경제학석사 학위논문

An Essay on Le Cam's Method

르 캄의 방법론에 대한 고찰

2020년 8월

서울대학교 대학원

경 제 학 부

이 희 준

An Essay on Le Cam's Method

지도교수 서 명 환

이 논문을 경제학석사 학위논문으로 제출함

2020 년 7 월

서울대학교 대학원

경제학부

이 희 준

이희준의 석사 학위논문을 인준함

2020 년 7 월

위 원 장 ______황 윤 재______ (인)

부위원장 ______서 명 환______ (인)

위 원 ______Okui, Ryo______ (인)

Abstract

An Essay on Le Cam's Method

Heejun Lee

Department of Economics

The Graduate School

Seoul National University

In this paper, we consider the threshold linear regression model $y_i = x_i^T\beta + x_i^T\delta \cdot 1[q_i > \gamma] + u_i$, where y_i is a dependent variable, $x_i \in \mathbb{R}^d$ are regressors, q_i is a threshold variable, u_i is a Gaussian noise, and β and δ are unknown regression vectors. This paper develops three lower bounds of minimax convergence rates for the estimation of the unknown threshold location γ for ℓ_1-loss under three different threshold types. We show that when there is a jump threshold, the convergence of minimax risk is lower bounded by $n^{2\alpha-1}$ rate, where n is the sample size and α means the diminishing rate of a threshold. In addition, we provide a lower bound of $n^{-1/2}$ rate for the kink threshold. Finally, we prove that if the threshold type is unknown, $n^{-1/3}$ rate becomes a lower bound of minimax risk. These rates are equivalent to the convergence rates of least-square based estimators.

Our proofs are based on Le Cam's method which is a technique that connects the minimax risk to the divergences among parameters. One of the widely used divergence measures to apply Le Cam's method is

Kullback-Leibler divergence. We briefly discuss Le Cam's method formulated with Kullback-Leibler divergence and the limitation of the Le Cam's method.

keywords : Minimax Criterion, Threshold Estimation
Student Number : 2017-26997

목　　차

1. Introduction ······ 1

2. Le Cam's Method ······ 3

2. 1 Minimax Framework ······ 3

2. 2 Le Cam's Method ······ 4

2. 3 KL-version Le Cam's Method ······ 6

3. Applications to Threshold Estimation ······ 8

3. 1 Jump Threshold ······ 10

3. 2 Kink Threshold ······ 11

3. 3 Unknown Threshold Type ······ 12

4. Conclusion ······ 14

References ······ 15

국문초록 ······ 17

An Essay on Le Cam's Method

Heejun Lee *

Seoul National University

July, 2020

Abstract

This paper develops three lower bounds of minimax convergence rates for the estimation of unknown threshold location for ℓ_1-loss under three different types of thresholds. In addition, we discuss Le Cam's method which is a technique that provides lower bounds of minimax risks.

1 Introduction

Non-linear effect of independent variables to the dependent variable is one of the major limitations of the linear regression model. The simplest form of non-linearity assumes that regression coefficients change at a threshold point. Consider the canonical threshold regression model:

$$y_i = x_i^T \beta + x_i^T \delta \mathbb{I}[q_i > \gamma] + u_i,$$

where y_i is a dependent variable, q_i is a threshold variable, γ is a location of threshold, x_i is a d-dimensional vector of regressors, u_i is a noise, and $\mathbb{I}[\cdot]$ is the indicator function.

*Department of Economics, Master's degree thesis.

There are extensive researches on the asymptotic behaviors of least-square estimators (LSE). One of the most notable facts is that the convergence rate of LSE is susceptible to the type of threshold. For example, constrained LSE converges in $n^{-1/2}$ rate for continuous threshold (Hansen, 2017), LSE converges in n^{-1} rate for fixed jump threshold (Chan, 1993), $n^{2\alpha-1}$ rate for diminishing jump threshold, $\delta_n = c \cdot n^{-\alpha}$ (Hansen, 2000), and surprisingly, LSE converges in $n^{-1/3}$ rate for continuous threshold (Hidalgo, Lee, and Seo, 2019).

After Wald (1939) suggested minimax criterion as a performance measure of statistical decision procedures, extensive literature adopts this criterion even in recent years; see, for example, Duchi, Jordan, and Wainwright (2018). However, it is infeasible to propose a minimax estimator under the classical minimax criterion except for extremely simplistic cases. Therefore, one usually focuses on finding the minimax convergence rate and relationships between the constant factor and nuisance parameters. Another approach to circumvent such difficulty is to develop an optimal estimator in the locally asymptotically minimax (LAM) sense (Hájek, 1972). Yu (2012, 2015) shows the optimality of Bayesian estimators under the LAM criterion for the threshold estimation problem where the threshold has a fixed discontinuity.

Our major goal in this paper is to justify the $n^{-1/3}$ convergence rate of LSE for the continuous threshold. Since we only consider the rate of convergence, we take the classical minimax criterion instead of the LAM criterion. In this direction, Wang and Samworth (2018) develops a minimax lower bound of $n^{2\alpha-1}$ rate under the high-dimensional, constant regressors, and diminishing jump threshold assumption. It also provides an upper bound of minimax risk that matches the lower bound up to $\log \log n$ factor.

Le Cam (1973) converts the problem of developing lower bound of minimax risk to the calculations of divergence measures among parameters. This method has been proven to be successful in capturing the minimax convergence rate for various estimation problems including threshold estimation (Wang and Samworth, 2018). We briefly discuss Le Cam's method and apply it to the threshold regression to provide minimax lower bounds of $n^{2\alpha-1}$, $n^{-1/2}$, and $n^{-1/3}$ rate for diminishing jump threshold, continuous threshold, and unknown type threshold, respectively.

2 Le Cam's Method

2.1 Minimax Framework

We formulate the minimax framework before proceeding to Le Cam's method. Let $\mathcal{X}$ be a sample space and $\mathcal{X}^{1:n}$ be a n-cartesian product of $\mathcal{X}$. Denote a *model* as $\mathcal{P}$ which is a class of probability measures on $\mathcal{X}$. For any probability measure $\mathbb{P} \in \mathcal{P}$, we denote a product measure of $\mathbb{P}$ as $\mathbb{P}^{1:n}$ for any sample size n. We assume that samples are independently and identically distributed (*i.i.d.*). Let Θ be a parameter space equipped with a metric ρ. Finally, we define a function $\theta : \mathcal{P} \to \Theta$. The model $\mathcal{P}$ may or may not be indexed by Θ.

For an unknown distribution $\mathbb{P}$, n-samples are drawn in *i.i.d.* manner. We denote these samples as $X_1, X_2, \cdots, X_n$ and define $X^{1:n} := (X_1, X_2, \cdots, X_n)$ for convenience. An estimator $\hat{\theta}$ is a measurable function from $\mathcal{X}^n$ to Θ defined for each sample size n. The performance of an estimator is measured by its *risk* at $\mathbb{P}$:

$$\mathcal{R}_n(\hat{\theta}, \theta(\mathbb{P}); \rho) := \mathbb{E}_{\mathbb{P}^{1:n}}[\rho(\hat{\theta}(X^{1:n}), \theta(\mathbb{P}))],$$

where $\mathbb{E}_{\mathbb{P}}$ stands for the expectation with respect to $\mathbb{P}$. Minimax criterion evaluates the performance of estimators based on the adversarial choice of $\mathbb{P} \in \mathcal{P}$. An estimator is called a *minimax estimator* if it minimizes maximal risk among all estimators, and the minimax risk is defined to be the maximal risk of a minimax estimator:

$$\mathfrak{M}_n(\theta(\mathcal{P}); \rho) := \inf_{\hat{\theta}} \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}^{1:n}}[\rho(\hat{\theta}(X^n), \theta(\mathbb{P}))].$$

2.2 Le Cam's Method

We explain Le Cam's method which will be employed to derive a lower bound of the minimax risk for threshold estimation. This technique is sometimes called *reduction to (Bayesian) testing argument*. The idea of Le Cam (1973) is that if there is a sequence of indistinguishable parameter sets, no estimator can converge faster than the minimal distance among the elements of parameter sets. To derive the sharpest lower bound, one wants to maximize the minimal distance under the indistinguishability restriction.

Mathematically speaking, we approximate the parameter space Θ with a $2\delta_n$-separated set $\{\theta_1^n, \theta_2^n, \cdots, \theta_{M_n}^n\}$ for each sample size n, which means $\rho(\theta_i^n, \theta_j^n) \geq 2\delta_n$ for all $i \neq j$. Consider the statistical problem of guessing the true parameter among $\{\theta_1^n, \theta_2^n, \cdots, \theta_{M_n}^n\}$. Let J_n be a uniform random variable distributed on $\{1, 2, \cdots, M_n\}$. Then, n-samples, $X^{1:n}$, are generated from $\mathbb{P}_{\theta_j^n}$ such that $\theta_j^n = \theta(\mathbb{P}_{\theta_j^n})$ if J_n is realized to be j. Observing $X^{1:n}$, one guesses the true parameter from which samples are generated. To quantify the difficulty of this problem, we denote a joint probability measure of $(X^{1:n}, J_n)$ as $\mathbb{Q}_n$ and any guessing function from $\mathcal{X}^n \to \{1, 2, \cdots, M_n\}$ as ψ. The indistinguishability is measured by the minimal average error probability:

$$\inf_{\psi} \mathbb{Q}_n[\psi(X^{1:n}) \neq J_n].$$

If the average error probability does not tend to 0, that is, the parameter sets remain indistinguishable, the convergence rate of the minimal distance δ_n gives a lower bound of minimax convergence rate.

Proposition 1 (Le Cam's Bound)**.** *For any choice of $2\delta_n$-separated sets, the sequence of minimax risks is lower bounded as*

$$\mathfrak{M}_n(\theta(\mathcal{P}); \rho) \geq \delta_n \inf_{\psi} \mathbb{Q}_n[\psi(X^n) \neq J_n].$$

If $\inf_{\psi} \mathbb{Q}_n[\psi(X^n) \neq J_n] \geq c$ for all n, for some positive constant c,

$$\mathfrak{M}_n(\theta(\mathcal{P}); \rho) \geq c\delta_n.$$

In this paper, we always set c to be $1/4$. For the proof of the proposition, refer Wainwright (2019) where we borrowed the notations and the problem setup.

In many cases of parametric estimation problems, the binary approximation, $M_n = 2$ for all n, is enough to capture the minimax rate. The minimal average error probability of binary approximation is simply expressed as:

$$\inf_{\psi} \mathbb{Q}_n[\psi(X^n) \neq J_n] = \frac{1 - d_{TV}(\mathbb{P}^{1:n}_{\theta_0^n}, \mathbb{P}^{1:n}_{\theta_1^n})}{2},$$

where $d_{TV}(\mathbb{P}, \mathbb{Q})$ stands for the total variation between $\mathbb{P}$ and $\mathbb{Q}$. However, the total variation behaves badly as the sample size grows. So, we formulate Le Cam's bound with Kullback-Leibler (KL) divergence.

2.3 KL-version Le Cam's Method

Unlike total variation, KL-divergence has a nice decoupling property:

$$d_{KL}(\mathbb{P}^{1:n}, \mathbb{Q}^{1:n}) = n d_{KL}(\mathbb{P}, \mathbb{Q}),$$

where $d_{KL}(\mathbb{P}, \mathbb{Q})$ is KL-divergence from $\mathbb{P}$ to $\mathbb{Q}$. To relate the Le Cam's bound to KL-divergence, we need the following inequality:

Lemma 2 (Pinsker's Inequality)**.** *Let $\mathbb{P}, \mathbb{Q}$ be any probability measures. Then the total variation is upper bounded by KL-divergence as*

$$d_{TV}(\mathbb{P}, \mathbb{Q}) \leq \sqrt{\frac{1}{2} d_{KL}(\mathbb{P}, \mathbb{Q})}.$$

Refer Tsybakov (2009) for the proof. Observing Le Cam's bound, the total variation representation of error probability, and Pinsker's inequality, it is straightforward to show that the following proposition holds.

Proposition 3 (KL-version Le Cam's Bound)**.** *Assume that there are two sequences of parameters, $\{\theta_0^1,\ \theta_0^2,\ \theta_0^3,\ \cdots\}$ and $\{\theta_1^1,\ \theta_1^2,\ \theta_1^3,\ \cdots\}$ such that $\rho(\theta_0^n, \theta_1^n) \geq 2\delta_n$ and*

$$d_{KL}(\theta_0^n, \theta_1^n) \leq \frac{1}{2n}.$$

Then, the minimax risk is lower bounded as

$$\mathfrak{M}_n(\theta(\mathcal{P}); \rho) \geq \frac{1}{4}\delta_n.$$

Proof. Note that we slightly abused the notation of KL-divergence. Then,

$$\begin{aligned}
\mathfrak{M}_n(\theta(\mathcal{P});\rho) &\geq \delta_n \frac{1 - d_{TV}(\mathbb{P}^{1:n}_{\theta^n_0}, \mathbb{P}^{1:n}_{\theta^n_1})}{2} \\
&\geq \delta_n \frac{1 - \sqrt{\frac{1}{2} d_{KL}(\mathbb{P}^{1:n}_{\theta^n_0}, \mathbb{P}^{1:n}_{\theta^n_1})}}{2} \\
&= \delta_n \frac{1 - \sqrt{\frac{n}{2} d_{KL}(\mathbb{P}_{\theta^n_0}, \mathbb{P}_{\theta^n_1})}}{2} \geq \frac{1}{4}\delta_n,
\end{aligned}$$

where the first inequality is a consequence of Le Cam's bound and total variation representation of minimal average error probability, the second inequality is Pinsker's inequality, and the last inequality follows from our construction of parameter sequences. □

Remark. *Consider the case where the metric ρ is equivalent to square-root KL-divergence. For example, ℓ_2-prediction error in the regression model with Gaussian noise is equivalent to square-root KL-divergence as we will show in the next chapter. Note that the best lower bound that can be obtained by KL-version Le Cam's method is of $n^{-1/2}$ rate. However, it is known that the minimax rate for ℓ_2-prediction error of the regression function is $n^{-1/3}$ for the* 1*-dimensional, bounded Lipschitz continuous regression function, which means that binary approximation is not enough to capture the complexity of nonparameteric estimation problem.*

After Hasminskii (1978), Le Cam's bound with $M_n \geq 3$ was widely used to derive a lower bound of minimax risk for the nonparametric estimation problem. The minimal average error probability of multiple parameter sets can be lower bounded by Fano's inequality (Fano, 1961), and Birgé (1983, 1986) formalized the method to apply Fano's inequality to minimax lower bound which is called generalized Fano's method.

Yang and Barron (1999) proved that if the metric ρ is equivalent to *square-root KL-divergence, packing entropy and covering entropy is equivalent, and metric entropy $M(\varepsilon)$ satisfies the richness condition, that is,* $\liminf_{\varepsilon\to 0} M(\varepsilon/2)/M(\varepsilon) > 1$*, then the equation*

$$n\varepsilon_n^2 = M(\varepsilon_n)$$

determines the minimax rate, ε_n. Assuming that $M(\varepsilon) \to \infty$ as $\varepsilon \to 0$, we can see that $\varepsilon_n / n^{-\frac{1}{2}} \to \infty$, which means that the minimax rate is slower than the $n^{-1/2}$ rate derived by KL-version Le Cam's bound. Since the parameter space of threshold estimation is a closed interval in $\mathbb{R}^1$, the richness condition does not hold, which is one of the reasons we use KL-version Le Cam's bound instead of the approximation by multiple parameters.

3 Applications to Threshold Estimation

In this chapter, we derive lower bounds of minimax risks for threshold estimations. Consider the threshold linear regression problem:

$$y_i = x_i^T\beta + x_i^T\delta\mathbb{I}[q_i > \gamma] + u_i.$$

The threshold location is denoted as γ and we assume that γ lies in a closed interval $\Gamma := [\eta, 1-\eta]$ in $[0,1]$ for $0 < \eta < 1/3$. For convenience, we decompose the vector of regressors and the threshold effect δ such that $x_i = (1, x_{2i}, q_i)$, $\delta = (\delta_1, \delta_2^T, \delta_3)^T$, and $x_i^T\delta = \delta_1 + x_{2i}^T\delta_2 + \delta_3 q_i$. The model $\mathcal{P}$ is determined by the collection of (β, δ, γ). We assume that $\theta : \mathcal{P} \to \Gamma$ is well-defined so that $\theta(\mathbb{P}_{(\beta,\delta,\gamma)}) = \gamma$, and $\rho(\gamma_1, \gamma_2) = |\gamma_2 - \gamma_1|$.

Also, we assume that (y_i, x_i, u_i) are drawn in *i.i.d.* manner, x_i and u_i are independent, and u_i is a univariate normal random variable with variance

σ^2. Let $f(\cdot)$ denote the density of the threshold variable q_i.

To exclude the case where the threshold effect is meaningless, we assume that $\mathbb{E}[x_{2i}x_{2i}^T|q_i = \gamma]$ is continuous and positive in Γ. Denote $D(\gamma) = \mathbb{E}[x_i x_i^T|q_i = \gamma]$. For the continuity of $D(\gamma)$, we assume that $\mathbb{E}[x_{2i}|q_i = \gamma]$ is continuous in Γ. Finally, we assume that $f(\cdot)$ is continuous in Γ and $\inf_{\gamma\in\Gamma} f(\gamma) > 0$, and denote $\bar{f} = \sup_{\gamma\in\Gamma} f(\gamma)$.

Before proceeding, we show that KL-divergence can be simply expressed by ℓ_2-norm for the regression model:

Lemma 4. *Assume that $x \sim \mathbb{P}_X$ and $u \sim \mathcal{N}(0, \sigma^2)$. Let $y = f(x) + u$. Then, KL-divergence from the joint distribution determined by the regression function f_0 to f_1 is*

$$d_{KL}(f_0, f_1) = \frac{1}{2\sigma^2}\|f_1 - f_0\|^2_{\ell_2(\mathbb{P}_X)}.$$

Proof. Assume that $\mathbb{P}_X$ has a density $p(\cdot)$. The joint density of (y, x) is

$$\frac{1}{\sigma}\phi(\frac{y - f(x)}{\sigma})p(x),$$

where ϕ is a density function of standard normal random variable. Then the KL-divergence from f_0 to f_1 is

$$\begin{aligned}
&d_{KL}(f_0, f_1) = \iint \frac{1}{\sigma}\phi(\frac{y - f_0(x)}{\sigma})p(x)\log\frac{\phi((y - f_0(x))/\sigma)}{\phi((y - f_1(x))/\sigma)}dydx \\
&= \iint \frac{1}{\sigma}\phi(\frac{y - f_0(x)}{\sigma})p(x)\big(-\frac{(y - f_0(x))^2 - (y - f_1(x))^2}{2\sigma^2}\big)dydx \\
&= \frac{1}{2\sigma^2}\iint \frac{1}{\sigma}\phi(\frac{y - f_0(x)}{\sigma})\big(2(f_0(x) - f_1(x))y - f_0^2(x) + f_1^2(x)\big)dyp(x)dx \\
&= \frac{1}{2\sigma^2}\int (f_0(x) - f_1(x))^2 p(x)dx.
\end{aligned}$$

□

3.1 Jump Threshold

We mention that this section is influenced by the proof of Wang and Samworth (2018) for the constant regressor case. To extend this proof, we impose a restriction on the parameter space as follows:

Assumption 1. *Fix $\kappa > 0$. Let $\delta^T D(\gamma)\delta \geq \kappa$ for some $\gamma \in \Gamma$.*

To include the diminishing threshold model of Hansen (2000), we let $\kappa = cn^{-\alpha}$ for some positive real number c. It means that our parameter space grows as the sample size increases. Then the following proposition follows:

Proposition 5. *Assume that **Assumption 1.** is satisfied. If $\kappa = cn^{-\alpha}$ for some $0 \leq \alpha < 1/2$, the minimax risk is lower bounded as*

$$\mathfrak{M}_n \geq \frac{\sigma^2}{32\bar{f}c^2} n^{2\alpha-1},$$

for sufficiently large n.

From our lower bound, we can see that the minimax risk increases as the noise σ^2 increases, the jump size c decreases, or $\bar{f}$ decreases. Note that the small $\bar{f}$ generally implies that there are fewer samples such that $q_i \in \Gamma$, so it is reasonable that the minimax risk is inversely related to $\bar{f}$.

Proof. In fact, our proof is redundant if regressors contain the constant term. However, our proof can be applied even in the case where the constant term is dropped.

To use KL-version Le Cam's bound, we want to find two sequences of parameters $\theta_0^n := (\beta_0^n, \delta_0^n, \gamma_0^n)$, $\theta_1^n := (\beta_1^n, \delta_1^n, \gamma_1^n)$ such that $d_{KL}(\theta_0^n, \theta_1^n) \leq \frac{1}{2n}$. If $\beta_0^n \neq \beta_1^n$ or $\delta_0^n \neq \delta_1^n$, slope coefficients contain some information on γ in general. Therefore, we set $\beta_0^n = \beta_1^n$ and $\delta_0^n = \delta_1^n$. After rescailing, we can choose (δ_0, γ_0) such that $\left(\delta_0^T D(\gamma_0)\delta_0\right)^{1/2} = 3c/2$. From the continuity of $D(\gamma)$, there exists a positive real number ε such that

$$c \leq \left(\delta_0^T D(\gamma)\delta_0\right)^{1/2} \leq 2c,$$

for all $\gamma \in [\gamma_0 - \varepsilon, \gamma_0 + \varepsilon] \cap \Gamma$. Let $\delta_0^n = \delta_1^n = \delta_0 \cdot n^{-\alpha}$, and choose $\beta_0^n = \beta_1^n = \beta$ arbitrarily. Note that

$$\kappa = cn^{-\alpha} \leq \left((\delta_0 \cdot n^{-\alpha})^T D(\gamma)(\delta_0 \cdot n^{-\alpha})\right)^{1/2} \leq 2\kappa.$$

Fix $\gamma_0^n = \gamma_0$ arbitrarily. If we let $\gamma_1^n \to \gamma_0$, then $\gamma_1^n \in [\gamma_0 - \varepsilon, \gamma_0 + \varepsilon] \cap \Gamma$ for sufficiently large n, so our construction of parameter sequences is justified.

Let $\gamma_1^n = \frac{\sigma^2}{4\bar{f}c^2} n^{2\alpha-1} + \gamma_0$. Since $2\alpha - 1 < 0$, $\gamma_1^n \to \gamma_0$. In addition,

$$\begin{aligned} d_{KL}(\theta_0^n, \theta_1^n) &= \frac{1}{2\sigma^2} \int_{[\gamma_0, \gamma_1^n]} (\delta_0 \cdot n^{-\alpha})^T D(\gamma)(\delta_0 \cdot n^{-\alpha}) f(\gamma) d\gamma \\ &\leq \frac{1}{2\sigma^2} (4c^2 n^{-2\alpha} \bar{f})(\gamma_1^n - \gamma_0) = \frac{1}{2n}. \end{aligned}$$

Applying KL-version Le Cam's bound, we get the desired result. □

3.2 Kink Threshold

We construct a probability model of the kink threshold case:

Assumption 2. *Let $\|\delta\|_2 \geq \kappa > 0$, $\delta_2 = 0$. Assume that $\delta_1 + \delta_3\gamma = 0$ for some γ.*

Note that κ is fixed over n, that is, we do not consider the diminishing kink case. We develop a minimax lower bound for this model:

Proposition 6. *Assume that* ***Assumption 2.*** *is satisfied. The minimax* *risk is lower bounded as*

$$\mathfrak{M}_n \geq \frac{\sigma}{\sqrt{128\eta \bar{f}\kappa}} n^{-1/2},$$

for sufficiently large n.

Relationships between the minimax risk and the nuisance parameters $\bar{f}, \sigma^2, \kappa$ are the same as in the jump threshold case. In addition, note that large η implies small parameter space Γ, so the estimation problem is easier for bigger η.

Proof. We start by constructing two sequences of parameters whose KL-divergence is smaller than $\frac{1}{2n}$ and $\rho(\gamma_0^n, \gamma_1^n)$ is as large as possible.

Let two parameter sequences be $\theta_0^n := (\beta, (-\delta_{30}\gamma_0^n, 0, \cdots, 0, \delta_{30}), \gamma_0^n)$, $\theta_1^n := (\beta, (-\delta_{31}\gamma_1^n, 0, \cdots, 0, \delta_{31}), \gamma_1^n)$. Let $\gamma_1^n < \gamma_0^n$. If we set $\delta_{30} = \delta_{31} = \kappa$, then $\|\delta_0^n\|_2, \|\delta_1^n\|_2 \geq \kappa$, hence the assumption 2 is satisfied. Set $\gamma_0^n = 1 - \eta$, $\gamma_1^n = \gamma_0^n - \left(\frac{\sigma^2}{2n\eta \bar{f}\kappa^2}\right)^{1/2}$. Observe that $\gamma_1^n \geq 1 - 2\eta$ for sufficiently large n.

Then, KL-divergence from θ_0^n to θ_1^n is

$$\begin{aligned} d_{KL}(\theta_0^n, \theta_1^n) &\leq \frac{1}{2\sigma^2} \int_{[\gamma_1^n, 1]} (\kappa(\gamma_0^n - \gamma_1^n))^2 f(\gamma) d\gamma \\ &\leq \frac{2\eta \bar{f}}{2\sigma^2} \kappa^2 (\gamma_0^n - \gamma_1^n)^2 = \frac{1}{2n}. \end{aligned}$$

Therefore, we can apply the KL-version Le Cam's bound. □

3.3 Unknown Threshold

In this section, we consider the case where we know there is a threshold but do not know whether it is a jump type threshold or kink type threshold. For this case, it is implausible to assume that the jump size is larger than a positive constant when the threshold is jump type. Instead, we impose a restriction on the size of threshold effect as we did in the kink threshold model.

Assumption 3. *Let $\|\delta\|_2 \geq \kappa > 0$. Assume that $\delta^T D(\gamma)\delta > 0$ or $\delta_1 + \delta_3\gamma = 0$; $\delta_2 = 0$ for some $\gamma \in \Gamma$.*

Note that we are assuming the fixed threshold case, that is, threshold effect δ cannot diminish because of the condition $\|\delta\|_2 \geq \kappa > 0$. For the fixed jump threshold, we derived n^{-1} lower bound, and for the fixed kink threshold, we derived a lower bound of $n^{-1/2}$ rate. We develop a slower $n^{-1/3}$ rate lower bound of the minimax risk for the unknown threshold model.

Proposition 7. *Assume that **Assumption 3.** is satisfied. The minimax risk is lower bounded as*

$$\mathfrak{M}_n \geq \frac{\sigma^{2/3}}{6\bar{f}^{1/3}\kappa^{2/3}} n^{-1/3}$$

Proof. To capture the complexity of the enlarged parameter space, we pick one parameter sequence to represent the kink threshold and the other one to represent the jump threshold. $\theta_0^n := (\beta, (-\delta_{30}\gamma_0^n, 0, \cdots, 0, \delta_{30}), \gamma_0^n)$, $\theta_1^n := (\beta, (-\delta_{31}\gamma_0^n, 0, \cdots, 0, \delta_{31}), \gamma_1^n)$. Let $\gamma_0^n < \gamma_1^n$. Set $\gamma_0^n = \eta$, $\delta_{30} = \delta_{31} = \kappa$. Then θ_0^n has a kink threshold, θ_1^n has a jump threshold, and θ_0^n, θ_1^n satisfy the assumption 3. Choose $\gamma_1^n = \eta + \left(\frac{3\sigma^2}{n\bar{f}\kappa^2}\right)^{1/3}$. KL-divergence from θ_0^n to θ_1^n is computed as

$$\begin{aligned} d_{KL}(\theta_0^n, \theta_1^n) &= \frac{1}{2\sigma^2}\int_{[\eta,\gamma_1^n]} (\kappa(\gamma-\eta))^2 \cdot f(\gamma)d\gamma \\ &\leq \frac{\bar{f}\kappa^2}{6\sigma^2}(\gamma_1^n - \eta)^3 = \frac{1}{2n}. \end{aligned}$$

Therefore, we can apply KL-version Le Cam's bound. Since $\frac{3^{1/3}}{8} \geq \frac{1}{6}$, we get the desired result. $\square$

4 Conclusion

At the early stage of this paper, we observed that the approximation by multiple sets can improve the KL-version Le Cam's bound if the parameter space is rich enough. Based on this observation, we used binary approximations of the parameter space to derive minimax lower bounds of various threshold models. To summarize, we proved that the minimax convergence rate is lower bounded by n^{-1} rate for the fixed jump threshold, $n^{-1/2}$ rate for the fixed kink threshold, and $n^{-1/3}$ rate for the unknown threshold type. Further studies on the upper bound of minimax risk will complete the proof for the minimax rate.

References

Birgé, L. (1983). Approximation dans les espaces metriques et theorie de léstimation. *Z. Wahrsch. Verw. Gebiete, 65*, 181-237.

Birgé, L. (1986). On estimating a density using Hellinger distance and some other strange facts. *Probability Theory Related Fields, 97*, 271-291.

Chan, K. S. (1993). Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model. *The Annals of Statistics, 21,* 520-533.

Duchi, J., Jordan, M., and Wainwright, M. (2018). Minimax optimal procedures for locally private estimation. *Journal of the American Statistical Association, 113*, 182-201.

Fano, R. M. (1961). *Transmission of Information : A Statistical Theory of communication.* MIT Press.

Hájek, J. (1972). Local asymptotic minimax and admissibility in estimation. *Proc. 6. Berkeley Symp. Math. Statistics and Probability, Univ. Calif. Press,* 172–194.

Hansen, B. E. (2000). Sample splitting and threshold estimation. *Econometrica, 68*, 575-603.

Hansen, B. E. (2017). Regression kink with an unknown threshold. *Journal of Business and Economic Statistics, 35*, 228-240.

Hasminskii, R. Z. (1978). A lower bound on the risks of nonparametric estimates of densities in the uniform metric. *Theory of Probability and Its Applications, 23*, 794-798.

Hidalgo, J., Lee, J., and Seo, M. (2019). Robust inference for threshold regression models. *Journal of Econometrics, 210(2)*, 291-309.

Le Cam, L. (1973). Convergence of estimates under dimensionality restrictions. *Annals of Statistics, 1(1)*, 38-53.

Tsybakov, A. B. (2009). *Introduction to Nonparametric Estimation.* Springer.

Wainwright, M. J. (2019). *High-dimensional statistics : A non-asymptotic view point.* Cambridge University Press.

Wald, A. (1939). Contributions to the theory of statistical estimation and testing hypotheses. *Annals of Mathematical Statistics, 10(4)*, 299–326.

Wang, T. and Samworth, R. (2018). High-dimensional changepoint estimation via sparse projection. *Journal of Royal Statistical Methodology : Series B, 80(1)*, 57-83.

Yang, Y. and Barron, A. (1999). Information-theoretic determination of minimax rates of convergence. *Annals of Statistics, 27(5)*, 1564-1599.

Yu, P. (2012). Likelihood estimation and inference in threshold regression. *Journal of Econometrics, 167*, 274-294.

Yu, P. (2015). Adaptive estimation of the threshold point in threshold regression. *Journal of Econometrics, 189(1)*, 83-100.

5 국문 초록

본 논문은 임계점이 있는선형 회귀 모형을 분석한다. 대표적인 임계점 회귀분석 모형은 다음과 같이 기술된다:

$$y_i = x_i^T \beta + x_i^T \delta \mathbb{I}[q_i > \gamma] + u_i.$$

임계점의 위치, γ를 추정함에 있어서의 근본적인 한계를 분석하는 것은 기존에 존재하는 추정 방법을 평가함에 보완적인 견해를 줄 수 있다. 본 논문은 임계점 추정에서 최대최소 수렴속도의 하한을 계산하여 임계점 모형에 관한 연구에 기여한다. 특히 기존에 존재하는 추정 방법들의 수렴 속도가 임계점의 성질에 영향을 받는다는 것에 주목하여 임계점의 성질에 따른 세 가지 모형을 제시하고 각각의 경우에 대해 최대최소 수렴속도의 하한을 구한다.

본 논문의 증명은 Le Cam (1973) 이 제시한 방법론에 기초한다. 따라서, 임계점 모형의 분석에 들어가기 앞서 Le Cam의 방법론을 적용하는 방법과 그 한계에 관하여 간략하게 논의한다.

핵심어 : 최대최소 추정, 임계점 모형

학번 : 2017 - 26997