Supplementary material of the paper

Plug-in Estimators for Conditional Expectations and Probabilities

A Families of Sets and Functions

Lemma A.1 (Proof of Proposition 2.1). \( C \) is a universal Donsker class if \( \mathbb{X} \) is \( \mathbb{R}^d \) or \([0, 1]^d\).

Proof. We apply Corollary 6.20 of Dudley (2014).

(i) The set \( C \) has finite VC dimension bounded by \( 2d + 1 \). Consider arbitrary \( 2d + 1 \) points \( x_1, x_2, \ldots \in \mathbb{X} \). Now, there is at least one element \( x_i \) which attains the maximum in dimension \( j \), \( 1 \leq j \leq d \), i.e. \( x_{ij} = \max_{j \leq d} x_{ij} \). Select for every dimension such an element and, in the same way, select \( d \) minimizers. Denote the joint set of these points with \( B \). Then every element from \( C \) that contains \( B \) also contains \( x_1, \ldots, x_{2d+1} \) and there is no set \( A \in C \) which fulfills \( A \cap \{ x_i : 1 \leq i \leq 2d + 1 \} = B \). \( \Box \)

(ii) The set \( \mathbb{X} \) is a Borel set. Hence \( \chi \mathbb{X} \in L^2(\mathbb{X}, P) \) for every probability measure \( P \) on the Borel sets and \( \chi \mathbb{X} \) is an envelope function of \( C \).

(iii) \( C \) is image admissible Suslin. Consider \( \mathbb{Y} = \mathbb{X} \times \{ 0, \infty \} \) equipped with the natural topology, which is the one induced by the Euclidean metric, and the corresponding Borel \( \sigma \)-algebra. \( \mathbb{Y} \) is a Polish space since it is a closed subset of the complete space \( \mathbb{R}^{d+1} \).

Consider now the map \( T(y, h) = \chi([y, y + h1]) \) if \([y, y + h1] \in C \) and \( \chi(0) \) otherwise. \( T \) maps \( \mathbb{Y} \) onto \( C \). We need to verify that \( (y, h, x) \mapsto (T(y, h))(x) \) is jointly measurable, that is for any Borel subset \( A \) of \( \mathbb{R}^d \)

\[
B = \{(y, h, x) : x, y \in X, h \in [0, \infty], (T(y, h))(x) \in A \}
\]

must be in the product \( \sigma \)-algebra. \( T(y, h)(x) \) attains either the value 0 or 1 hence there are four events we need to consider. First \( A = \emptyset \) which implies \( B = \emptyset \) and \([0, 1] \subseteq A \) which implies \( B = \mathbb{Y} \) are always in the \( \sigma \)-algebra.

So consider now a set \( A \) such that \( 1 \in A \) but not 0, then

\[
B = \{(y, h, x) : x, y \in X, h \in [0, \infty], \chi([y, y + h1])(x) = 1 \}
\]

\[
= \{(y, h, x) : x, y \in X, h \in [0, \infty], x \in [y, y + h1] \}.
\]

This set is closed in the natural topology of \( \mathbb{X} \times [0, \infty] \times [0, \infty] \) since, if \( (y, h, x) \) is not in \( B \) then with \( \epsilon < d \langle x, [y, y + h1] \rangle \), where \( d \) is the Euclidean metric, we have that the open ball

\[
\{ (u, \eta, \xi) : d(y, u), d(h, \eta), d(x, \xi) < \epsilon / 3, u, \eta, \xi \in \mathbb{X}, \eta \in [0, \infty] \} \subseteq \mathbb{Y} \times \mathbb{X} \backslash B
\]

contains \( (y, h, x) \). And, since,

\[
(\mathbb{Y} \times \mathbb{X}) \backslash B = \{(y, h, x) : x, y \in X, h \in [0, \infty], (T(y, h))(x) = 0 \}
\]

the latter set is open. Hence, both sets are in the Borel algebra \( \mathcal{B}(\mathbb{Y} \times \mathbb{X}) \).

It remains to show that the product algebra equals the Borel algebra, that is,

\[
\mathcal{B}(\mathbb{Y}) \otimes \mathcal{B}(\mathbb{X}) = \mathcal{B}(\mathbb{Y} \times \mathbb{X}).
\]

This follows from Fremlin (2003)[4A3D(ci)] if \( \mathbb{Y} \times \mathbb{X} \) is a hereditary Lindelöf space. Though, every second countable space like \( \mathbb{Y} \times \mathbb{X} \) is hereditary Lindelöf and the result follows. \( \Box \)

Lemma A.2 (Proof of Lemma 2.3). Let \( \mathbb{X} \) be any set and \( \mathcal{G} \) be a \( \sigma \)-algebra of subsets of \( \mathbb{X} \). \( \mathcal{G} \) is a VC-class if, and only if, \( \mathcal{G} \) is a finite family of sets.

Proof. Any finite collection of sets is a VC-class. For the other direction assume \( \mathcal{G} \) is infinite.

(i) There exists a countably infinite sequence of disjoint sets in \( \mathcal{G} \). We prove by induction that for any \( n \in \mathbb{N} \) there exist \( n \) disjoint sets in \( \mathcal{G} \). The induction hypothesis is trivially fulfilled for \( n = 1 \). For the induction step let us assume that
$A_1, \ldots, A_n \in \mathcal{G}$ are not empty and mutually disjoint. There exists an element $B \in \mathcal{G}$ that is not contained in $\sigma(A_1, \ldots, A_n)$ since otherwise $\mathcal{G} = \sigma(A_1, \ldots, A_n)$ and $\mathcal{G}$ would be finite. Take such an element $B$. If $B \setminus \bigcup_{i \leq n} A_i \in \mathcal{G}$ is not empty then add this to the sequence as $A_{n+1}$. $A_{n+1}$ is then obviously disjoint from all $A_1, \ldots, A_n$. If $B \setminus \bigcup_{i \leq n} A_i = \emptyset$ then $B \subseteq \bigcup_{i \leq n} A_i$. Furthermore, there is some $i \leq n$ such that $A_i \setminus B \neq \emptyset \neq A_i \cap B$, because otherwise $B$ would be a union of a subset of $A_1, \ldots, A_n$. Now, remove $A_i$ from the sequence and add $A_i \setminus B$ and $A_i \cap B$ to the sequence. This way we gain $n+1$ disjoint elements that are all contained in $\mathcal{G}$. This implies now directly that there is a countably infinite sequence of disjoint sets contained in $\mathcal{G}$. Q.E.D.

(ii) By (i) we can choose a sequence $\{A_n\}_{n \in \mathbb{N}}$ of disjoint and non-empty subsets of $\mathcal{G}$. By countable choice we can select a sequence $\{x_n\}_{n \in \mathbb{N}}$ such that $x_n \in A_n$. Consider any $k \in \mathbb{N}$, points $x_1, \ldots, x_k$ and any subset of these, say $\{x_{n_i} : i \leq l, 0 \leq n_1 < n_2 < \ldots < n_l \leq k\}$ for $l \leq k$ and consider the corresponding sequence of sets $\{A_{n_i} : i \leq l\}$ then $\{x_{n_i} : i \leq l\} = \{x_{n_i} \cap A_{n_i} : i \leq l\}$ and the set $\{x_i\}_{i \leq k}$ is shattered. Since this argument applies to any $k \in \mathbb{N}$ we know that $\mathcal{G}$ is not a VC-class.

Lemma A.3 (Proof of Lemma 5). Let $(\Omega, \mathcal{A}, P)$ be a probability space and $C \subset \mathcal{A}$ a disjoint family of sets such that for each $A \in C$ there exists $\{A_n\}_{n \in \mathbb{N}}$ in $C$ with $\Omega \setminus A = \bigcup_{n \in \mathbb{N}} A_n$ and $\emptyset \in C$. For any measure $Q$ for which there exists a constant $c > 0$ such that for all $A \in C$, $|Q(A) - P(A)| \leq cP(A)$, we have that $\sup_{A \in \sigma(C)} \frac{|Q(A) - P(A)|}{P(A)} \leq c$.

Proof. We apply the monotone class theorem. $A, B \in \mathcal{C}$ then either $A \cap B = \emptyset \in \mathcal{C}$ or $A = B \in \mathcal{C}$. Define

\[ \mathcal{D} := \{ A : A = \bigcup_{n \in \mathbb{N}} \mathcal{E}, \mathcal{E} \subseteq C \text{ a countable family, } |Q(A) - P(A)| \leq cP(A) \} . \]

$\mathcal{D}$ is a Dynkin class: (1) $\emptyset \in \mathcal{D}$; (2) $A \in \mathcal{D}$ then by assumption $\Omega \setminus A = \bigcup_{n \in \mathbb{N}} A_n$ for some elements $A_n \in \mathcal{C}$ and because the $A_n$ are disjoint we have

\[ |Q(\Omega \setminus A) - P(\Omega \setminus A)| \leq \sum_{n \in \mathbb{N}} |Q(A_n) - P(A_n)| \leq c \sum_{n \in \mathbb{N}} P(A_n) = cP(\Omega \setminus A) \]

and $\Omega \setminus A \in \mathcal{D}$. (3) If $\{A_n\}_{n \in \mathbb{N}}$ is a disjoint sequence in $\mathcal{D}$, then

\[ |Q\left( \bigcup_{n \in \mathbb{N}} A_n \right) - P\left( \bigcup_{n \in \mathbb{N}} A_n \right)| \leq \sum_{n \in \mathbb{N}} |Q(A_n) - P(A_n)| \leq cP\left( \bigcup_{n \in \mathbb{N}} A_n \right) . \]

Since each $A_n$ is a countable family of elements of $C$ we know that $\bigcup_{n \in \mathbb{N}} A_n$ is also a countable family of elements of $C$ and therefore $\bigcup_{n \in \mathbb{N}} A_n \in \mathcal{D}$. The result follows now from the monotone class theorem since $\mathcal{C} \subseteq \mathcal{D}$.

Corollary A.1 (Proof of Proposition 2.1). Let $([0, 1]^d, \mathcal{A}, P)$ be a probability space such that $P$ has a density $p$ that is lower bounded by $b > 0$. Let $\{\lambda_n\}_{n \geq 1}$ be a non-decreasing sequence in $\mathbb{N}_+$ such that $\lambda_n \to \infty$ and $|\lambda_n|_{\sigma(C)} \in \mathcal{O}_P^{n^2}(\sqrt{\log(n)}2^n\lambda_n^2/2)$. Furthermore, for any Borel set $A$ and $\epsilon > 0$ there exists an $n \in \mathbb{N}$ and $B \in \sigma(C_{\lambda})$ such that $P(A \Delta B) \leq \epsilon$.

Proof. The universal approximation property of the family of sets $C$ is well known. We provide here for completeness a simple proof. The set $\sigma(C_n)$ contains many intervals. In particular, to every element $x'$ in $\{ x : x_i \in 0, i \leq n \}$, where $l_n = (\sum_{i=1}^{\mu_n} d_i/2^i : d_i \in \{0, 1\})$, and any element $x''$ in $\{ x : x_i \in r_n, i \leq n \}$, where $r_n = \{ 1 - \sum_{i=1}^{\mu_n} d_i/2^i : d_i \in \{0, 1\} \}$, corresponds an interval $I = [x', x''] \in \sigma(C_n)$. Both, $\bigcup_{n \geq 1} l_n$ and $\bigcup_{n \geq 1} r_n$ lie dense in $[0, 1]$. This implies that any half-open interval $[a, b)$, $0 \leq a < b \leq 1$ for all $i \leq d$, can be approximated arbitrary well in Lebesgue measure, i.e. for $\epsilon > 0$ and with $\mu$ denoting Lebesgue-measure, there exists an $n \in \mathbb{N}$ and an $I \in \sigma(C_n)$ such that $[a, b) \subseteq I$ and $\mu(I \setminus [a, b)) \leq \epsilon$. Consider now any Borel subset $A$ of $[0, 1]^d$ and $\epsilon > 0$. Let $\{I_n\}_{n \geq 1}$ be a sequence of half-open intervals in $[0, 1]^d$ such that $A \subseteq \bigcup_{n \geq 1} I_n$ and $\mu(\bigcup_{n \geq 1} I_n) \leq \mu(A) + \epsilon/4$. Furthermore, select for each $I_n$ an half-open interval $I_n' \in \bigcup_{m \geq 1} \sigma(C_m)$ such that $I_n \subseteq I_n'$ and $\mu(I_n' \setminus I_n) \leq \epsilon/2^{n+2}$ then $A \subseteq \bigcup_{n \geq 1} I_n'$ and

\[ \mu\left( \bigcup_{n \geq 1} I_n' \setminus I_n \right) \leq \mu\left( \bigcup_{n \geq 1} I_n' \setminus \bigcup_{n \geq 1} I_n \right) + \mu\left( \bigcup_{n \geq 1} I_n \right) \leq \mu(A) + \epsilon/4 + \mu\left( \bigcup_{n \geq 1} (I_n' \setminus I_n) \right) \leq \mu(A) + \epsilon/2 . \]

Choose an $N$ such that $\mu(\bigcup_{n \geq N} I_n') \leq \epsilon/2$ and define $B = \bigcup_{n < N} I_n' \in \bigcup_{m \geq 1} \sigma(C_m)$. Then $\mu(B \setminus A) \leq \mu(\bigcup_{n \geq N} I_n') \leq \epsilon/2$ and $\mu(A \setminus B) \leq \mu(\bigcup_{n \geq N} I_n') \leq \epsilon/2$. Hence, $\mu(A \Delta B) \leq \epsilon$. Since $P$ is absolutely continuous with respect to Lebesgue-measure we can choose for every $\epsilon > 0$ a $\delta > 0$ such that $\mu(A \Delta B) \leq \delta$ implies $P(A \Delta B) \leq \epsilon$. And the second part of the proposition follows.
B Conditioning

**Proposition B.1** (Proof of Proposition 3.1). If $C \subseteq B_\mathcal{S}$ is a finite set with $\inf_{B \in C} P(B) > 0$, $\mathcal{F}$ is a subset of $L^1(P)$ uniformly bounded in supremum norm and $\mathcal{F}_C$ is a $P$-Donsker class then

$$\sup_{f \in \mathcal{F}} \sup_{B \in C} |E_n(f \mid B) - E(f \mid B)| \in O_P^*(n^{-1/2}).$$

Furthermore, if $C' \subseteq B_\mathcal{S}$, is such that $\mathcal{C}'_C$ is a $P$-Donsker class then

$$\sup_{A \in C'} \sup_{B \in C} |P_n(A \mid B) - P(A \mid B)| \in O_P^*(n^{-1/2}).$$

**Proof.** For a finite family of measurable sets $C$ the corresponding set of indicator functions $\mathcal{C}$ is always a $P$-Donsker class since for a single element the standard CLT provides the necessary statement and finite unions of $P$-Donsker classes are again $P$-Donsker due to (Dudley, 2014)[Thm 4.34]. Hence,

$$\sup_{f \in \mathcal{F}, B \in C} \left| \int_B f \, dP_n - \int_B f \, dP \right| = \sup_{f \in \mathcal{F}, B \in C} \left| \int f \times \chi(B) \, dP_n - \int f \times \chi(B) \, dP \right| = O_P^*(n^{-1/2})$$

and

$$\sup_{B \in C} |P_n(B) - P(B)| = O_P^*(n^{-1/2}).$$

By definition this implies that for $\epsilon > 0$ there exists an $M_1$ such that $\Pr^*\{\sup_{B \in C} |P(B) - P_n(B)| > M_1 n^{-1/2}\} < \epsilon/2$ for all $n \geq 1$. Let $N := \lceil (2M_1 / \inf_{B \in C} P(B))^2 \rceil$. Because $\inf_{B \in C} P(B) > 0$, for all $n \geq N$ we have for any $A \in C$ that

$$\{P_n(A) < \inf_{B \in C} P(B)/2\} \subseteq \{P_n(A) < P(A)/2\} = \{P_n(A) = P(A) > P(A)/2\}$$

$$\subseteq \{\sup_{B \in C} |P(B) - P_n(B)| > P(A)/2\} \subseteq \{\sup_{B \in C} |P(B) - P_n(B)| > \inf_{B \in C} P(B')/2\}$$

$$\subseteq \{\sup_{B \in C} |P(B) - P_n(B)| > M_1 n^{-1/2}\}.$$ 

Similarly, there exists an $M_2$ such that $\Pr^*\{\sup_{f \in \mathcal{F}, B \in C} \left| \int_B f \, dP_n - \int_B f \, dP \right| > M_2 n^{-1/2}\} < \epsilon/2$. The events

$$\Omega_n := \left\{ \sup_{B \in C} |P(B) - P_n(B)| \leq M_1 n^{-1/2} \right\} \cap \left\{ \sup_{f \in \mathcal{F}, B \in C} \left| \int_B f \, dP_n - \int_B f \, dP \right| \leq M_2 n^{-1/2} \right\}$$

have outer probability $\Pr^*(\Omega_n) \geq 1 - \epsilon$ and for all $n \geq N$ and $B \in C$, $\Omega_n \subseteq \{P_n(B) \geq \inf_{B' \in C} P(B')/2\}$. In the event $\Omega_n, n \geq N$, we know that $P_n(B) > 0$ and

$$E_n(f \mid B) - E(f \mid B) = \left( \int_B f \, dP_n \right) / P_n(B) - \left( \int_B f \, dP \right) / P(B)$$

$$= \left( P(B) \int_B f \, dP_n - P_n(B) \int_B f \, dP \right) / (P_n(B) P(B))$$

$$= \left( P(B) \left( \int_B f \, dP_n - \int_B f \, dP \right) + (P(B) - P_n(B)) \int_B f \, dP \right) / (P_n(B) P(B)).$$

Therefore, for $n \geq N$ in the event $\Omega_n$,

$$n^{1/2} \sup_{f \in \mathcal{F}, B \in C} \left| \left( \int_B f \, dP_n \right) / P_n(B) - \left( \int_B f \, dP \right) / P(B) \right| \leq 2 \left( M_2 + b M_1 \right) / c^2,$$

where $b := \sup_{x \in \mathcal{S}, f \in \mathcal{F}} |f(x)|$ and $c := \inf_{B \in C} P(B) > 0$.

For any $n < N$ and $B$ with $P_n(B) = 0$ the estimate $E_n(f \mid B) = 0$ by definition and

$$\sup_{f \in \mathcal{F}} |E_n(f \mid B) - E(f \mid B)| = \sup_{f \in \mathcal{F}} \left| \left( \int_B f \, dP \right) / P(B) \right| \leq b < \infty.$$
For any $n < N$ with $P_n(B) > 0$ we have that

$$n^{1/2} \sup_{f \in \mathcal{F}, B \in \mathcal{C}} \left| \left( \int_B f \, dP_n \right) / P_n(B) - \left( \int_B f \, dP \right) / P(B) \right| \leq 2n^{1/2}b < \infty$$

and with the constant $M := \max\{2(M_2 + bM_1)/c^2, 2N^{1/2}b\}$ we have

$$\Pr^* \{ \sup_{f \in \mathcal{F}, B \in \mathcal{C}} |E_n(f \mid B) - E(f \mid B)| > Mn^{-1/2} \} \leq \epsilon.$$

This is sufficient to prove the first claim. The second claim follows from the first by substituting $\mathcal{C}'$ for $\mathcal{F}$.

**Proposition B.2** (Proof of Proposition 3.2). If $\mathcal{C} \subseteq \mathcal{B}_\delta$, $\mathcal{C}$ and $\mathcal{F}_C$ are $P$-Donsker classes, $\mathcal{F}$ is a subset of $\mathcal{L}^1(P)$ uniformly class bounded in supremum norm and $P$ has a density which is lower bounded by a constant $b > 0$ then with $\mathcal{C}_n := \{ C : C \in \mathcal{C}, \mu(C) \geq n^{-\alpha}\}$ and $\alpha \in [0, 1/2)$

$$\sup_{f \in \mathcal{F}, B \in \mathcal{C}_n} |E_n(f \mid B) - E(f \mid B)| \in O_p(n^{\alpha-1/2}).$$

Furthermore, if $\mathcal{C}' \subseteq \mathcal{B}_\delta$ is such that $\mathcal{C}'$ is a $P$-Donsker class then

$$\sup_{A \in \mathcal{C}', B \in \mathcal{C}_n} |P_n(A \mid B) - P(A \mid B)| \in O_p(n^{\alpha-1/2}).$$

**Proof.** As in the proof of Proposition 3.1, $\sup_{f \in \mathcal{F}, B \in \mathcal{C}} \left| \int_B f \, dP_n - \int_B f \, dP \right| \in O_p(n^{-1/2})$ and $\sup_{B \in \mathcal{C}} |P_n(B) - P(B)| \in O_p(n^{-1/2})$. Hence, for a given $\epsilon > 0$ there exists a $M_1$ such that

$$\Pr^* \{ \sup_{B \in \mathcal{C}_n} |P_n(B) - P(B)| > M_1n^{-1/2} \} \leq \frac{\epsilon}{2}.$$}

In particular, by assumption $\inf_{B \in \mathcal{C}_n} P(B')/2 \geq b \inf_{B \in \mathcal{C}_n} \mu(B')/2 \geq (b/2)n^{-\alpha}$ we have for $N := \left( (2M_1/b)^{1/(1/2-\alpha)} \right)$ and all $n \geq N$ such that

$$\{ P_n(B) < \inf_{B' \in \mathcal{C}_n} P(B')/2 \} \subseteq \{ \sup_{B \in \mathcal{C}_n} |P_n(B) - P_n(B)| > b \inf_{B \in \mathcal{C}_n} P(B')/2 \} \subseteq \{ \sup_{B \in \mathcal{C}_n} |P_n(B) - P_n(B)| > (b/2)n^{-\alpha} \}$$

and, since the last event has an outer probability strictly less than $\epsilon/2$, there exists an $M_2$ such that for all $n \geq N$ the events

$$\Omega_n := \left\{ \sup_{B \in \mathcal{C}} |P_n(B) - P_n(B)| \leq M_1n^{-1/2} \right\} \cap \left\{ \sup_{f \in \mathcal{F}, B \in \mathcal{C}_n} \left| \int_B f \, dP_n - \int_B f \, dP \right| \leq M_2n^{-1/2} \right\}$$

have outer probability $\Pr^*(\Omega_n) \geq 1 - \epsilon$ and $\{ P_n(B) \geq \inf_{B' \in \mathcal{C}_n} P(B')/2 \} \supseteq \Omega_n$ for all $n \geq N$.

Using the bound (8)

$$\left| \left( \int_B f \, dP_n \right) / P_n(B) - \left( \int_B f \, dP \right) / P(B) \right| \leq \left( \int_B f \, dP_n - \int_B f \, dP \right) / P_n(B) + \left( P(B) - P_n(B) \right) \int_B f \, dP / (P_n(B)P(B)).$$

Let $c = \sup_{x \in \mathcal{X}} \sup_{f \in \mathcal{F}} |f(x)| < \infty$ then

$$\Pr^* |f(x)| < \infty \text{ and, since } P_n(B) \geq (b/2)n^{-\alpha} \text{ on } \Omega_n, \text{ for } n \geq N$$

$$\sup_{f \in \mathcal{F}, B \in \mathcal{C}_n} |E_n(f \mid B) - E(f \mid B)| \leq (2M_2/b)n^{\alpha-1/2} + (2cM_1/b)n^{\alpha-1/2}.$$

As in the the proof of Proposition 3.1 the errors for $n < N$ can be bounded since $\inf_{n < N} \inf_{B \in \mathcal{C}_n} P(B) \geq bN^{-\alpha}$ and the first result follows. Substituting $\mathcal{C}'$ for $\mathcal{F}$ yields the second claim.

**Lemma B.1.** If $\mathcal{G}$ is a $\sigma$-algebra consisting of finitely many elements then there exists a unique partition $\mathcal{P}_G$ such that each element of $\mathcal{B}$ can be represented as a finite union of elements of $\mathcal{P}_G$. 

Steffen Grünewälder
**Proposition B.3** (Proof of Proposition 3.3). If \( \mathcal{G} \subseteq \mathcal{B}_S \) is a \( \sigma \)-algebra consisting of finitely many sets, \( \mathcal{F} \) is a subset of \( \mathcal{L}^1(P) \) uniformly bounded in supremum norm, \( \mathcal{G}_\mathcal{F} \) is a \( P \)-Donsker class then

\[
\sup_{f \in \mathcal{F}} \| E_n(f \mid \mathcal{G}) - E(f \mid \mathcal{G}) \|_{\mathcal{L}^1(P)} \in O_P^\star(n^{-1/2}).
\]

Furthermore, if \( C \subseteq \mathcal{B}_S \) is such that \( \mathcal{C}_\mathcal{G} \) is a \( P \)-Donsker class then

\[
\sup_{A \in \mathcal{C}} \| P_n(A \mid \mathcal{G}) - P(A \mid \mathcal{G}) \|_{\mathcal{L}^1(P)} \in O_P^\star(n^{-1/2}).
\]

**Proof.** If \( P(B) = 0 \) for some \( B \in \mathcal{G} \) then \( \int_B E_n(f \mid \mathcal{G}) - E(f \mid \mathcal{G}) \) \( dP = 0 \). Let \( \mathcal{G}' = \{ B : B \in \mathcal{G}, P(B) > 0 \} \) then from Proposition 3.1 it follows that

\[
\sup_{f \in \mathcal{F}, B \in \mathcal{G}'} \left| \int_B E_n(f \mid \mathcal{G}) - E(f \mid \mathcal{G}) \right| \in O_P^\star(n^{-1/2}).
\]

For \( \epsilon > 0 \), choose \( M, N \) such that \( E_n = \sup_{f \in \mathcal{F}} \sup_{B \in \mathcal{G}'} \left| \int_B E_n(f \mid \mathcal{G}) - E(f \mid \mathcal{G}) \right| \leq M n^{-1/2} \) has probability \( P(E_n) \geq 1 - \epsilon \) for all \( n \geq N \). For \( B \in \mathcal{G}' \), let \( B_1, \ldots, B_m \in \mathcal{G}_\mathcal{F} \subseteq \mathcal{G} \) such that \( B = B_1 \cup \ldots \cup B_m \), then \( P \)-a.s. \( E(f \mid \mathcal{G}) = \sum_{i \leq m} E(f \mid B_i) \chi_{B_i} \) and

\[
\left| \int_B (E_n(f \mid \mathcal{G}) - E(f \mid \mathcal{G})) \right| \leq \sum_{i \leq m} \left| \int_{B \cap B_i} (E_n(f \mid B_i) - E(f \mid B_i)) \right| dP
\]

\[
\leq \sum_{i \leq m} \left| E_n(f \mid B_i) - E(f \mid B_i) \right| P(B \cap B_i) \leq M n^{-1/2} \sum_{i \leq m} P(B \cap B_i)
\]

\[
= M n^{-1/2} P(B) \leq M n^{-1/2}.
\]

Furthermore,

\[
\| E_n(f \mid \mathcal{G}) - E(f \mid \mathcal{G}) \|_{\mathcal{L}^1(P)} \leq 2 \sup_{B \in \mathcal{G}'} \left| \int_B E_n(f \mid \mathcal{G}) - E(f \mid \mathcal{G}) \right| \leq 2M n^{-1/2}.
\]

This implies the result. The second result follows again from the first. \( \square \)

**Proposition B.4** (Proof of Proposition 3.4). Let \( (0, 1]^d, \mathcal{B}, P \) be a probability space such that \( P \) has a density \( p \) that is lower bounded by \( \epsilon > 0 \) and let \( \{ \lambda_n \}_{n \geq 1} \) be a non-decreasing sequence in \( \mathbb{N}_+ \) such that

\[
\lambda_n \in o\left( \frac{1}{3d \log(2) \log\left( \frac{n}{\sqrt{\log(n)}} \right)} \right).
\]

If \( \mathcal{F} \) is a subset of \( \mathcal{L}^1(P) \) uniformly bounded in supremum norm which fulfills Equation 4, then

\[
\sup_{f \in \mathcal{F}} \| E_n(f \mid \mathcal{G}_{\lambda_n}) - E(f \mid \mathcal{G}_{\lambda_n}) \|_{\infty} \in O_P^\star(\sqrt{\log(n)} 2^{(3/2) d \lambda_n n^{-1/2}}).
\]
Furthermore, if \( C \subseteq \mathcal{B}_S \), is such that \( C \) fulfills Equation 4 then

\[
\sup_{A \in C} \| P_n(A \mid \mathcal{G}_{\lambda_n}) - P(A \mid \mathcal{G}_{\lambda_n}) \|_\infty \in O_p(\sqrt{\log(n)}2^{(3/2)d\lambda_n}n^{-1/2}).
\]

**Proof.** By assumption and from Corollary 2.1 we know that

\[
\sup_{f \in F} \sup_{B \in \mathcal{B}_{\lambda_n}} \left| \int_B f \, dP_n - \int_B f \, dP \right| \in O_p(\sqrt{\log(n)}2^{d\lambda_n/2}n^{-1/2}),
\]

\[
\sup_{B \in \mathcal{B}_{\lambda_n}} \left| P_n(B) - P(B) \right| \in O_p(\sqrt{\log(n)}2^{d\lambda_n/2}n^{-1/2}).
\]

Furthermore, \( \inf_{B \in \mathcal{B}_{\lambda_n}} P(B) \geq b2^{-d\lambda_n} \). By combining the technique in the proof of Proposition 3.2 with the assumption on the rate of \( \lambda_n \) we have for any \( B \in \mathcal{B}_{\lambda_n} \) that

\[
\left\{ P_n(B) < \inf_{B' \in \mathcal{B}_{\lambda_n}} P(B')/2 \right\} \subseteq \left\{ \sup_{B \in \mathcal{B}_{\lambda_n}} \left| P(B) - P_n(B) \right| > M \sqrt{\log(n)}2^{d\lambda_n/2}n^{-1/2} \right\}
\]

for some constant \( N, M \) and all \( n \geq N \). The same line of reasoning as in Proposition 3.2 then shows

\[
\sup_{f \in F} \sup_{B \in \mathcal{B}_{\lambda_n}} \left| E_n(f \mid B) - E(f \mid B) \right| \in O_p(\sqrt{\log(n)}2^{(3/2)d\lambda_n}n^{-1/2}).
\]

Substituting in the definition of \( E_n(f \mid \mathcal{G}_{\lambda_n}) \) and \( E(f \mid \mathcal{G}_{\lambda_n}) \) gives us the first result,

\[
\sup_{f \in F} \| E_n(f \mid \mathcal{G}_{\lambda_n}) - E(f \mid \mathcal{G}_{\lambda_n}) \|_\infty \leq \sup_{f \in F} \sup_{B \in \mathcal{B}_{\lambda_n}} \left| E_n(f \mid B) - E(f \mid B) \right| \in O_p(\sqrt{\log(n)}2^{(3/2)d\lambda_n}n^{-1/2}).
\]

The second claim is directly implied by this result.