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 $n^{-\frac{1}{4}-\varepsilon}$ for Lebesgue almost all $t \in [0, \tau]$ for $\varepsilon > \frac{1}{12}$. This corresponds to a convergence rate of slightly better than $n^{-\frac{1}{3}}$. 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 $n^{-\frac{1}{4}}$, where m is the Lebesgue measure on $[0, \tau]$. The precise statement is given in Theorem 1.1. This is useful if one wishes to obtain convergence rates for a hazard function which one has not explicitly considered, but only the cumulative hazard function, such as in a Cox regression. For example for the Cox:

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

Under standard regularity conditions, the last term is $O_P\left(n^{-\frac{1}{2}}\right)$ (parametric rate) and the first term is $O_P\left(n^{-\frac{1}{2}}\right)$ (nonparametric rate) if we can show that the Breslow estimator is bounded so that dominated convergence can be applied. If we assume that the covariates are bounded $\exp\left(\hat{\beta}_n x\right)$ can be moved outside.

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 $\left\|\hat{P}_n(t \mid \cdot) - P(t \mid \cdot)\right\|_{\mu_0} = o_P\left(n^{-\frac{1}{4}-\varepsilon}\right)$ for Lebesgue almost all $t \in [0, \tau]$ for $\varepsilon > \frac{1}{12}$, then there exists an estimator $\hat{p}_n(t \mid x)$ of $p(t \mid x) = P'(t \mid x)$ such that

$$\|\hat{p}_n - p\|_{\mu_0 \otimes m} = o_P\left(n^{-\frac{1}{4}}\right). \tag{3}$$

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 $b(K_n) = \max_{1 \leq k \leq K_n}(t_k - t_{k-1})$. Choose $x_1 > 0$ and $x_2 < 0$ such that $0 < x_1 < \frac{3}{4}\varepsilon - \frac{1}{16}$ and $2(x_1 - \varepsilon) < x_2 < -\frac{2}{3}\varepsilon - \frac{1}{6} < 0$. Here, we will let $K_n = \lfloor n^{x_1} \rceil)$ and $b(K_n) = \lfloor n^{x_2} \rceil$. Then $K_n \to \infty$ as $n \to \infty$ because $\varepsilon > \frac{1}{12}$ and $b(K_n) \to 0$ as $n \to \infty$ because $\varepsilon > 0$. 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.

$$\begin{aligned}
\|\hat{p}_n - \tilde{p}_n\|_{\mu_0 \otimes m} &= \left\| \sum_{k=1}^{K_n} \mathbb{1}\{\cdot \in (t_k, t_{k+1}]\} \frac{\left(\hat{P}_n(t_{k+1} \mid \cdot) - P(t_{k+1} \mid \cdot)\right) - \left(\hat{P}_n(t_k \mid \cdot) - P(t_k \mid \cdot)\right)}{t_{k+1} - t_k} \right\|_{\mu_0 \otimes m} \\
&\leq \sum_{k=1}^{K_n} \left\| \mathbb{1}\{\cdot \in (t_k, t_{k+1}]\} \frac{\hat{P}_n(t_{k+1} \mid \cdot) - P(t_{k+1} \mid \cdot) - \left(\hat{P}_n(t_k \mid \cdot) - P(t_k \mid \cdot)\right)}{t_{k+1} - t_k} \right\|_{\mu_0 \otimes m} \qquad (9) \\
&\leq \sum_{k=1}^{K_n} \frac{1}{\sqrt{t_{k+1} - t_k}} \left(\left\|\hat{P}_n(t_{k+1} \mid \cdot) - P(t_{k+1} \mid \cdot)\right\|_{\mu_0} + \left\|\hat{P}_n(t_k \mid \cdot) - P(t_k \mid \cdot)\right\|_{\mu_0} \right) \\
&= o\left(n^{x_1 - \frac{1}{2}x_2}\right) o_P\left(n^{-\frac{1}{4}-\varepsilon}\right) = o_P\left(n^{\varepsilon - \frac{1}{4} - \varepsilon}\right) = o_P\left(n^{-\frac{1}{4}}\right).
\end{aligned}$$

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 $p(t \mid x) - p\big(\xi_{k,x} \mid x\big) = \big(t - \xi_{k,x}\big)p'\big(\xi'_{k,t,x} \mid x\big)$. This implies that we can bound the second term on the right-hand side as

$$\begin{aligned}
\|\tilde{p}_n - p\|_{\mu_0\otimes m} &= \left\|\sum_{k=1}^{K_n} \mathbb{1}\{\cdot \in (t_k, t_{k+1}]\} p\big(\xi_{k,\cdot} \mid \cdot\big) - p(\cdot \mid \cdot)\right\|_{\mu_0\otimes m} \\
&= \left\|\sum_{k=1}^{K_n} \mathbb{1}\{\cdot \in (t_k, t_{k+1}]\}\big(p\big(\xi_{k,\cdot} \mid \cdot\big) - p(\cdot \mid \cdot)\big)\right\|_{\mu_0\otimes m} \\
&= \left\|\sum_{k=1}^{K_n} \mathbb{1}\{\cdot \in (t_k, t_{k+1}]\}\big(\cdot - \xi_{k,\cdot}\big)p'\big(\xi'_{k,\cdot,\cdot} \mid \cdot\big)\right\|_{\mu_0\otimes m} \\
&\leq K \sum_{k=1}^{K_n} \big(t_{k+1} - t_k\big)\sqrt{t_{k+1} - t_k} \\
&= K \sum_{k=1}^{K_n} (b(k))^{\frac{3}{2}} = o\Big(n^{x_1 + \frac{3}{2}x_2}\Big) = o\Big(n^{-\frac{1}{4}}\Big).
\end{aligned} \tag{10}$$

so that we have

$$\|\hat{p}_n - p\|_{\mu_0\otimes m} = o_P\Big(n^{-\frac{1}{4}}\Big). \tag{11}$$

□