
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

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

Consider a sequence of estimators $\hat{P}_n(t | x)$ of $P(t | x)$ which are defined on $[0, \tau]$. We assume that $\hat{P}_n(0 | x) = P(0 | x) = 0$. 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 $p(t | x) = P'(t | x)$ of the derivative of $P(t | x)$, 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 $\|\hat{P}_n(t | x) - P(t | x)\|_{\mu_0} = o_P(n^{-\frac{1}{2}})$ 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 $\|R_n\|_{\mu_0 \otimes m} = o_P(n^{-\frac{1}{2}})$ if $\|R_{2,n}\|_{\mu_0 \otimes m} = o_P(n^{-\frac{3}{8}-\varepsilon})$ for some small $\varepsilon > 0$ (a little bit slower than parametric rate) ($\gamma + 3\gamma = \frac{1}{2}, \gamma + z = \frac{1}{2} \Rightarrow z = \frac{3}{8}$), so it may still be possible to use flexible machine learning.

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:

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

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

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

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

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

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

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

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

Then evidently, we have

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

Furthermore, let

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

By the triangle inequality, we have

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

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

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

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

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

so that we have

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

□