A note on the L2-convergence rates of derivatives

Johan Sebastian Ohlendorff
johan.ohlendorff@sund.ku.dk
University of Copenhagen

Abstract

In this brief note, …

1 Main section

Let the $L_2(\nu)$-norm of a function $f \in L_2(\nu)$ be defined as

$$\|f\|_\nu = \sqrt{\int f^2 d\nu}. \tag{1}$$

Consider a sequence of estimators $\hat{P}_n(t \mid x)$ of $P(t \mid x)$ which are defined on $[0, \tau]$. We assume that $\hat{P}_n(0 \mid x) = P(0 \mid x) = 0$. We let μ_0 denote an appropriate measure for the covariates x. These are assumed to have the $L_2(\mu_0)$-convergence rate $n^{-\gamma-\varepsilon}$ for Lebesgue almost all $t \in [0, \tau]$ for $\varepsilon > 3\gamma$ and $\gamma > 0$. We are interested in constructing an estimator $p(t \mid x) = P'(t \mid x)$ of the derivative of $P(t \mid x)$, which also has the $L_2(\mu_0 \otimes m)$-convergence rate $n^{-\gamma}$, where m is the Lebesgue measure on $[0, \tau]$. The precise statement is given in Theorem 1.1.

Let us look at parametric models as an example. If $\left\|\hat{P}_n(t \mid x) - P(t \mid x)\right\|_{\mu_0} = o_P\left(n^{-\frac{1}{2}}\right)$ and $\|R_n\|_{\mu_0 \otimes m} \leq K \|R_{1,n}\|_{\mu_0 \otimes m} \|R_{2,n}\|_{\mu_0 \otimes m}$, where $\|R_{1,n}\|_{\mu_0 \otimes m} = \|\hat{p}_n - p\|_{\mu_0 \otimes m}$, then $\|R_n\|_{\mu_0 \otimes m} = o_P\left(n^{-\frac{1}{2}}\right)$ if $\|R_{2,n}\|_{\mu_0 \otimes m} = o_P\left(n^{-\frac{5}{12}-\varepsilon}\right)$ for some small $\varepsilon > 0$ (a little bit slower than parametric rate).

This is useful if one wishes to obtain convergence rates for a hazard function which one has not explicitly considered such as in a Cox regression. As an example for the Cox:

$$\begin{aligned} \sqrt{\int \left(\hat{\Lambda}(t \mid x) - \Lambda(t \mid x)\right)^2 \mu_0(x)} &\leq \sqrt{\int \left(\left(\hat{\Lambda}_0(t \mid x) - \Lambda_0(t)\right) \exp\left(\hat{\beta}_n x\right)\right)^2 \mu_0(dx)} \\ &+ \sqrt{\Lambda_0^2(t) \int \left(\exp\left(\hat{\beta}_n x\right) - \exp(\beta x)\right)^2 \mu_0(dx)} \end{aligned} \tag{2}$$

Under standard regularity conditions, the last term is $O_P\left(n^{-\frac{1}{2}}\right)$ (parametric rate) and the first term is $O_P\left(n^{-\frac{1}{2}}\right)$ (parametric rate). The first may be shown to be $O_P\left(n^{-\frac{1}{2}}\right)$ using empirical process theory (note the theorem also works with bounded in probability).

Theorem 1.1: Let $\hat{P}_n(t \mid x)$ be a sequence of estimators of $P(t \mid x)$ defined on $[0, \tau]$ fulfilling that $\hat{P}_n(0 \mid x) = P(0 \mid x) = 0$. Suppose that $P(t \mid x) \in C^2([0, \tau])$ for μ_0-almost all x and that there exists a constant $K > 0$ such that $p'(t \mid x) \leq K$ for μ_0-almost all x and $t \in [0, \tau]$. If $\left\|\hat{P}_n(t \mid \cdot) - P(t \mid \cdot)\right\|_{\mu_0} = o_P(n^{-\gamma-\varepsilon})$ for Lebesgue almost all $t \in [0, \tau]$ for $\varepsilon > 3\gamma$, then there exists an estimator $\hat{p}_n(t \mid x)$ of $p(t \mid x) = P'(t \mid x)$ such that

$$\|\hat{p}_n - p\|_{\mu_0 \otimes m} = o_P(n^{-\gamma}). \tag{3}$$

The estimator $\hat{p}_n(t \mid x)$ fulfills on a grid $0 = t_1 < ... < t_{K_n} = \tau$ with some mesh $b(n) = \max_{1 \leq k \leq K_n}(t_k - t_{k-1}) \to 0$ as $n \to \infty$ and $K_n \to \infty$ as $n \to \infty$ determined by ε such

$$\int_0^{t_k} \hat{p}_n(s \mid x) ds = \hat{P}_n(t_k \mid x). \tag{4}$$

Proof: Consider a partition $0 = t_1 < ... < t_{K_n} = t$ of $[0, t]$ with mesh $b(K_n) = \max_{1 \leq k \leq K_n}(t_k - t_{k-1})$. Let $K_n = \lfloor n^z \rceil$ for some $2\gamma < z < \frac{2}{3}\varepsilon$ and $b(K_n) = \tau \lfloor n^{-z} \rceil$. Then $K_n \to \infty$ as $n \to \infty$ and $b(K_n) \to 0$ as $n \to \infty$. We will show the theorem by constructing an explicit estimator $\hat{p}_n(t \mid x)$ by approximating the derivative via a secant. Let

$$\hat{p}_n(t \mid x) = \sum_{k=1}^{K_n} \mathbb{1}\{t \in (t_k, t_{k+1}]\} \frac{\hat{P}_n(t_{k+1} \mid x) - \hat{P}_n(t_k \mid x)}{t_{k+1} - t_k} \tag{5}$$

Then evidently, we have

$$\int_0^{t_k} \hat{p}_n(s \mid x) ds = \sum_{j=1}^{k-1} \frac{\hat{P}_n(t_{k+1} \mid x) - \hat{P}_n(t_k \mid x)}{t_{k+1} - t_k}(t_{k+1} - t_k) = \hat{P}_n(t_k \mid x). \tag{6}$$

Furthermore, let

$$\tilde{p}_n(t \mid x) = \sum_{k=1}^{K_n} \mathbb{1}\{t \in (t_k, t_{k+1}]\} \frac{P(t_{k+1} \mid x) - P(t_k \mid x)}{t_{k+1} - t_k}. \tag{7}$$

By the triangle inequality, we have

$$\|\hat{p}_n - p\|_{\mu_0 \otimes m} \leq \|\hat{p}_n - \tilde{p}_n\|_{\mu_0 \otimes m} + \|\tilde{p}_n - p\|_{\mu_0 \otimes m}. \tag{8}$$

We start with the first term on the right-hand side.

$$\begin{aligned}
\|\hat{p}_n - \tilde{p}_n\|_{\mu_0 \otimes m} &= \left\| \sum_{k=1}^{K_n} \mathbb{1}\{\cdot \in (t_k, t_{k+1}]\} \frac{\left(\hat{P}_n(t_{k+1} \mid \cdot) - P(t_{k+1} \mid \cdot)\right) - \left(\hat{P}_n(t_k \mid \cdot) - P(t_k \mid \cdot)\right)}{t_{k+1} - t_k} \right\|_{\mu_0 \otimes m} \\
&\leq \sum_{k=1}^{K_n} \left\| \mathbb{1}\{\cdot \in (t_k, t_{k+1}]\} \frac{\hat{P}_n(t_{k+1} \mid \cdot) - P(t_{k+1} \mid \cdot) - \left(\hat{P}_n(t_k \mid \cdot) - P(t_k \mid \cdot)\right)}{t_{k+1} - t_k} \right\|_{\mu_0 \otimes m} \\
&\leq \sum_{k=1}^{K_n} \frac{1}{\sqrt{t_{k+1} - t_k}} \left(\left\|\hat{P}_n(t_{k+1} \mid \cdot) - P(t_{k+1} \mid \cdot)\right\|_{\mu_0} + \left\|\hat{P}_n(t_k \mid \cdot) - P(t_k \mid \cdot)\right\|_{\mu_0} \right) \\
&= o\left(n^{z - \frac{1}{2}(-z)}\right) o_P(n^{-\gamma-\varepsilon}) = o_P\left(n^{\frac{3}{2}z - \gamma - \varepsilon}\right) = o_P(n^{-\gamma}).
\end{aligned} \tag{9}$$

There exists by the mean value theorem a $\xi_{k,x} \in (t_k, t_{k+1})$ such that $\frac{P(t_{k+1} \mid x) - P(t_k \mid x)}{t_{k+1} - t_k} = p(\xi_{k,x} \mid x)$ for μ_0-almost all x. Furthermore, there exists also a $\xi'_{k,t,x}$ between t and $\xi_{k,x}$ such that $p(t \mid x) - p(\xi_{k,x} \mid x) = (t - \xi_{k,x}) p'(\xi'_{k,t,x} \mid x)$. This implies that we can bound the second term on the right-hand side as

$$
\begin{aligned}
\|\tilde{p}_n - p\|_{\mu_0 \otimes m} &= \left\| \sum_{k=1}^{K_n} \mathbb{1}\{\cdot \in (t_k, t_{k+1}]\} p\big(\xi_{k,\cdot} \mid \cdot\big) - p(\cdot \mid \cdot) \right\|_{\mu_0 \otimes m} \\
&= \left\| \sum_{k=1}^{K_n} \mathbb{1}\{\cdot \in (t_k, t_{k+1}]\} \big(p\big(\xi_{k,\cdot} \mid \cdot\big) - p(\cdot \mid \cdot)\big) \right\|_{\mu_0 \otimes m} \\
&= \left\| \sum_{k=1}^{K_n} \mathbb{1}\{\cdot \in (t_k, t_{k+1}]\} \big(\cdot - \xi_{k,\cdot}\big) p'\big(\xi'_{k,\cdot,\cdot} \mid \cdot\big) \right\|_{\mu_0 \otimes m} \qquad (10) \\
&\leq K \sum_{k=1}^{K_n} (t_{k+1} - t_k) \sqrt{t_{k+1} - t_k} \\
&= K \sum_{k=1}^{K_n} (b(k))^{\frac{3}{2}} = o\Big(n^{z + \frac{3}{2}(-z)}\Big) = o\Big(n^{-\frac{1}{2}z}\Big) = o(n^{-\gamma}).
\end{aligned}
$$

so that we have

$$
\|\hat{p}_n - p\|_{\mu_0 \otimes m} = o_P(n^{-\gamma}). \qquad (11)
$$

$\square$