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)**.** *$\mathcal{C}$ 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. **P** Consider arbitrary 2d + 1 points $\mathbf{x}_1, \mathbf{x}_2, \ldots \in \mathbb{X}$. Now, there is at least one element $\mathbf{x}_i$ which attains the maximum in dimension j, $1 \leq j \leq d$, i.e. $\mathbf{x}_{ij} = \max_{l \leq d} \mathbf{x}_{lj}$. 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 $\mathbf{x}_1, \ldots, \mathbf{x}_{2d+1}$ and there is no set $A \in C$ which fulfills $A \cap \{x_i : 1 \leq i \leq 2d+1\} = B$. **Q**

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

(iii) $\mathcal{C}$ is *image admissible Suslin*. **P** Let $\mathbb{Y} = \mathbb{X} \times [0, \infty[$ equipped with the natural topology, which is the one induced by the Euclidean metric, and the corresponding Borel σ-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 $T(\mathbf{y}, h) = \chi([\mathbf{y}, \mathbf{y} + h\mathbf{1}])$ if $[\mathbf{y}, \mathbf{y} + h\mathbf{1}] \in C$ and $\chi(\emptyset)$ otherwise. T maps $\mathbb{Y}$ onto $\mathcal{C}$. We need to verify that $(\mathbf{y}, h, x) \mapsto (T(\mathbf{y}, h))(x)$ is jointly measurable, that is for any Borel subset A of $\mathbb{R}$

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

must be in the product σ-algebra. $T(\mathbf{y}, h)(x)$ 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 σ-algebra.

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

$$\begin{aligned} B &= \{(\mathbf{y}, h, \mathbf{x}) : \mathbf{x}, \mathbf{y} \in \mathbb{X}, h \in [0, \infty[, \chi([\mathbf{y}, \mathbf{y} + h\mathbf{1}])(\mathbf{x}) = 1\} \\ &= \{(\mathbf{y}, h, \mathbf{x}) : \mathbf{x}, \mathbf{y} \in \mathbb{X}, h \in [0, \infty[, \mathbf{x} \in [\mathbf{y}, \mathbf{y} + h\mathbf{1}]\}. \end{aligned}$$

This set is closed in the natural topology of $\mathbb{X} \times [0, \infty[\times \mathbb{X}$ since, if $(\mathbf{y}, h, \mathbf{x})$ is not in B then with $\epsilon < d(\mathbf{x}, [\mathbf{y}, \mathbf{y} + h\mathbf{1}])$, where d is the Euclidean metric, we have that the open ball

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

contains $(\mathbf{y}, h, \mathbf{x})$. And, since,

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

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. **Q** □

Lemma A.2 (Proof of Lemma 2.3)**.** *Let $\mathbb{X}$ be any set and $\mathcal{G}$ be a σ-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 $\mathcal{G}$. **P** 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

$A_1, \dots, A_n \in \mathcal{G}$ are not empty and mutually disjoint. There exists an element $B \in \mathcal{G}$ that is not contained in $\sigma(A_1, \dots, A_n)$ since otherwise $\mathcal{G} = \sigma(A_1, \dots, A_n)$ and $\mathcal{G}$ would be finite. Take such an element B. If $B \setminus \bigcup_{i \le n} A_n \in \mathcal{G}$ is not empty then add this to the sequence as A_{n+1}. A_{n+1} is then obviously disjoint from all $A_1, \dots, A_n$. If $B \setminus \bigcup_{i \le n} A_n = \emptyset$ then $B \subsetneq \bigcup_{i \le n} A_n$. Furthermore, there is some $i \le n$ such that $A_i \setminus B \neq \emptyset \neq A_i \cap B$, because otherwise B would be a union of a subset of $A_1, \dots, A_n$. 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**

(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 $x_1, \dots, x_k$ and any subset of these, say $\{x_{n_i} : i \le l, 0 \le n_1 < n_2 \dots < n_l \le k\}$ for $l \le k$ and consider the corresponding sequence of sets $\{A_{n_i} : i \le l\}$ then $\{x_{n_i} : i \le l\} = \{x_n\}_{n \le k} \cap \bigcup \{A_{n_i} : i \le l\}$ and the set $\{x_i\}_{i \le k}$ 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 $\mathcal{C} \subset \mathcal{A}$ a disjoint family of sets such that for each $A \in \mathcal{C}$ there exists $\{A_n\}_{n \in \mathbb{N}}$ in $\mathcal{C}$ with $\Omega \setminus A = \bigcup_{n \in \mathbb{N}} A_n$ and $\emptyset \in \mathcal{C}$. For any measure Q for which there exists a constant $c > 0$ such that for all $A \in \mathcal{C}$, $|Q(A) - P(A)| \le cP(A)$, we have that* $\sup_{A \in \sigma(\mathcal{C})} |Q(A) - P(A)|/P(A) \le c$.

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

$$\mathcal{D} := \{A : A = \bigcup \mathcal{E}, \mathcal{E} \subseteq \mathcal{C} \text{ a countable family}, |Q(A) - P(A)| \le cP(A)\}.$$

$\mathcal{D}$ 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

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

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

$$\left| Q\Big(\bigcup_{n \in \mathbb{N}} A_n\Big) - P\Big(\bigcup_{n \in \mathbb{N}} A_n\Big) \right| \le \sum_{n \in \mathbb{N}} |Q(A_n) - P(A_n)| \le cP\Big(\bigcup_{n \in \mathbb{N}} A_n\Big).$$

Since each A_n is a countable family of elements on $\mathcal{C}$ we know that $\bigcup_{n \in \mathbb{N}} A_n$ is also a countable family of elements of $\mathcal{C}$ 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 $\{\lambda_n\}_{n \ge 1}$ be a non-decreasing sequence in $\mathbb{N}_+$ such that* $\lim_{n \to \infty} \lambda_n = \infty$ *then* $\|\nu_n(A)\|_{\sigma(\mathcal{C}_{\lambda_n})} \in O_P^*(\sqrt{\log(n)} 2^{d\lambda_n/2})$*. Furthermore, for any Borel set A and $\epsilon > 0$ there exists an $n \in \mathbb{N}$ and $B \in \sigma(\mathcal{C}_{\lambda_n})$ such that $P(A \Delta B) \le \epsilon$.*

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

$$\mu(\bigcup_{n \ge 1} \mathbf{I}'_n) \le \mu(\bigcup_{n \ge 1} \mathbf{I}'_n \setminus \bigcup_{n \ge 1} \mathbf{I}_n) + \mu(\bigcup_{n \ge 1} \mathbf{I}_n) \le \mu(A) + \epsilon/4 + \mu(\bigcup_{n \ge 1} (\mathbf{I}'_n \setminus \mathbf{I}_n)) \le \mu(A) + \epsilon/2.$$

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

B Conditioning

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

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

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

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

Proof. For a finite family of measurable sets $\mathcal{C}$ the corresponding set of indicator functions $\mathscr{C}$ 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,

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

and

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

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

$$\begin{aligned}
&\{P_n(A) < \inf_{B\in\mathcal{C}} P(B)/2\} \subseteq \{P_n(A) < P(A)/2\} = \{P(A) - P_n(A) > P(A)/2\} \\
&\subseteq \{\sup_{B\in\mathcal{C}} |P(B) - P_n(B)| > P(A)/2\} \subseteq \{\sup_{B\in\mathcal{C}} |P(B) - P_n(B)| > \inf_{B'\in\mathcal{C}} P(B')/2\} \\
&\subseteq \{\sup_{B\in\mathcal{C}} |P(B) - P_n(B)| > M_1 n^{-1/2}\}.
\end{aligned}$$

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

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

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

$$\begin{aligned}
&E_n(f \mid B) - E(f \mid B) \qquad (8)\\
&= \left(\int_B f\, dP_n\right) \Big/ P_n(B) - \left(\int_B f\, dP\right) \Big/ P(B) \\
&= \left(P(B) \int_B f\, dP_n - P_n(B) \int_B f\, dP\right) \Big/ (P_n(B)P(B)) \\
&= \left(P(B) \left(\int_B f\, dP_n - \int_B f\, dP\right) + (P(B) - P_n(B)) \int_B f\, dP\right) \Big/ (P_n(B)P(B)).
\end{aligned}$$

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

$$n^{1/2} \sup_{f\in\mathcal{F},B\in\mathcal{C}} \left| \left(\int_B f\, dP_n\right) \Big/ P_n(B) - \left(\int_B f\, dP\right) \Big/ P(B) \right| \leq 2\,(M_2 + bM_1)/c^2,$$

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

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

$$\sup_{f\in\mathcal{F}} |E_n(f \mid B) - E(f \mid B)| = \sup_{f\in\mathcal{F}} \left| \left(\int_B f\, dP\right) \Big/ P(B) \right| \leq b < \infty.$$

For any $n < N$ with $P_n(B) > 0$ we have that

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

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

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

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

Proposition B.2 (Proof of Proposition 3.2)**.** *If $\mathcal{C} \subseteq \mathscr{B}_{\mathbb{S}}$, $\mathscr{C}$ and $\mathcal{F}_{\mathcal{C}}$ are P-Donsker classes, $\mathcal{F}$ is a subset of $\mathcal{L}^1(P)$ uniformly 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 $\alpha \in [0, 1/2)$*

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

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

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

Proof. As in the 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 $\Pr^*\{\sup_{B \in \mathcal{C}} |P(B) - P_n(B)| > M_1 n^{-1/2}\} < \epsilon/2$. In particular, since 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 $N := \lceil (2M_1/b)^{1/(1/2 - \alpha)} \rceil$ and all $n \geq N$, $B \in \mathcal{C}_n$ that

$$\begin{aligned}
\{P_n(B) < \inf_{B' \in \mathcal{C}_n} P(B')/2\} &\subseteq \{\sup_{B \in \mathcal{C}_n} |P(B) - P_n(B)| > \inf_{B' \in \mathcal{C}_n} P(B')/2\} \\
&\subseteq \{\sup_{B \in \mathcal{C}_n} |P(B) - P_n(B)| > (b/2)n^{-\alpha}\} \subseteq \{\sup_{B \in \mathcal{C}_n} |P(B) - P_n(B)| > M_1 n^{-1/2}\} \\
&\subseteq \{\sup_{B \in \mathcal{C}} |P(B) - P_n(B)| > M_1 n^{-1/2}\}.
\end{aligned}$$

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

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

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

Using the bound (8)

$$\begin{aligned}
&\left| \left(\int_B f \, dP_n \right) \Big/ P_n(B) - \left(\int_B f \, dP \right) \Big/ P(B) \right| \\
&\leq \left| \left(\int_B f \, dP_n - \int_B f \, dP \right) \Big/ P_n(B) \right| + \left| (P(B) - P_n(B)) \int_B f \, dP \Big/ (P_n(B) P(B)) \right|.
\end{aligned}$$

Let $c = \sup_{x \in \mathbb{S}} \sup_{f \in \mathcal{F}} |f(x)| < \infty$ then $\left| \int_B f \, dP \right| \leq c P(B)$ and, since $P_n(B) \geq (b/2)n^{-\alpha}$ on Ω_n, for $n \geq N$

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

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 $\mathscr{C}'$ for $\mathcal{F}$ yields the second claim. □

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

Proof. **(i)** Uniqueness. Assume there are two partitions $\mathcal{P}_{\mathcal{G}}$ and $\mathcal{P}'_{\mathcal{G}}$. There must be an element $B \in \mathcal{P}'_{\mathcal{G}}$ that is not in $\mathcal{P}'_{\mathcal{G}}$ (or vice versa) since otherwise $\mathcal{P}'_{\mathcal{G}} = \mathcal{P}_{\mathcal{G}}$.This B is equal to $A_1 \cup \ldots \cup A_n$ for some $n \in \mathbb{N}$ and disjoint, non-empty, elements $A_1, \ldots, A_n \in \mathcal{P}_{\mathcal{G}}$. Now $A_1 \subsetneq B$ but A_1 cannot be a finite union of elements of $\mathcal{P}'_{\mathcal{G}}$ since $\mathcal{P}'_{\mathcal{G}}$ is a partition.

(ii) Enumerate $\mathcal{G}$ as $A_1, \ldots, A_n$ where n is the cardinality of $\mathcal{G}$. We construct a partition iterartively. I.e. we construct for each $i \leq n$ a family of sets $\mathcal{B}_i = \{B_{1i}, \ldots, B_{m_i,i}\}$ that is disjoint and such that $A_1, \ldots A_i$ can be represented as unions of the $B_{j,i}$ elements. For $i = 1$ let $B_{11} = A_1$. Now assume we have such a family of sets for i and we want to get a family of disjoint sets for $A_1, \ldots, A_{i+1}$. Consider the following family of set

$$\mathcal{B}_{i+1} = \{A_{i+1} \backslash \bigcup \mathcal{B}_i\} \cup \{B \backslash A_{i+1} : B \in \mathcal{B}_i\} \cup \{A_{i+1} \cap B : B \in \mathcal{B}_i\}.$$

Clearly ever $B \in \mathcal{B}_i$ can still be represented as a union of elements since $B = (B \backslash A_{i+1}) \cup (B \cap A_{i+1})$. Similarly, $A_{i+1} = (A_{i+1} \backslash \bigcup \mathcal{B}_i) \cup (A_{i+1} \cap \bigcup \mathcal{B}_i)$ can be represented. The family of sets is also disjoint: any element of the form $B \backslash A_{i+1}$ is certainly disjoint from any $B' \cap A_{i+1}$ and $A_{i+1} \backslash B'' \supset A_{i+1} \backslash \bigcup \mathcal{B}_i$. Also, since the elements in $\mathcal{B}_i$ are disjoint any two $B \backslash A_{i+1}$ and $B' \backslash A_{i+1}$ will be disjoint. Finally, any $A_{i+1} \cap B$ is disjoint of $A_{i+1} \backslash \bigcup \mathcal{B}_i$ and $(A_{i+1} \cap B) \cap (A_{i+1} \cap B') = \emptyset$. This concludes the induction. □

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

$$\sup_{f \in \mathcal{F}} \| E_n\left(f \mid \mathcal{G}\right) - E\left(f \mid \mathcal{G}\right) \|_{\mathcal{L}^1(P)} \in O_P^*(n^{-1/2}).$$

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

$$\sup_{A \in \mathcal{C}} \| P_n\left(A \mid \mathcal{G}\right) - P\left(A \mid \mathcal{G}\right) \|_{\mathcal{L}^1(P)} \in O_P^*(n^{-1/2}).$$

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

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

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

$$\begin{aligned}
&\Big|\int_B (E_n(f|\mathcal{G}) - E(f|\mathcal{G}))\, dP\Big| \leq \sum_{i \leq m} \Big|\int_{B \cap B_i} (E_n(f|B_i) - E(f|B_i))\, dP\Big| \\
&\leq \sum_{i \leq m} |E_n(f|B_i) - E(f|B_i)| P(B \cap B_i) \leq M n^{-1/2} \sum_{i \leq m} P(B \cap B_i) \\
&= M n^{-1/2} P(B) \leq M n^{-1/2}.
\end{aligned}$$

Furthermore,

$$\| E_n(f|\mathcal{G}) - E(f|\mathcal{G}) \|_{\mathcal{L}^1(P)} \leq 2 \sup_{B \in \mathcal{G}} \Big| \int_B E_n(f|\mathcal{G}) - E(f|\mathcal{G})\, dP \Big| \leq 2 M n^{-1/2}.$$

This implies the result. The second result follows again from the first. □

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 $b > 0$ and let $\{\lambda_n\}_{n \geq 1}$ be a non-decreasing sequence in $\mathbb{N}_+$ such that*

$$\lambda_n \in o\bigg(\frac{1}{3d \log(2)} \log\Big(\frac{n}{\sqrt{\log(n)}}\Big)\bigg).$$

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

$$\sup_{f \in \mathcal{F}} \| E_n\left(f \mid \mathcal{G}_{\lambda_n}\right) - E\left(f \mid \mathcal{G}_{\lambda_n}\right) \|_\infty \in O_P^*(\sqrt{\log(n)} 2^{(3/2)d\lambda_n} n^{-1/2}).$$

Furthermore, if $\mathcal{C} \subseteq \mathcal{B}_{\mathbb{S}}$, is such that $\mathcal{C}$ fulfills Equation 4 then

$$\sup_{A\in\mathcal{C}} \| P_n \left(A \,\middle|\, \mathcal{G}_{\lambda_n}\right) - P \left(A \,\middle|\, \mathcal{G}_{\lambda_n}\right) \|_\infty \in O_P^*(\sqrt{\log(n)} 2^{(3/2)d\lambda_n} n^{-1/2}).$$

Proof. By assumption and from Corollary 2.1 we know that

$$\sup_{f\in\mathcal{F}} \sup_{B\in\mathcal{G}_{\lambda_n}} \Big| \int_B f\, dP_n - \int_B f dP \Big| \in O_P^*(\sqrt{\log(n)} 2^{d\lambda_n/2} n^{-1/2}),$$
$$\sup_{B\in\mathcal{G}_{\lambda_n}} |P_n(B) - P(B)| \in O_P^*(\sqrt{\log(n)} 2^{d\lambda_n/2} n^{-1/2}).$$

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

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

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

$$\sup_{f\in\mathcal{F}} \sup_{B\in\mathcal{G}_{\lambda_n}} |E_n(f|B) - E(f|B)| \in O_P^*(\sqrt{\log(n)} 2^{(3/2)d\lambda_n} n^{-1/2}).$$

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

$$\sup_{f\in\mathcal{F}} \|E_n(f|\mathcal{G}_{\lambda_n}) - E(f|\mathcal{G}_{\lambda_n})\|_\infty \leq \sup_{f\in\mathcal{F}} \sup_{B\in\mathcal{G}_{\lambda_n}} |E_n(f|B) - E(f|B)| \in O_P^*(\sqrt{\log(n)} 2^{(3/2)d\lambda_n} n^{-1/2}).$$

The second claim is directly implied by this result. □