quad E\|X_i\|_2^4 < \infty,$$

that is the second moment of the $X_i$ is positive definite.

(a) Let $L(\theta) = E[l_\theta(Y|X)]$ denote the population logistic loss. Show that the second order derivative evaluated at $\theta_0$ is positive definite. You may assume that the order of differentiation and integration may be exchanged.

(b) Under these assumptions show that $\hat{\theta}$ is consistent estimator of $\theta_0$ when $n \to \infty$. Provide details of your work (definitions, theorems used should be cited from the notes).

(c) Provide an asymptotic distribution of $\sqrt{n}(\hat{\theta} - \theta_0)$. You may assume here that $\hat{\theta}$ is consistent. Assume that $d = 1$ and even simpler setting where $x \in \{-1, 1\}$.

(d) Describe the effect of $\theta_0$ on the efficiency of $\hat{\theta}$. 

---

**Date:** September, 2019.