Case Study

Xinjie Qian

September 30, 2021

Outline

1. Partly Linear Logistic Regression

2. Testing for a Change-point

Partly Linear Logistic Regression

As discussed in Chapter 1 and Chapter 4, the observed data are n independent realizations of the random triplet (Y, Z, U), where $Z \in \mathbb{R}^p$ and $U \in \mathbb{R}$ are covariates which are not linearly dependent, Y is a dichotomous outcome with conditional expectation $\nu[\beta' Z + \eta(U)]$, $\beta \in \mathbb{R}^p$, Z is restricted to a bounded set, $U \in [0, 1]$, $\nu(t) = 1/(1 + e^{-t})$, and where η is an unknown smooth function. Hereafter, for simplicity, we will also assume that $p = 1$. We further assume , for some integer $k \geq 1$, that the first $k - 1$ derivatives of η exist and are absolutely continuous with $J^2(\eta) = \int_0^1 [\eta^{(k)}(t)]^2 dt < \infty$.

Partly Linear Logistic Regression

To estimate β and η based on an i.i.d. sample $X_i = (Y_i, Z_i, U_i), i = 1, ..., n$, we can use the following penalized log-likelihood:

$$\tilde{L}_n(\beta, \eta) = n^{-1} \sum_{i=1}^{n} \log p_{\beta,\eta}(X_i) - \hat{\lambda}_n^2 J^2(\eta),$$

where the conditional density at $Y = y$ given the covariates $(Z, U) = (z, u)$ has the form

$$p_{\beta,\eta}(x) = \{\nu[\beta z + \eta(u)]\}^y \{1 - \nu[\beta z + \eta(u)]\}^{1-y}$$

and $\hat{\lambda}_n$ is chosen to satisfy $\hat{\lambda}_n = o_P(n^{-1/4})$ and $\hat{\lambda}_n^{-1} = O_P(n^{k/(2k+1)})$. Denote $\hat{\beta}_n$ and $\hat{\eta}_n$ to be the maximizers of $\tilde{L}_n(\beta, \eta)$, let $P_{\beta,\eta}$ denote expectation under the model, and let β_0 and η_0 to be the true values of the parameters.

Partly Linear Logistic Regression

Now we want to prove that both $\hat{\beta}_n$ and $\hat{\eta}_n$ are uniformly consistent. First, Theorem 9.21 readily implies, after some rescaling of $\mathcal{H}_c$, that

$$\log N_{[]}(\epsilon, \mathcal{H}_c, L_2(P)) \leq M\epsilon^{-1/k}, \tag{15.1}$$

for all $\epsilon > 0$, where M only depends on c and P. This readily yields that $\mathcal{H}_c$ is Donsker.
Define $\theta \equiv (\beta, \eta)$, and let $\ell_\theta(x) \equiv y\omega_\theta(x) - \log(1 + e^{\omega_\theta(x)})$ and $\omega_\theta(x) \equiv z\beta + \eta(u)$.

Theorem 15.1

Under the given assumptions, $J(\hat{\eta}_n) = O_P(1)$ and

$$\|\omega_{\hat{\theta}_n}(X) - \omega_{\theta_0}(X)\|_{P,2} = O_P(n^{-k/(2k+1)}). \tag{15.2}$$

Proof of Theorem 15.1

Proof

We first need to establish a fairly precise bound on the bracketing entropy of

$$\mathcal{G} \equiv \left\{ \frac{\ell_\theta(X) - \ell_{\theta_0}(X)}{1 + J(\eta)} : |\beta - \beta_0| \le c_1, \|\eta - \eta_0\|_\infty \le c_1, J(\eta) < \infty \right\},$$

which satisfies $\mathcal{G} \subset \mathcal{G}_1 + \mathcal{G}_2(\mathcal{G}_1)$, where

$$\mathcal{G}_1 = \left\{ \frac{\omega_\theta(X) - \omega_{\theta_0}(X)}{1 + J(\eta)} : |\beta - \beta_0| \le c_1, \|\eta - \eta_0\|_\infty \le c_1, J(\eta) < \infty \right\},$$

$\mathcal{G}_2$ consists of all functions $t \mapsto a^{-1} \log(1 + e^{at})$ with $a \ge 1$, and $\mathcal{G}_2(\mathcal{G}_1) \equiv \{g_2(g_1) : g_1 \in \mathcal{G}_1, g_2 \in \mathcal{G}_2\}$.

Proof of Theorem 15.1

Proof (cont).
By (15.1) combined with the properties of bracketing entropy and the fact that $J(\eta_0)$ is a finite constant, it is not hard to verify that there exists an $M_0 < \infty$ such that $\log N_{[]}(\epsilon, \mathcal{G}_1, L_2(P)) \leq M_0\epsilon^{-1/k}$. Now by Exercise 15.6.1 combined with Lemma 15.2 below, we obtain that there exists a $K_1 < \infty$ such that $\log N_{[]}(\epsilon, \mathcal{G}_2(\mathcal{G}_1), L_2(P)) \leq K_1\epsilon^{-1/k}$, for every $\epsilon > 0$. Combining this with preservation properties of bracketing entropy (see Lemma 9.25), we obtain that there exists an $M_1 < \infty$ such that $\log N_{[]}(\epsilon, \mathcal{G}, L_2(P)) \leq M_1\epsilon^{-1/k}$ for all $\epsilon > 0$.

Proof of Theorem 15.1

Proof (cont).

Combining previous result with Theorem 15.3 below, we obtain that

$$|(\mathbb{P}_n - P)(\ell_{\hat{\theta}_n}(X) - \ell_{\theta_0}(X))| = O_P(n^{-1/2}(1 + J(\hat{\eta}))) \times \left[\left\| \frac{\ell_{\hat{\theta}_n}(X) - \ell_{\theta_0}(X)}{1 + J(\hat{\eta}_n)} \right\|_{P,2}^{1-1/(2k)} \vee n^{-(2k-1)/[2(2k+1)]} \right]. \tag{15.3}$$

Now note that by a simple Taylor expansion and the boundedness constraints on the parameters, there exists a $c_1 > 0$ and a $c_2 < \infty$ such that

$$P(\ell_{\hat{\theta}_n}(X) - \ell_{\theta_0}(X)) \le -c_1 P[\omega_{\hat{\theta}_n}(X) - \omega_{\theta_0}(X)]^2$$

and

$$|\ell_{\hat{\theta}_n}(X) - \ell_{\theta_0}(X)| \le c_2 |\omega_{\hat{\theta}_n}(X) - \omega_{\theta_0}(X)|,$$

almost surely.

Proof of Theorem 15.1

Proof (cont).
Combining this with (15.3) and a simple Taylor expansion, we can readily establish that

$$\begin{aligned}
\lambda_n^2 J^2(\hat{\eta}_n) &\leq \lambda_n^2 J^2(\eta_0) + (\mathbb{P} - P)(\ell_{\hat{\theta}_n}(X) - \ell_{\theta_0}(X)) \\
&\quad + P(\ell_{\hat{\theta}_n}(X) - \ell_{\theta_0}(X)) \\
&\leq O_P(\lambda_n^2) + O_P(n^{-1/2}(1 + J(\hat{\eta}))) \\
&\quad \times \left[\left\| \frac{\omega_{\hat{\theta}_n}(X) - \omega_{\theta_0}(X)}{1 + J(\hat{\eta}_n)} \right\|_{P,2}^{1-1/(2k)} \vee n^{-(2k-1)/[2(2k+1)]} \right] \\
&\quad - c_1 P[\omega_{\hat{\theta}_n}(X) - \omega_{\theta_0}(X)]^2,
\end{aligned}$$

from which we can deduce that both

Proof of Theorem 15.1

Proof (cont).

$$\frac{J^2(\hat{\eta}_n)}{1+J(\hat{\eta}_n)} = O_P(1) + O_P(n^{(2k-1)/[2(2k+1)]}) \times \left[\left\| \frac{\omega_{\hat{\theta}_n}(X) - \omega_{\theta_0}(X)}{1+J(\hat{\eta}_n)} \right\|_{P,2}^{1-1/(2k)} \vee n^{-(2k-1)/[2(2k+1)]} \right] \quad (15.4)$$

and

$$\left\| \frac{\omega_{\hat{\theta}_n}(X) - \omega_{\theta_0}(X)}{1+J(\hat{\eta}_n)} \right\|_{P,2}^{2} = O_P(\lambda_n^2) + O_P(n^{-1/2}) \times \left[\left\| \frac{\omega_{\hat{\theta}_n}(X) - \omega_{\theta_0}(X)}{1+J(\hat{\eta}_n)} \right\|_{P,2}^{1-1/(2k)} \vee n^{-(2k-1)/[2(2k+1)]} \right]. \quad (15.5)$$

Proof of Theorem 15.1

Proof (cont).
Letting $A_n \equiv n^{k/(2k+1)}(1 + J(\hat{\eta}_n))^{-1}\|\omega_{\hat{\theta}_n}(X) - \omega_{\theta_0}(X)\|_{P,2}$, we obtain from (15.5) that $A_n^2 = O_P(1) + O_P(1)A_n^{1-1/(2k)}$. Solving this yields $A_n = O_P(1)$, which implies

$$\frac{\|\omega_{\hat{\theta}_n}(X) - \omega_{\theta_0}(X)\|_{P,2}}{1 + J(\hat{\eta}_n)} = O_P(n^{-k/(2k+1)}).$$

Applying this to (15.4) now yields $J(\hat{\eta}_n) = O_P(1)$, which implies $\|\omega_{\hat{\theta}_n}(X) - \omega_{\theta_0}(X)\|_{P,2} = O_P(n^{-k/(2k+1)})$, and the proof is complete.

Partly Linear Logistic Regression

We now use (15.2) to obtain that both $\hat{\beta}_n$ and $\hat{\eta}_n$ are uniformly consistent and optimality of the L_2 convergence rate of $\hat{\eta}_n$.

Recall that $\tilde{h}_1(u) \equiv E[Z|U = u]$ and that $P[Z - \tilde{h}_1(U)]^2 > 0$ by assumption. Since $E[(Z - \tilde{h}_1(U))g(U)] = 0$ for all $g \in L_2(U)$, (15.2) implies $|\hat{\beta}_n - \beta_0| = O_P(n^{-k/(2k+1)})$ and thus $\hat{\beta}_n$ is consistent. These results now imply that $P[\hat{\eta}_n(U) - \eta_0(U)]^2 = O_P(n^{-2k/(2k+1)})$, and thus we have L_2 optimality of $\hat{\eta}_n$ because of the assumptions on the density of U. Uniform consistency of $\hat{\eta}_n$ now follows form the fact that $J(\hat{\eta}_n) = O_P(1)$ forces $u \mapsto \hat{\eta}_n(u)$ to be uniformly equicontinuous in probability.

Partly Linear Logistic Regression

Lemma 15.2

For a probability measure P, let $\mathcal{F}_1$ be a class of measurable functions $f : \mathcal{X} \mapsto \mathbb{R}$, and let $\mathcal{F}_2$ denote a class of nondecreasing functions $f_2 : \mathbb{R} \mapsto [0, 1]$ that are measurable for every probability measure. Then,

$$\log N_{[]}(\epsilon, \mathcal{F}_2(\mathcal{F}_1), L_2(P)) \leq 2 \log N_{[]}(\epsilon/3, \mathcal{F}_1, L_2(P)) + \sup_Q \log N_{[]}(\epsilon/3, \mathcal{F}_2, L_2(Q)),$$

for all $\epsilon > 0$, where the supremum is over all probability measures Q.

Proof of Lemma 15.2

Proof.

Fix $\epsilon > 0$, and let $\{[f_k, g_k], 1 \leq k \leq n_1\}$ ba a minimal $L_2(P)$ bracketing $\epsilon/3$-cover for $\mathcal{F}_1$, where f_k is the lower- and g_k is the upper-boundary function for the bracket. For each f_k, construct a minimal $L_2(Q_{k,1})$ bracketing $\epsilon/3$-cover for $\mathcal{F}_1(f_k(x))$, where $Q_{k,1}$ is the distribution of $f_k(X)$. Let $n_2 = \sup_Q \log N_{[]}(\epsilon/3, \mathcal{F}_2, L_2(Q))$, and choose a corresponding minimal cover $\{[f_{k,j,1}, g_{k,j,1}], l \leq j \leq n_2\}$. Construct a similar cover $\{[f_{k,j,2}, g_{k,j,2}], l \leq j \leq n_2\}$ for each $\mathcal{F}_1(g_k(x)), 1 \leq k \leq n_1$.

Proof of Lemma 15.2

Proof (cont).
Let $h_1 \in \mathcal{F}_1$ and $h_2 \in \mathcal{F}_2$ be arbitrary; let $[f_k, g_k]$ be the bracket containing h_1; let $[f_{k,j,1}, g_{k,j,1}]$ be the bracket containing $h_2(f_k)$; and let $[f_{k,j,2}, g_{k,j,2}]$ be the bracket containing $h_2(g_k)$. Then $[f_{k,j,1}(f_k), g_{k,j,2}(g_k)]$ is an $L_2(P)$ ϵ-bracket which satisfies $f_{k,j,1}(f_k) \leq h_2(f_k) \leq h_2(h_1) \leq h_2(g_k) \leq g_{k,j,2}(g_k)$. Thus, since f_1 and f_2 were arbitrary, the number of $L_2(P)$ ϵ-brackets needed to completely cover $\mathcal{F}_2(\mathcal{F}_1)$ is bounded by

$$N_{[]}^2(\epsilon/3, \mathcal{F}_1, L_2(P)) \times \sup_Q N_{[]}(\epsilon/3, \mathcal{F}_2, L_2(Q)),$$

and the desired result follows.

Partly Linear Logistic Regression

Theorem 15.3

Let $\mathcal{F}$ be a uniformly bounded class of measurable functions such that for some measurable f_0, $\sup_{f\in\mathcal{F}} \|f - f_0\|_\infty < \infty$. Moreover, assume that $\log N_{[]}(\epsilon, \mathcal{F}, L_2(P)) \leq K_1\epsilon^{-\alpha}$ for some $K_0 < \infty$ and $\alpha \in (0,2)$ and for all $\epsilon > 0$. Then

$$\sup_{f\in\mathcal{F}} \left[\frac{|(\mathbb{P} - P)(f - f_0)|}{\|f - f_0\|_{P,2}^{1-\alpha/2} \vee n^{-(2-\alpha)/[2(2+\alpha)]}} \right] = O_P(n^{-1/2}).$$

Note: This is a result presented on Page 79 of van de Geer (2000).

Testing for a Change-point

Recall the change-point model example of Section 14.5.1 and consider testing the null hypothesis $H_0 : \alpha = \beta$. Under this null, the change-point parameter ζ is not identifiable, and thus $\hat{\zeta}$ is not consistent. This means that testing H_0 is an important concern, since it is unlikely we would know in advance whether H_0 were true. The statistic we propose using is $T_n \equiv sup_{\zeta \in [a,b]} |U_n(\zeta)|$, where

$$U_n(\zeta) \equiv \frac{\sqrt{n\hat{F}_n(\zeta)(1-\hat{F}_n(\zeta))}}{\hat{\sigma}_n}\left(\frac{\sum_{i=1}^n 1\{Z_i \le \zeta\}Y_i}{n\hat{F}_n(\zeta)} - \frac{\sum_{i=1}^n 1\{Z_i \le \zeta\}Y_i}{n(1-\hat{F}_n(\zeta))}\right),$$

$\hat{\sigma}_n^2 \equiv n^{-1}\sum_{i=1}^n (Y_i - \hat{\alpha}_n 1\{Z_i \le \hat{\zeta}_n\} - \hat{\beta}_n 1\{Z_i > \hat{\zeta}_n\})^2$, and where $\hat{F}_n(t) \equiv \mathbb{P}_n 1\{Z \le t\}$.

Testing for a Change-point

We will study the asymptotic limiting behavior of this statistic under the sequence of contiguous alternative hypotheses $H_{1n} : \beta = \alpha + \eta/\sqrt{n}$, where the distribution of Z and ϵ does not change with n. Thus
$Y_i = \epsilon + \alpha_0 + (\eta/\sqrt{n})1\{Z_i > \zeta_0\}, i = 1, ..., n$.
We will first show that under H_{1n},

$$U_n(\zeta) = a_0(\zeta)B_n(\zeta) + \nu_0(\zeta) + r_n(\zeta), \tag{15.6}$$

where $a_0(\zeta) \equiv \sigma^{-1}\sqrt{F(\zeta)(1 - F(\zeta))}, F(\zeta) \equiv P1\{Z \leq \zeta\}$,
$B_n(\zeta) \equiv \frac{\sqrt{n}\mathbb{P}_n[1\{\zeta_0 < Z \leq \zeta\}\epsilon]}{F(\zeta)} - \frac{\sqrt{n}\mathbb{P}_n[1\{Z > \zeta \vee \zeta_0\}\epsilon]}{1-F(\zeta)}$,
$\nu_0(\zeta) \equiv \eta a_0(\zeta)\left(\frac{P[\zeta_0 < Z \leq \zeta]}{F(\zeta)} - \frac{P[Z > \zeta \vee \zeta_0]}{1-F(\zeta)}\right)$,
and $\sup_{\zeta \in [a,b]} |r_n(\zeta)| \xrightarrow{P} 0$.

Testing for a Change-point

The first step is to note that $\hat{F}_n$ is uniformly consistent on $[a, b]$ for F and that both $\inf_{\zeta \in [a,b]} F(\zeta) > 0$ and $\inf_{\zeta \in [a,b]}(1 - F(\zeta)) > 0$. Since $\eta/\sqrt{n} \to 0$, we also have that both

$$V_n(\zeta) \equiv \frac{\sum_{i=1}^n 1\{Z_i \le \zeta\} Y_i}{\sum_{i=1}^n 1\{Z_i \le \zeta\}}, \text{and} \quad W_n(\zeta) \equiv \frac{\sum_{i=1}^n 1\{Z_i \ge \zeta\} Y_i}{\sum_{i=1}^n 1\{Z_i \ge \zeta\}}$$

are uniformly consistent, over $\zeta \in [a, b]$. Since $\hat{\zeta}_n \in [a, b]$ with probability 1 by assumption, we have $\hat{\sigma}_n^2 \xrightarrow{P} \sigma^2$ even though $\hat{\zeta}_n$ may not be consistent. Now the fact that both $\mathcal{F}_1 \equiv \{1\{Z \le \zeta\} : \zeta \in [a, b]\}$ and $\mathcal{F}_2 \equiv \{1\{Z > \zeta\} : \zeta \in [a, b]\}$ are Donsker yields the final conclusion, after some simple calculations.

Testing for a Change-point

Reapplying the fact that $\mathcal{F}_1$ and $\mathcal{F}_2$ are Donsker yields that $a_0(\zeta)B_n(\zeta) \rightsquigarrow \mathcal{B}_0(\zeta)$ in $\ell^\infty([a,b])$, where $\mathcal{B}_0$ is a tight, mean zero Gaussian process with covariance

$$H_0(\zeta_1, \zeta_2) \equiv \sqrt{\frac{F(\zeta_1 \wedge \zeta_2)(1 - F(\zeta_1 \vee \zeta_2))}{F(\zeta_1 \vee \zeta_2)(1 - F(\zeta_1 \wedge \zeta_2))}}.$$

Thus, from (15.6), we have the $U_n \rightsquigarrow \mathcal{B}_0 + \nu_0$, it is easy to verify that $\nu_0(\zeta_0) \neq 0$. Hence, if we use critical values based on $\mathcal{B}_0$, the statistic T_n will asymptotically have the correct size under H_0 as well as have arbitrarily large power under H_{1n} as $|\eta|$ gets larger.

Testing for a Change-point

Thus what we need now is a computationally easy method for obtaining the critical values of $\sup_{\zeta\in[a,b]}|\mathcal{B}_0(\zeta)|$. Define

$$A_i^n(\zeta) \equiv \sqrt{\hat{F}_n(\zeta)(1-\hat{F}_n(\zeta))}\Big(\frac{1\{Z_i \le \zeta\}}{\hat{F}_n(\zeta)} - \frac{1\{Z_i > \zeta\}}{1-\hat{F}_n(\zeta)}\Big),$$

and consider the weighted "bootstrap" $\mathcal{B}_n(\zeta) \equiv n^{-1/2}\sum_{i=1}^n A_i^n(\zeta)\omega_i$, where $\omega_1, ..., \omega_n$ are i.i.d. standard normals independent of the data. The continuous mapping theorem applied to the following lemma verifies that this bootstrap will satisfy $\sup_{\zeta\in[a,b]}|\mathcal{B}_n(\zeta)| \underset{\omega}{\overset{P}{\rightsquigarrow}} \sup_{\zeta\in[a,b]}|\mathcal{B}_0(\zeta)|$ and thus can be used to obtain the needed critical values:

Testing for a Change-point

Lemma 15.4

$\mathcal{B}_n \underset{\omega}{\overset{P}{\rightsquigarrow}} \mathcal{B}_0$ in $\ell^\infty([a,b])$.

Proof

Let $\mathcal{G}$ be the class of nondecreasing functions $G : [a,b] \mapsto [a_0, b_0]$ of ζ, where $a_0 \equiv F(a)/2$ and $b_0 \equiv 1/2 + F(b)/2$, and note that by the uniform consistency of $\hat{F}_n$, $\hat{F}_n \in \mathcal{G}$ with probability arbitrarily close to 1 for all n large enough. Also note that

$$\left\{ \sqrt{G(\zeta)(1-G(\zeta))} \Big(\frac{1\{Z \leq \zeta\}}{G(\zeta)} - \frac{1\{Z > \zeta\}}{1 - G(\zeta)} \Big) : \zeta \in [a,b], G \in \mathcal{G} \right\}$$

is Donsker, since $\{a\mathcal{F} : a \in K\}$ is Donsker for any compact $K \subset \mathbb{R}$ and any Donsker class $\mathcal{F}$.

Proof of Lemma 15.4

Proof (cont).

Thus by the multiplier central limit theorem, Theorem 10.1, we have that $\mathcal{B}_n$ converges weakly to a tight Gaussian process unconditionally. Thus we know that $\mathcal{B}_n$ is asymptotically tight conditionally, i.e., we know that

$$P^*\left(\sup_{\zeta_1,\zeta_2\in[a,b]:|\zeta_1-\zeta_2|\leq\delta_n}|\mathcal{B}_n(\zeta_1)-\mathcal{B}_n(\zeta_2)|>\tau\middle|X_1,...,X_n\right)=o_P(1),$$

for all $\tau > 0$ and all sequences $\delta_n \downarrow 0$. All we need to verify now is that the finite dimensional distributions of $\mathcal{B}_n$ converge to the appropriate limiting multivariate normal distributions conditionally.

Proof of Lemma 15.4

Proof (cont).

Since $\omega_1, \ldots, \omega_n$ are i.i.d. standard normal, it is easy to see that $\mathcal{B}_n$ is conditionally a Gaussian process with mean zero and, after some algebra (see Exercise 15.6.8), with covariance

$$\hat{H}_n(\zeta_1, \zeta_2) \equiv \sqrt{\frac{\hat{F}_n(\zeta_1 \wedge \zeta_2)(1 - \hat{F}_n(\zeta_1 \vee \zeta_2))}{\hat{F}_n(\zeta_1 \vee \zeta_2)(1 - \hat{F}_n(\zeta_1 \wedge \zeta_2))}}. \tag{15.7}$$

Since $\hat{H}_n$ can easily be shown to converge to $H_0(\zeta_1, \zeta_2)$ uniformly over $[a, b] \times [a, b]$, the proof is complete.

Testing for a Change-point

We note that the problem of testing a null hypothesis under which some of the parameters are no longer identifiable is quite challenging, especially when infinite-dimensional parameters are present. An example of the latter setting is in transformation regressionmodels with a change-point (Kosorok and Song, 2007). Obtaining the critical value of the appropriate statistic for these models is much more difficult than it is for the simple example we have presented in this section.