Empirical Processes: Probability Inequalities and Rates of Convergence

Ahnaf Rafi

August 24, 2025

1 Introduction

Let $n \in \mathbb{N}$, $(\Omega, \mathscr{A}, \mathrm{Pr})$ be an underlying probability space and $(\mathcal{X}, \mathscr{X})$ be a measurable space. Let Q be a signed measure on $(\mathcal{X}, \mathscr{X})$, and define

$$Qg := \int g(x) Q(\mathrm{d}x) \quad \text{if } g \in \mathscr{L}_1(Q),$$
$$\text{and} \quad \|Q\|_{\mathcal{F}} := \sup_{f \in \mathcal{F}} |Qf| \quad \text{if} \quad \mathcal{F} \subseteq \mathscr{L}_1(Q).$$

Let $X_1, \ldots, X_n$ be independent $\mathcal{X}$-valued random elements defined on $(\Omega, \mathscr{A}, \mathrm{Pr})$ and denote the Pr-law of X_i by P_i. Define the measures:

$$\mathbb{P}(A) \equiv \mathbb{P}_n(A) := \frac{1}{n} \sum_{i=1}^{n} \delta_{\{X_i\}}(A) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}\{X_i \in A\}$$
$$\text{and} \quad \overline{P}(A) \equiv \overline{P}_n(A) := \frac{1}{n} \sum_{i=1}^{n} P_i(A).$$

For $\mathcal{F} \subseteq \mathscr{L}_1(\overline{P})$, the empirical process is the map $f \mapsto \left(\mathbb{P} - \overline{P}\right)[f] := \frac{1}{n} \sum_{i=1}^{n} \left(f\left(X_i\right) - P_i f\right)$.

Our main goal here is to provide upper bounds for the tail outer probability $\mathrm{Pr}^* \left\{ \left\| \mathbb{P} - \overline{P} \right\|_{\mathcal{F}} > y \right\}$ as a function of y, n and $\mathcal{F}$. Dependence on $\mathcal{F}$ can appear in two ways. The probability bounds can depend on the supremum second moment of $\mathcal{F}$:

$$\kappa_2(Q, \mathcal{F}) := \sup_{f \in \mathcal{F}} \sqrt{Q\left[f^2\right]} \quad \text{for a positive measure } Q.$$

Furthermore, the bounds will depend on the complexity of $\mathcal{F}$; as in the rest of the empirical process literature, we use covering numbers under $\mathscr{L}_p$ norms, see Definition 1.1 below. Throughout, we will consider the case where functions in $\mathcal{F}$ are uniformly bounded in magnitude by 1.

Definition 1.1 ($\mathscr{L}_p(Q)$-covering numbers)**.** Let Q be a positive measure on $(\mathcal{X}, \mathscr{X})$ and denote

$$\|f\|_{p,Q} = \begin{cases} (Q[|f|^p])^{\frac{1}{p}} & \text{if } 1 \le p < \infty, \\ \inf\{t > 0 : |f| \le t \ Q\text{-a.e.}\} & \text{if } p = \infty, \end{cases}$$

$$\mathscr{L}_p(Q) = \{f \text{ measurable such that } \|f\|_{p,Q} < \infty\}.$$

For $\mathcal{F} \subseteq \mathscr{L}_p(Q)$ and $\varepsilon > 0$, define

$$N(\varepsilon, \mathscr{L}_p(Q), \mathcal{F}) := \min\left\{k \in \mathbb{N} : \exists f_1, \ldots, f_k \in \mathcal{F} \text{ s.t. } \min_{1 \le j \le k} (Q[|f - f_j|^p])^{\frac{1}{p}} < \varepsilon \ \forall f \in \mathcal{F}\right\}.$$

Upon characterizing $\Pr^*\left\{\|\mathbb{P} - \overline{P}\|_{\mathcal{F}} > y\right\}$, a second goal is then to explore how these probability bounds can then be used to derive rates of convergence. A tertiary aim will be to characterize how and when the chaining method offers an improvement over more crude methods using only covering numbers, for both probability inequalities and convergence rate results.

2 Preliminary probability inequalities

Our first main result is the following bound.

Theorem 2.1. *Let $\mathcal{F}$ satisfy $\sup_{f \in \mathcal{F}, y \in \mathcal{X}} |f(y)| \le 1$ and let $x, t > 0$ satisfy*

$$x \ge \frac{1}{\sqrt{8n}} \sup_{f \in \mathcal{F}} \sqrt{\frac{1}{n} \sum_{i=1}^n \operatorname{Var}[f(X_i)]} \quad \text{and} \quad t \ge \kappa_2\left(\overline{P}, \mathcal{F}\right).$$

Then

$$\begin{aligned} &\Pr^*\left\{\left\|\mathbb{P} - \overline{P}\right\|_{\mathcal{F}} > 8x\right\} \le \min\{1, \mathrm{E}[\phi(x, t; \mathbb{P}, \mathcal{F})]\}, \quad \text{where} \\ &\phi(x, t; \mathbb{P}, \mathcal{F}) = 8\exp\left(-\frac{nx^2}{128t^2}\right) N(x, \mathscr{L}_1(\mathbb{P}), \mathcal{F}) + 16 N(t, \mathscr{L}_2(\mathbb{P}), \mathcal{F}) \exp\left(-nt^2\right). \end{aligned} \tag{2.1}$$

In (2.1), the trivial upper bound of one comes from the fact that $\Pr^*$ is an outer probability. We prove the bound due to the function ϕ in Theorem 2.1 as follows. First, we bound $\Pr^*\left\{\left\|\mathbb{P}-\overline{P}\right\|_{\mathcal{F}}>y\right\}$ by a tail outer probability for a symmetrized empirical process. Then, we bound the outer probability "conditional" on symmetrization variables by an squared-exponential decay. This is the first summand in ϕ in (2.1). The squared-exponential decay function in the second step depends on the largest empirical second moment $\kappa_2(\mathbb{P},\mathcal{F})$. The second summand entering ϕ in (2.1) results from a bound on deviation probabilities for $\kappa_2(\mathbb{P},\mathcal{F})/\kappa_2(\overline{P},\mathcal{F})$.

The symmetrized empirical process is

$$\mathbb{P}^\circ f \equiv \mathbb{P}_n^\circ f := \frac{1}{n}\sum_{i=1}^{n} U_i f\left(X_i\right), \tag{2.2}$$

where $U_1,\ldots,U_n \overset{\text{iid}}{\sim}$ Rademacher with $\mathbf{U}:=(U_1,\ldots,U_n)$ independent to $\mathbf{X}:=(X_1,\ldots,X_n)$.

Remark 2.1. For random elements V_1 and V_2 and an arbitrary (not necessarily measurable) map $T\left(V_1,V_2\right)$, $\Pr_{V_1,*}\left[T\left(V_1,V_2\right)\right]$ and $\Pr^*_{V_1}\left[T\left(V_1,V_2\right)\right]$ are the inner and outer probabilities respectively over V_1, taking V_2 as fixed. Analogous notation applies to inner and outer expectations.

Lemma 2.1. *For* $y \geq \sqrt{\frac{8}{n}}\cdot\sup_{f\in\mathcal{F}}\sqrt{\frac{1}{n}\sum_{i=1}^{n}\operatorname{Var}\left[f\left(X_i\right)\right]}$,

$$\Pr^*\left\{\left\|\mathbb{P}-\overline{P}\right\|_{\mathcal{F}}>y\right\} \leq 4\Pr^*\left\{\left\|\mathbb{P}^\circ\right\|_{\mathcal{F}}>\frac{y}{4}\right\}. \tag{2.3}$$

Proof of Lemma 2.1. Lemma 2.1 is an application of Lemma B.3, which is a slightly generalized variant of Lemma 2.3.7 in van der Vaart and Wellner (2023, p. 176). For details of the application here, see Section A.1. □

Lemma 2.2. *Suppose $\mathcal{F}$ satisfies $\kappa_2^2\left(\mathbb{P},\mathcal{F}\right)<\infty$ almost surely (e.g. $\mathcal{F}$ has an envelope in $\mathscr{L}_2(P)$). For $t>0$,*

$$\Pr^*_{\mathbf{U}}\left\{\left\|\mathbb{P}^\circ\right\|_{\mathcal{F}}>t\right\} \leq \min\left\{1, 2\exp\left(-\frac{nt^2}{8\kappa_2^2(\mathbb{P},\mathcal{F})}\right)N\left(t/2,\mathscr{L}_1(\mathbb{P}),\mathcal{F}\right)\right\}. \tag{2.4}$$

Proof of Lemma 2.2. See Section A.2. □

Lemma 2.3. *Let $\mathcal{F}$ satisfy $\sup_{f\in\mathcal{F},x\in\mathcal{X}}|f(x)|\le 1$. For $t>0$ such that $t\ge\kappa_2\left(\overline{P},\mathcal{F}\right)$,*

$$\Pr{}^*\left\{\sup_{f\in\mathcal{F}}\sqrt{\mathbb{P}\left[f^2\right]}>8t\right\}\le \mathrm{E}\left[\min\left\{1,4N\left(t,\mathscr{L}_2(\mathbb{P}),\mathcal{F}\right)\exp\left(-nt^2\right)\right\}\right]. \tag{2.5}$$

Proof of Lemma 2.3. See Section A.3 □

We can now proceed to the proof of Theorem 2.1.

Proof of Theorem 2.1. By (2.3) in Lemma 2.1, for $x\ge\frac{1}{\sqrt{8n}}\sup_{f\in\mathcal{F}}\sqrt{\frac{1}{n}\sum_{i=1}^n\mathrm{Var}\left[f\left(X_i\right)\right]}$ and $t>0$,

$$\begin{aligned}\Pr{}^*\left\{\left\|\mathbb{P}-\overline{P}\right\|_{\mathcal{F}}>8x\right\}\le 4\Pr{}^*\left\{\|\mathbb{P}^\circ\|_{\mathcal{F}}>2x\right\}&\le 4\Pr{}^*\left\{\|\mathbb{P}^\circ\|_{\mathcal{F}}>2x,\kappa_2^2(\mathbb{P},\mathcal{F})\le 64t^2\right\}\\&+4\Pr{}^*\left\{\|\mathbb{P}^\circ\|_{\mathcal{F}}>2x,\kappa_2^2(\mathbb{P},\mathcal{F})>64t^2\right\}.\end{aligned}$$

Hence,

$$\Pr{}^*\left\{\left\|\mathbb{P}-\overline{P}\right\|_{\mathcal{F}}>8x\right\}\le 4\left(\Pr{}^*\left\{\|\mathbb{P}^\circ\|_{\mathcal{F}}>2x,\kappa_2^2(\mathbb{P},\mathcal{F})\le 64t^2\right\}+\Pr{}^*\left\{\kappa_2(\mathbb{P},\mathcal{F})>8t\right\}\right). \tag{2.6}$$

By (2.4) in Lemma 2.2,

$$\Pr{}^*_{\mathbf{U}}\left\{\|\mathbb{P}^\circ\|_{\mathcal{F}}>2x,\kappa_2^2(\mathbb{P},\mathcal{F})\le 64t^2\right\}\le 2\exp\left(-\frac{nx^2}{128t^2}\right)N\left(x,\mathscr{L}_1(\mathbb{P}),\mathcal{F}\right). \tag{2.7}$$

Then (2.1) follows from combining (2.7) with (2.6) and using (2.5) in Lemma 2.3. □

3 A crude use of Lemma 2.2

From Lemma 2.2, we can immediately derive a crude probability bound for $\|\mathbb{P}-\overline{P}\|_{\mathcal{F}}$ when $\mathcal{F}$ is uniformly bounded, e.g. if $\sup_{f\in\mathcal{F},x\in\mathcal{X}}|f(x)|\le 1$.

Lemma 3.1. *Let $\mathcal{F}$ satisfy $\sup_{f\in\mathcal{F},x\in\mathcal{X}}|f(x)|\le 1$. For $t>0$,*

$$\Pr{}^*_{\mathbf{U}}\left\{\|\mathbb{P}^\circ\|_{\mathcal{F}}>t\right\}\le\min\left\{1,2\exp\left(-nt^2/8\right)N\left(t/2,\mathscr{L}_1(\mathbb{P}),\mathcal{F}\right)\right\}. \tag{3.1}$$

Therefore,

$$\Pr{}^*\left\{\|\mathbb{P}^\circ\|_{\mathcal{F}}>t\right\}\le\mathrm{E}\left[\min\left\{1,2\exp\left(-nt^2/8\right)N\left(t/2,\mathscr{L}_1(\mathbb{P}),\mathcal{F}\right)\right\}\right]. \tag{3.2}$$

Proof of Lemma 3.1. Note that (3.2) follows from (3.1). Furthermore, $\kappa_2^2(\mathbb{P}, \mathcal{F}) \leq 1$ since functions in $\mathcal{F}$ are uniformly bounded in magnitude by 1. Thus (3.1) follows from (2.4) in Lemma 2.2. □

We lose some generality by directly using the uniform sup norm bound. In particular, the bounds (3.1) and (3.2) are far from tight if $\mathcal{F}$ has "small" second moments, i.e. if $\kappa_2(\mathbb{P}, \mathcal{F})$ and $\kappa_2(\overline{P}, \mathcal{F})$ are small. In contrast, Theorem 2.1 accounts for cases of "small" and "large" $\kappa_2(\mathbb{P}, \mathcal{F})$ and $\kappa_2\left(\overline{P}, \mathcal{F}\right)$.

4 Implications for VC classes

We will use the following result on covering numbers for Vapnik-Červonenkis (VC) classes of functions.

Lemma 4.1 (van der Vaart and Wellner (2023), Theorem 2.6.7, p. 206)**.** *Let $\mathcal{F}$ be a class of functions with a positive envelope F with VC-dimension $V(\mathcal{F}) < \infty$. There is a universal constant $K \in (0, \infty)$ such that for any $r \in [1, \infty)$, any probability measure Q with $\|F\|_{Q,r} > 0$ and any $0 < \varepsilon < 1$,*

$$N\left(\varepsilon \|F\|_{Q,r}, \mathscr{L}_r(Q), \mathcal{F}\right) \leq K \cdot V(\mathcal{F}) (16e)^{V(\mathcal{F})} \left(\frac{1}{\varepsilon}\right)^{rV(\mathcal{F})}. \tag{4.1}$$

Theorem 4.1. *Let $\mathcal{F}$ be a VC class of functions with VC dimension $V(\mathcal{F}) < \infty$ and $x, t > 0$ satisfying*

$$\sup_{f \in \mathcal{F}, y \in \mathcal{X}} |f(y)| \leq 1, \quad x \geq \frac{1}{\sqrt{8n}} \sup_{f \in \mathcal{F}} \sqrt{\frac{1}{n} \sum_{i=1}^{n} \mathrm{Var}\left[f\left(X_i\right)\right]} \quad \textit{and} \quad t \geq \kappa_2\left(\overline{P}, \mathcal{F}\right).$$

Then for a universal constant $K \in (0, \infty)$ not depending on $\mathcal{F}$, $\{P_i\}_{i=1}^n$ or n,

$$\begin{aligned} &\Pr{}^*\left\{\left\|\mathbb{P} - \overline{P}\right\|_{\mathcal{F}} > 8x\right\} \\ &\leq 16K \cdot V(\mathcal{F})(16e)^{V(\mathcal{F})} \left[\exp\left\{-\frac{nx^2}{128t^2} + V(\mathcal{F}) \log(1/x)\right\} + \exp\left\{-nt^2 + 2V(\mathcal{F}) \log(1/t)\right\}\right]. \end{aligned} \tag{4.2}$$

Proof of Theorem 4.1. By (2.1) in Theorem 2.1,

$$\begin{aligned} \Pr{}^*\left\{\left\|\mathbb{P} - \overline{P}\right\|_{\mathcal{F}} > 8x\right\} \leq & \, 8 \exp\left(-\frac{nx^2}{128t^2}\right) \mathrm{E}\left[N\left(x, \mathscr{L}_1(\mathbb{P}), \mathcal{F}\right)\right] \\ & + 16 \exp\left(-nt^2\right) \mathrm{E}\left[N\left(t, \mathscr{L}_2(\mathbb{P}), \mathcal{F}\right)\right]. \end{aligned}$$

By (4.1) in Lemma 4.1,

$$\Pr^* \left\{ \left\| \mathbb{P} - \overline{P} \right\|_{\mathcal{F}} > 8x \right\} \leq 8KV(\mathcal{F})(16e)^{V(\mathcal{F})} \exp\left(-\frac{nx^2}{128t^2} + V(\mathcal{F}) \log(1/x) \right) + 16KV(\mathcal{F})(16e)^{V(\mathcal{F})} \exp\left(-nt^2 + 2V(\mathcal{F}) \log(1/t) \right).$$

Now bound 8 by 16 to get (4.2). □

References

Pollard, David. 1984. *Convergence of Stochastic Processes.* Springer Series in Statistics. Springer New York.

van der Vaart, Aad, and Jon Wellner. 2023. *Weak Convergence and Empirical Processes: With Applications to Statistics.* 2nd ed. Springer Series in Statistics. Springer New York.

A Proofs of results in Section 2

A.1 Proof of Lemma 2.1

Define $Z_i(f) = (f(X_i) - \mathrm{E}[f(X_i)])/n$. By Lemma B.3, the lower bound on y justifies the use of $\beta = 1/2$ and $\alpha = y/2$ in Lemma B.2; whence (2.3) follows from an application of (B.3) in Lemma B.2. □

A.2 Proof of Lemma 2.2

Treat $\mathbf{X}$ as fixed. The trivial upper bound of 1 follows from the fact that $\Pr^*_{\mathbf{U}}$ is an outer probability. It will be useful to note the following: since $U_i \in \{-1, 1\}$,

$$\forall g \in \mathcal{F}, \quad |\mathbb{P}^\circ g| = \left| \frac{1}{n} \sum_{i=1}^n U_i g(X_i) \right| \leq \frac{1}{n} \sum_{i=1}^n |g(X_i)| = \mathbb{P}|g|. \tag{A.1}$$

Let $N = N(t/2, \mathscr{L}_1(\mathbb{P}), \mathcal{F})$ and select $\mathcal{F}_* = \{g_1, \ldots, g_N\} \subseteq \mathcal{F}$ such that

$$\text{defining } g_{*,f} := \operatorname*{argmin}_{g \in \mathcal{F}_*} \mathbb{P}|f - g|, \text{ we have } \sup_{f \in \mathcal{F}} \mathbb{P}|f - g_{*,f}| \leq t/2. \tag{A.2}$$

Combining (A.1) and (A.2),

$$\|\mathbb{P}^\circ\|_{\mathcal{F}} = \sup_{f\in\mathcal{F}} |\mathbb{P}^\circ f| \leq \max_{j=1,\ldots,N} |\mathbb{P}^\circ g_j| + \sup_{f\in\mathcal{F}} \mathbb{P}\,|f - g_{*,f}| \leq \max_{j=1,\ldots,N} |\mathbb{P}^\circ g_j| + (t/2).$$

Thus $\mathbf{1}\{\|\mathbb{P}^\circ\|_{\mathcal{F}} > t\} \leq \mathbf{1}\left\{\max_{j=1,\ldots,N} |\mathbb{P}^\circ g_j| > \frac{t}{2}\right\}$ whence

$$\Pr^*_{\mathbf{U}}\{\|\mathbb{P}^\circ\|_{\mathcal{F}} > t\} \leq \Pr_{\mathbf{U}}\left\{\max_{j=1,\ldots,N} |\mathbb{P}^\circ g_j| > t/2\right\} \leq N \max_{j=1,\ldots,N} \Pr_{\mathbf{U}}\{|\mathbb{P}^\circ g_j| > t/2\}.$$

By (C.1) in Lemma C.2, since $\mathbb{P}g_j^2 \leq \kappa_2^2(\mathbb{P},\mathcal{F})$, it follows that for each $j = 1,\ldots,N$,

$$\Pr_{\mathbf{U}}\{|\mathbb{P}^\circ g_j| > t/2\} \leq 2\exp\left(-\frac{nt^2}{8\cdot\mathbb{P}g_j^2}\right) \leq 2\exp\left(-\frac{nt^2}{8\cdot\kappa_2^2(\mathbb{P},\mathcal{F})}\right).$$

Now (2.4) follows by combining the two previous displayed equations:

$$\Pr^*_{\mathbf{U}}\{\|\mathbb{P}^\circ\|_{\mathcal{F}} > t\} \leq 2N\exp\left(-\frac{nt^2}{8\cdot\kappa_2^2(\mathbb{P},\mathcal{F})}\right).$$

□

A.3 Proof of Lemma 2.3

A.3.1 Setup and auxiliary results

In this proof, let $\mathbf{W} = (W_1,\ldots,W_n)$ be an independent copy of $\mathbf{X} = (X_1,\ldots,X_n)$, and let

$$\mathbb{Q} \equiv \mathbb{Q}_n := \frac{1}{n}\sum_{i=1}^n \delta_{\{W_i\}} \quad \text{and} \quad \mathbb{S} \equiv \mathbb{S}_n := \frac{1}{2}(\mathbb{P}_n + \mathbb{Q}_n) \equiv \frac{1}{2}(\mathbb{P} + \mathbb{Q}). \tag{A.3}$$

We might expect that our previous symmetrization approach to be useful here. However the square root in the definition of κ_2 complicates symmetrization by Rademacher variables; thus a carefully constructed alternative treatment of "symmetrization" is required. We proceed via a randomized selection process: let $\mathbf{T} = (T_1,\ldots,T_n)$ be a vector of iid Bernoulli(1/2) variables satisfying $\mathbf{T} \perp\!\!\!\perp$

$(\mathbf{X}, \mathbf{W})$, and define

$$\begin{aligned}
&\xi_i = T_i X_i + (1 - T_i) W_i \quad \text{and} \quad \omega_i = T_i W_i + (1 - T_i) X_i, \\
&\mathbb{P}' = \frac{1}{n} \sum_{i=1}^{n} \delta_{\{\xi_i\}} \quad \text{and} \quad \mathbb{Q}' = \frac{1}{n} \sum_{i=1}^{n} \delta_{\{\omega_i\}}.
\end{aligned} \tag{A.4}$$

Notice that $(\mathbb{P}, \mathbb{Q}) \sim (\mathbb{P}', \mathbb{Q}')$ and that $\mathbb{S} \equiv \frac{1}{2}(\mathbb{P}' + \mathbb{Q}')$.

Lemma A.1. *For every $t > 0$ such that $t \geq \kappa_2\left(\overline{P}, \mathcal{F}\right)$,*

$$\Pr^* \left\{ \sup_{f \in \mathcal{F}} \left| \sqrt{\mathbb{P}\left[f^2\right]} \right| > 8t \right\} \leq \frac{4}{3} \Pr^* \left\{ \sup_{f \in \mathcal{F}} \left| \sqrt{\mathbb{P}'\left[f^2\right]} - \sqrt{\mathbb{Q}'\left[f^2\right]} \right| > 6t \right\}. \tag{A.5}$$

Proof of Lemma A.1. See Section A.3.3. □

Lemma A.2. *Let f be a measurable real function such that $\sup_{x \in \mathcal{X}} |f(x)| \leq 1$. For any $t > 0$,*

$$\Pr_{\mathbf{T}} \left\{ \left| \sqrt{\mathbb{P}'\left[f^2\right]} - \sqrt{\mathbb{Q}'\left[f^2\right]} \right| > t \right\} \leq 2 \exp\left(-nt^2/2\right). \tag{A.6}$$

Proof of Lemma A.2. See Section A.3.4. □

Lemma A.3. *Let $\mathcal{F}$ satisfy $\sup_{f \in \mathcal{F}, x \in \mathcal{X}} |f(x)| \leq 1$. For $t > 0$ such that $t \geq \kappa_2\left(\overline{P}, \mathcal{F}\right)$,*

$$\begin{aligned}
\Pr_{\mathbf{T}}^* \left\{ \sup_{f \in \mathcal{F}} \left| \sqrt{\mathbb{P}'\left[f^2\right]} - \sqrt{\mathbb{Q}'\left[f^2\right]} \right| > 6t \right\} &\leq \min\left\{1, 2N\left(t, \mathscr{L}_2(\mathbb{P}), \mathcal{F}\right) \exp\left(-nt^2\right)\right\} \\
&\quad + \min\left\{1, N\left(t, \mathscr{L}_2(\mathbb{Q}), \mathcal{F}\right) \exp\left(-nt^2\right)\right\},
\end{aligned} \tag{A.7}$$

and so,

$$\Pr^* \left\{ \sup_{f \in \mathcal{F}} \left| \sqrt{\mathbb{P}'\left[f^2\right]} - \sqrt{\mathbb{Q}'\left[f^2\right]} \right| > 6t \right\} \leq \mathrm{E}\left[\min\left\{1, 3N\left(t, \mathscr{L}_2(\mathbb{P}), \mathcal{F}\right) \exp\left(-nt^2\right)\right\}\right]. \tag{A.8}$$

Proof of Lemma A.3. See Section A.3.5. □

A.3.2 Proof of Lemma 2.3

(2.5) follows directly from combining (A.5) in Lemma A.1 with (A.8) in Lemma A.3. □

A.3.3 Proof of Lemma A.1

First, note that by $\mathbb{P} \sim \mathbb{P}'$,

$$\Pr^*\left\{\sup_{f\in\mathcal{F}}\left|\sqrt{\mathbb{P}\left[f^2\right]}\right| > 8t\right\} = \Pr^*\left\{\sup_{f\in\mathcal{F}}\left|\sqrt{\mathbb{P}'\left[f^2\right]}\right| > 8t\right\}$$

Define $Z_1(f) = \sqrt{\mathbb{P}'\left[f^2\right]}$ and $Y_1(f) = \sqrt{\mathbb{Q}'\left[f^2\right]}$. We will apply (B.1) in Lemma B.2 with $m = 1$, $y = 8t$, $\alpha = 2t$ and $\beta = 3/4$. For $t > 0$,

$$\sup_{f\in\mathcal{F}} \Pr\left\{|Y_1(f)| > 2t\right\} \leq \sup_{f\in\mathcal{F}} \frac{\mathrm{E}\left[Y_1^2(f)\right]}{4t} = \sup_{f\in\mathcal{F}} \frac{\mathrm{E}\left[\mathbb{Q}'\left[f^2\right]\right]}{4t^2} = \sup_{f\in\mathcal{F}} \frac{\overline{P}\left[f^2\right]}{4t^2} = \frac{\kappa_2^2\left(\overline{P},\mathcal{F}\right)}{4t^2}.$$

Thus if additionally $t \geq \kappa_2\left(\overline{P},\mathcal{F}\right)$, it follows that $\sup_{f\in\mathcal{F}} \Pr\left\{|Y_1(f)| > 2t\right\} \leq 1/4$, which is equivalent to $\inf_{f\in\mathcal{F}} \Pr\left\{|Y_1(f)| \leq 2t\right\} \geq 3/4 = \beta$. Now, (A.5) follows directly from (B.1) in Lemma B.2 with $m = 1$, $y = 8t$, $\alpha = 2t$ and $\beta = 3/4$.

□

A.3.4 Proof of Lemma A.2

Note that

$$\left|\sqrt{\mathbb{P}'\left[f^2\right]} - \sqrt{\mathbb{Q}'\left[f^2\right]}\right| = \frac{\left|\mathbb{P}'\left[f^2\right] - \mathbb{Q}'\left[f^2\right]\right|}{\sqrt{\mathbb{P}'\left[f^2\right]} + \sqrt{\mathbb{Q}'\left[f^2\right]}}.$$

Furthermore,

$$\left(\sqrt{\mathbb{P}'\left[f^2\right]} + \sqrt{\mathbb{Q}'\left[f^2\right]}\right)^2 = \mathbb{P}'\left[f^2\right] + \mathbb{Q}'\left[f^2\right] + 2\sqrt{\mathbb{P}'\left[f^2\right]\mathbb{Q}'\left[f^2\right]} \geq \mathbb{P}'\left[f^2\right] + \mathbb{Q}'\left[f^2\right] = 2\mathbb{S}\left[f^2\right].$$

Combine the previous two displays:

$$\left|\sqrt{\mathbb{P}'\left[f^2\right]} - \sqrt{\mathbb{Q}'\left[f^2\right]}\right| \leq \frac{\left|\mathbb{P}'\left[f^2\right] - \mathbb{Q}'\left[f^2\right]\right|}{\sqrt{2\mathbb{S}\left[f^2\right]}}.$$

Then,

$$\Pr_{\mathbf{T}}\left\{\left|\sqrt{\mathbb{P}'\left[f^2\right]} - \sqrt{\mathbb{Q}'\left[f^2\right]}\right| > t\right\} \leq \Pr_{\mathbf{T}}\left\{\left|\mathbb{P}'\left[f^2\right] - \mathbb{Q}'\left[f^2\right]\right| > t\sqrt{2\mathbb{S}\left[f^2\right]}\right\}$$

Now, we can express $\mathbb{P}'\left[f^2\right] - \mathbb{Q}'\left[f^2\right]$ as a Rademacher symmetrization. Define $U_i = 2T_i - 1$, so that $U_i \sim$ Rademacher. Then,

$$\begin{aligned}\mathbb{P}'\left[f^2\right] - \mathbb{Q}'\left[f^2\right] &= \frac{1}{n}\sum_{i=1}^{n}\left[T_i f^2\left(X_i\right) + \left(1-T_i\right) f^2\left(W_i\right) - T_i f^2\left(W_i\right) - \left(1-T_i\right) f^2\left(X_i\right)\right] \\ &= \frac{1}{n}\sum_{i=1}^{n} U_i\left[f^2\left(X_i\right) - f^2\left(W_i\right)\right].\end{aligned}$$

Combine these last two displays:

$$\Pr_{\mathbf{T}}\left\{\left|\sqrt{\mathbb{P}'\left[f^2\right]} - \sqrt{\mathbb{Q}'\left[f^2\right]}\right| > t\right\} \le \Pr_{\mathbf{T}}\left\{\left|\sum_{i=1}^{n} U_i\left[f^2\left(X_i\right) - f^2\left(W_i\right)\right]\right| > nt\sqrt{2\mathbb{S}\left[f^2\right]}\right\}.$$

By Lemma C.2

$$\Pr_{\mathbf{T}}\left\{\left|\sqrt{\mathbb{P}'\left[f^2\right]} - \sqrt{\mathbb{Q}'\left[f^2\right]}\right| > t\right\} \le 2\exp\left(-\frac{n^2 t^2 \mathbb{S}\left[f^2\right]}{\sum_{i=1}^{n}\left[f^2\left(X_i\right) - f^2\left(W_i\right)\right]^2}\right).$$

Now (A.6) follows since

$$\begin{aligned}\sum_{i=1}^{n}\left[f^2\left(X_i\right) - f^2\left(W_i\right)\right]^2 &= \sum_{i=1}^{n}\left[f^4\left(X_i\right) + f^4\left(W_i\right) - 2f^2\left(X_i\right)f^2\left(W_i\right)\right] \\ &\le \sum_{i=1}^{n}\left[f^4\left(X_i\right) + f^4\left(W_i\right)\right] = 2n\mathbb{S}\left[f^4\right] \le 2n\mathbb{S}\left[f^2\right], \text{ by } |f| \le 1.\end{aligned}$$

□

A.3.5 Proof of Lemma A.3

Let $M = N\left(\sqrt{2}t, \mathscr{L}_2(\mathbb{S}), \mathcal{F}\right)$ and select $\mathcal{F}_* = \{g_1, \ldots, g_M\} \subseteq \mathcal{F}$ such that

$$\text{defining } g_{*,f} := \operatorname*{argmin}_{g\in\mathcal{F}_*} \mathbb{S}\left[(f-g)^2\right], \text{ we have } \sup_{f\in\mathcal{F}}\sqrt{\mathbb{S}\left[\left(f-g_{*,f}\right)^2\right]} \le \sqrt{2}t. \tag{A.9}$$

By the triangle inequality and (A.9), setting $g = g_{*,f}$ for brevity of notation,

$$\left|\sqrt{\mathbb{P}'\left[f^2\right]} - \sqrt{\mathbb{P}'\left[g^2\right]}\right| \le \sqrt{\mathbb{P}'\left[(f-g)^2\right]} \le \sqrt{2\mathbb{S}\left[(f-g)^2\right]} \le 2t.$$

Since the same bound holds for $\mathbb{Q}'$,

$$\begin{aligned}\left|\sqrt{\mathbb{P}'\left[f^2\right]}-\sqrt{\mathbb{Q}'\left[f^2\right]}\right| &\leq \left|\sqrt{\mathbb{P}'\left[f^2\right]}-\sqrt{\mathbb{P}'\left[g^2\right]}\right|+\left|\sqrt{\mathbb{Q}'\left[f^2\right]}-\sqrt{\mathbb{Q}'\left[g^2\right]}\right|+\left|\sqrt{\mathbb{P}'\left[g^2\right]}-\sqrt{\mathbb{Q}'\left[g^2\right]}\right| \\ &\leq 4t+\left|\sqrt{\mathbb{P}'\left[g^2\right]}-\sqrt{\mathbb{Q}'\left[g^2\right]}\right|.\end{aligned}$$

Therefore,

$$\begin{aligned}\Pr^*_{\mathbf{T}}\left\{\sup_{f\in\mathcal{F}}\left|\sqrt{\mathbb{P}'\left[f^2\right]}-\sqrt{\mathbb{Q}'\left[f^2\right]}\right|>6t\right\} &\leq \Pr^*_{\mathbf{T}}\left\{\sup_{f\in\mathcal{F}}\left|\sqrt{\mathbb{P}'\left[g^2_{*,f}\right]}-\sqrt{\mathbb{Q}'\left[g^2_{*,f}\right]}\right|>2t\right\} \\ &= \Pr^*_{\mathbf{T}}\left\{\max_{j=1,\ldots,M}\left|\sqrt{\mathbb{P}'\left[g^2_j\right]}-\sqrt{\mathbb{Q}'\left[g^2_j\right]}\right|>2t\right\} \\ &\leq M\max_{j=1,\ldots,M}\Pr^*_{\mathbf{T}}\left\{\left|\sqrt{\mathbb{P}'\left[g^2_j\right]}-\sqrt{\mathbb{Q}'\left[g^2_j\right]}\right|>2t\right\}\end{aligned}$$

By Lemma A.2,

$$\Pr^*_{\mathbf{T}}\left\{\sup_{f\in\mathcal{F}}\left|\sqrt{\mathbb{P}'\left[f^2\right]}-\sqrt{\mathbb{Q}'\left[f^2\right]}\right|>6t\right\}\leq\min\left\{1,2N\left(\sqrt{2}t,\mathscr{L}_2(\mathbb{S}),\mathcal{F}\right)\exp\left(-2nt^2\right)\right\}.$$

By Lemma C.6,

$$\Pr^*_{\mathbf{T}}\left\{\sup_{f\in\mathcal{F}}\left|\sqrt{\mathbb{P}'\left[f^2\right]}-\sqrt{\mathbb{Q}'\left[f^2\right]}\right|>6t\right\}\leq\min\left\{1,2N\left(t,\mathscr{L}_2(\mathbb{P}),\mathcal{F}\right)N\left(t,\mathscr{L}_2(\mathbb{Q}),\mathcal{F}\right)\exp\left(-2nt^2\right)\right\}.$$

By Lemma C.7,

$$\begin{aligned}\Pr^*_{\mathbf{T}}\left\{\sup_{f\in\mathcal{F}}\left|\sqrt{\mathbb{P}'\left[f^2\right]}-\sqrt{\mathbb{Q}'\left[f^2\right]}\right|>6t\right\} &\leq \min\left\{1,2N\left(t,\mathscr{L}_2(\mathbb{P}),\mathcal{F}\right)\exp\left(-nt^2\right)\right\} \\ &\quad+\min\left\{1,N\left(t,\mathscr{L}_2(\mathbb{Q}),\mathcal{F}\right)\exp\left(-nt^2\right)\right\}.\end{aligned}$$

This proves (A.7), and "integrating out **X** and **W**" by an application of (B.1) then proves (A.8). □

B Useful Symmetrization Results

Lemma B.1 (van der Vaart and Wellner (2023), Lemma 1.2.6, p. 10)**.** *Let X_1 and X_2 be random elements on a common probability space. For any arbitrary real-valued map T,*

$$\mathrm{E}_* \left[T\left(X_1, X_2\right)\right] \leq \mathrm{E}_{X_1,*}\mathrm{E}_{X_2,*}\left[T\left(X_1, X_2\right)\right] \leq \mathrm{E}^*_{X_1}\mathrm{E}^*_{X_2}\left[T\left(X_1, X_2\right)\right] \leq \mathrm{E}^*\left[T\left(X_1, X_2\right)\right].$$

Lemma B.2. *Let $\mathbf{Z} = (Z_1, \ldots, Z_m)$, $\mathbf{Y} = (Y_1, \ldots, Y_m)$, and $\mathbf{U} = (U_1, \ldots, U_m)$ be mutually independent random elements defined on a common probability space. Let $\mu_1, \ldots, \mu_n : \mathcal{F} \to \mathbb{R}$ be arbitrary maps. Assume $\mathbf{U}$ takes values in $\{-1, 1\}^m$, that the Z_i's are mutually independent real-valued stochastic processes indexed by a set $\mathcal{F}$ and that $\mathbf{Y}$ is an independent copy of $\mathbf{Z}$. For $\alpha > 0$, denote $\beta_m(\alpha) = \inf_{f\in\mathcal{F}} \Pr\{|\sum_{i=1}^m Z_i(f)| \leq \alpha\}$. For any $y > \alpha > 0$, and any $0 \leq \beta \leq \beta_m(\alpha)$,*

$$\beta \cdot \Pr^*\left\{\left\|\sum_{i=1}^m Z_i\right\|_{\mathcal{F}} > y\right\} \leq \Pr^*\left\{\left\|\sum_{i=1}^m \left(Z_i - Y_i\right)\right\|_{\mathcal{F}} > y - \alpha\right\} \tag{B.1}$$

$$\leq \Pr^*\left\{\left\|\sum_{i=1}^m U_i\left(Z_i - Y_i\right)\right\|_{\mathcal{F}} > y - \alpha\right\} \tag{B.2}$$

$$\leq 2\Pr^*\left\{\left\|\sum_{i=1}^m U_i\left(Z_i - \mu_i\right)\right\|_{\mathcal{F}} > \frac{y-\alpha}{2}\right\}. \tag{B.3}$$

Proof of Lemma B.2. We start with the first inequality in (B.1). It suffices to prove this inequality for $\beta = \beta_m(\alpha)$. On $\{\|\sum_{i=1}^m Z_i\|_{\mathcal{F}} > y\}$, we can select $\widehat{f}(\mathbf{Z})$ such that $\left|\sum_{i=1}^m Z_i\left(\widehat{f}(\mathbf{Z})\right)\right| > y$; let $\widehat{f}(\cdot)$ be equal to an arbitrary member of $\mathcal{F}$ on $\{\|\sum_{i=1}^m Z_i\|_{\mathcal{F}} \leq y\}$. Under the hypothesis $\left|\sum_{i=1}^m Y_i\left(\widehat{f}(\mathbf{Z})\right)\right| \leq \alpha$,

$$\left\|\sum_{i=1}^m \left(Z_i - Y_i\right)\right\|_{\mathcal{F}} \geq \left|\sum_{i=1}^m \left(Z_i\left(\widehat{f}(\mathbf{Z})\right) - Y_i\left(\widehat{f}(\mathbf{Z})\right)\right)\right| \geq \left|\sum_{i=1}^m Z_i\left(\widehat{f}(\mathbf{Z})\right)\right| - \left|\sum_{i=1}^m Y_i\left(\widehat{f}(\mathbf{Z})\right)\right| > y - \alpha.$$

Therefore,

$$\mathbf{1}\left\{\left\|\sum_{i=1}^m Z_i\right\|_{\mathcal{F}} > y\right\} \cdot \mathbf{1}\left\{\left|\sum_{i=1}^m Y_i\left(\widehat{f}(\mathbf{Z})\right)\right| \leq \alpha\right\} \leq \mathbf{1}\left\{\left\|\sum_{i=1}^m \left(Z_i - Y_i\right)\right\|_{\mathcal{F}} > y - \alpha\right\}.$$

Fix $\mathbf{Z}$ and take outer expectations with respect to $\mathbf{Y}$

$$\mathbf{1}\left\{\left\|\sum_{i=1}^{m} Z_i\right\|_{\mathcal{F}} > y\right\} \cdot \Pr{}_{\mathbf{Y}}^{*}\left\{\left|\sum_{i=1}^{m} Y_i\left(\widehat{f}(\mathbf{Z})\right)\right| \leq \alpha\right\} \leq \Pr{}_{\mathbf{Y}}^{*}\left\{\left\|\sum_{i=1}^{m}(Z_i - Y_i)\right\|_{\mathcal{F}} > y - \alpha\right\}. \quad \text{(B.4)}$$

Note that for fixed $\mathbf{Z}$,

$$\beta_m(\alpha) \leq \Pr{}_{\mathbf{Y}}\left\{\left|\sum_{i=1}^{m} Y_i\left(\widehat{f}(\mathbf{Z})\right)\right| \leq \alpha\right\} = \Pr{}_{\mathbf{Y}}^{*}\left\{\left|\sum_{i=1}^{m} Y_i\left(\widehat{f}(\mathbf{Z})\right)\right| \leq \alpha\right\},$$

where the last equality follows since $\sum_{i=1}^{m} Y_i(f)$ is measurable for fixed f. Combine the last display with (B.4):

$$\beta_m(\alpha) \cdot \mathbf{1}\left\{\left\|\sum_{i=1}^{m} Z_i\right\|_{\mathcal{F}} > y\right\} \leq \Pr{}_{\mathbf{Y}}^{*}\left\{\left\|\sum_{i=1}^{m}(Z_i - Y_i)\right\|_{\mathcal{F}} > y - \alpha\right\}.$$

Now the first inequality in (B.1) follows from integrating out $\mathbf{Z}$ and using Lemma B.1:

$$\begin{aligned}\beta_m(\alpha) \cdot \Pr{}_{\mathbf{Z}}^{*}\left\{\left\|\sum_{i=1}^{m} Z_i\right\|_{\mathcal{F}} > y\right\} &\leq \mathrm{E}_{\mathbf{Z}}^{*}\left[\Pr{}_{\mathbf{Y}}^{*}\left\{\left\|\sum_{i=1}^{m}(Z_i - Y_i)\right\|_{\mathcal{F}} > y - \alpha\right\}\right] \\ &\leq \Pr{}^{*}\left\{\left\|\sum_{i=1}^{m}(Z_i - Y_i)\right\|_{\mathcal{F}} > y - \alpha\right\}.\end{aligned}$$

For (B.2), first note that for each value of $\mathbf{U}$, $\sum_{i=1}^{m} U_i\left(Z_i - Y_i\right)$ and $\sum_{i=1}^{m}\left(Z_i - Y_i\right)$ have the same distribution since $\mathbf{Z}$ and $\mathbf{Y}$ are independent copies of each other, the Z_i's are all mutually independent, and the U_i's are all either -1 or 1. Thus for $\mathbf{U}$ held fixed, integrating out $(\mathbf{Z}, \mathbf{Y})$ yields

$$\Pr{}_{(\mathbf{Z},\mathbf{Y})}^{*}\left\{\left\|\sum_{i=1}^{m} U_i\left(Z_i - Y_i\right)\right\|_{\mathcal{F}} > y - \alpha\right\} = \Pr{}^{*}\left\{\left\|\sum_{i=1}^{m}\left(Z_i - Y_i\right)\right\|_{\mathcal{F}} > y - \alpha\right\}.$$

Hence, (B.2) follows from integrating out $\mathbf{U}$ in the above and using Lemma B.1:

$$\begin{aligned}\Pr{}^{*}\left\{\left\|\sum_{i=1}^{m} U_i\left(Z_i - Y_i\right)\right\|_{\mathcal{F}} > y - \alpha\right\} &\geq \mathrm{E}_{\mathbf{U}}\left[\Pr{}_{(\mathbf{Z},\mathbf{Y})}^{*}\left\{\left\|\sum_{i=1}^{m} U_i\left(Z_i - Y_i\right)\right\|_{\mathcal{F}} > y - \alpha\right\}\right] \\ &= \Pr{}^{*}\left\{\left\|\sum_{i=1}^{m}\left(Z_i - Y_i\right)\right\|_{\mathcal{F}} > y - \alpha\right\}.\end{aligned}$$

For (B.3), by the triangle inequality

$$\left\| \sum_{i=1}^{m} U_i \left(Z_i - Y_i \right) \right\|_{\mathcal{F}} \leq \left\| \sum_{i=1}^{m} U_i \left(Z_i - \mu_i \right) \right\|_{\mathcal{F}} + \left\| \sum_{i=1}^{m} U_i \left(Y_i - \mu_i \right) \right\|_{\mathcal{F}} .$$

Now (B.3) follows from the union bound:

$$\begin{aligned}
\Pr{}^* \left\{ \left\| \sum_{i=1}^{m} U_i \left(Z_i - Y_i \right) \right\|_{\mathcal{F}} > y - \alpha \right\} &\leq \Pr{}^* \left\{ \left\| \sum_{i=1}^{m} U_i \left(Z_i - \mu_i \right) \right\|_{\mathcal{F}} + \left\| \sum_{i=1}^{m} U_i \left(Y_i - \mu_i \right) \right\|_{\mathcal{F}} > y - \alpha \right\} \\
&\leq \Pr{}^* \left\{ \left\| \sum_{i=1}^{m} U_i \left(Z_i - \mu_i \right) \right\|_{\mathcal{F}} > \frac{y - \alpha}{2} \right\} \\
&\quad + \Pr{}^* \left\{ \left\| \sum_{i=1}^{m} U_i \left(Y_i - \mu_i \right) \right\|_{\mathcal{F}} > \frac{y - \alpha}{2} \right\} \\
&= 2\Pr{}^* \left\{ \left\| \sum_{i=1}^{m} U_i \left(Z_i - \mu_i \right) \right\|_{\mathcal{F}} > \frac{y - \alpha}{2} \right\} .
\end{aligned}$$

□

Lemma B.3. *Let the conditions in the premise of Lemma B.2 hold and in addition, assume that* $Z_1, \ldots, Z_n$ *are independent mean zero and finite variance processes indexed by* $\mathcal{F}$*. Let* $\beta \in (0, 1)$ *and* $y > \alpha > \left(\sup_{f \in \mathcal{F}} \sqrt{\sum_{i=1}^{m} \mathrm{Var}\left[Z_i(f) \right]} \right) / \sqrt{1 - \beta}$*. Then* (B.1)*,* (B.2) *and* (B.3) *all hold.*

Proof of Lemma B.3. We are done if we show $\beta \leq \beta_m(\alpha) := \inf_{f \in \mathcal{F}} \Pr \left\{ \| \sum_{i=1}^{m} Z_i(f) | \leq \alpha \right\}$ or equivalently that $1 - \beta_m(\alpha) \leq 1 - \beta$. To that end, by $\alpha > \left(\sup_{f \in \mathcal{F}} \sqrt{\sum_{i=1}^{m} \mathrm{Var}\left[Z_i(f) \right]} \right) / \sqrt{1 - \beta}$, and Chebyshev's inequality,

$$\begin{aligned}
1 - \beta > \frac{\sup_{f \in \mathcal{F}} \sum_{i=1}^{m} \mathrm{Var}\left[Z_i(f) \right]}{\alpha^2} &\geq \sup_{f \in \mathcal{F}} \Pr \left\{ \left\| \sum_{i=1}^{m} Z_i(f) \right| > \alpha \right\} \\
&= 1 - \inf_{f \in \mathcal{F}} \Pr \left\{ \left\| \sum_{i=1}^{m} Z_i(f) \right| \leq \alpha \right\} = 1 - \beta_m(\alpha).
\end{aligned}$$

□

C Useful Miscellany

Lemma C.1 (Hoeffding's Inequality)**.** *Let $V_1, \dots, V_n$ be independent random variables such that $a_i \le V_i \le b_i$ almost surely for some $a_i, b_i \in \mathbb{R}$. For all $t > 0$,*

$$\Pr\left(\left|\sum_{i=1}^{n}\left(V_i - \mathrm{E}\left[V_i\right]\right)\right| \ge t\right) \le 2\exp\left(-\frac{2t^2}{\sum_{i=1}^{n}\left(b_i - a_i\right)^2}\right).$$

Lemma C.2 (Hoeffding's Inequality Special Case)**.** *Let $U_1, \dots, U_n$ be independent mean zero random variables all supported in $[-1,1]$. For fixed non-stochastic numbers $g_1, \dots, g_n$ and any $t > 0$,*

$$\Pr\left\{\left|\frac{1}{n}\sum_{i=1}^{n} U_i g_i\right| > t\right\} \le 2\exp\left(-\frac{n^2 t^2}{2\sum_{i=1}^{n} g_i^2}\right) \le 2\exp\left(-\frac{n t^2}{2\max_{i=1,\dots,n} g_i^2}\right) \tag{C.1}$$

Proof of Lemmas C.1 and C.2. Various standard references have proofs of Hoeffding's Inequality. For example, the reader can look at Theorem 2 and its proof in Pollard (1984, Appendix B, pp. 191-192). We therefore focus on Lemma C.2. The second inequality in (C.1) is a straightforward consequence of the first inequality in (C.1). The first inequality in (C.1) follows directly from Hoeffding's Inequality, i.e. Lemma C.1, by defining $V_i = U_i g_i$, $a_i = -g_i$ and $b_i = g_i$ and noting that $\mathrm{E}[V_i] = 0$ by $\mathrm{E}[U_i] = 0$. □

Lemma C.3. *Let Q be a measure, A be a measurable set for which $QA > 0$. If $Q[A \cap \{u \le 0\}] = 0$ (i.e. on A, u is positive Q-a.e.), then $Q[u \cdot \mathbf{1}_A] > 0$.*

Proof of Lemma C.3. Suppose for contradiction that both $Q[A \cap \{u \le 0\}] = 0$, and $Q[u \cdot \mathbf{1}_A] \le 0$. By the hypothesis $Q[A \cap \{u \le 0\}] = 0$, $Q[u \cdot \mathbf{1}_A] = Q\left[u \cdot \mathbf{1}_{A \cap \{u > 0\}}\right]$. By the latter condition, $Q[A \cap \{u > 0\}] = 0$. But then $QA = Q[A \cap \{u \le 0\}] + Q[A \cap \{u > 0\}] = 0 + 0 = 0$, which is the desired contradiction □

Lemma C.4. *For measures Q_1, Q_2, let $Q = Q_1 + Q_2$ and $h_\ell = \mathrm{d}Q_\ell / \mathrm{d}Q$. Furthermore, let $A = \{h_1 < 1/2$ and $h_2 < 1/2\}$. Then $QA = 0$.*

Proof of Lemma C.4. For the purpose of contradiction, suppose $QA > 0$. Then let $u_\ell = (1/2) - h_\ell$. Clearly $Q[A \cap \{u_\ell \le 0\}] = 0$. By Lemma C.3, $Q[u_\ell \cdot \mathbf{1}_A] > 0$ and so $Q[h_\ell \cdot \mathbf{1}_A] < \frac{1}{2}QA$. But $Q[h_\ell \cdot \mathbf{1}_A] = Q_\ell A$ by definition of h_ℓ. Combined, $Q_\ell A < \frac{1}{2}QA$ for both $\ell = 1, 2$. But this implies $QA = Q_1 A + Q_2 A < \frac{1}{2}QA + \frac{1}{2}QA = QA$, which is a contradiction. □

Lemma C.5. *Let* Q_1, Q_2 *be measures,* $Q = Q_1 + Q_2$*,* $h_\ell = \mathrm{d}Q_\ell/\mathrm{d}Q$ *and* $A_\ell = \{h_\ell \geq 1/2\}$*. Let* $f \in \mathscr{L}_2(Q_1) \cap \mathscr{L}_2(Q_2)$ *and for* $\delta > 0$*, suppose* $\varphi_\ell \in \mathscr{L}_2(Q_\ell)$ *satisfy* $Q_\ell\left[|f - \varphi_\ell|^2\right] \leq \delta^2$*. Define* $g = \varphi_1 \cdot \mathbf{1}_{A_1} + \varphi_2 \cdot (1 - \mathbf{1}_{A_1}) \cdot \mathbf{1}_{A_2}$*. Then* $Q\left[|f - g|^2\right] \leq 4\delta^2$*.*

Proof of Lemma C.5. Let $A = A_1^c \cap A_2^c = \{h_1 < 1/2 \text{ and } h_2 < 1/2\}$. By Lemma C.4, $QA = 0$. Then $f = f \cdot \mathbf{1}_{A_1} + f \cdot (1 - \mathbf{1}_{A_1}) \cdot \mathbf{1}_{A_2} + f \cdot \mathbf{1}_A$ and so

$$f - g = (f - \varphi_1) \cdot \mathbf{1}_{A_1} + (f - \varphi_2) \cdot (1 - \mathbf{1}_{A_1}) \cdot \mathbf{1}_{A_2} + f \cdot \mathbf{1}_A.$$

Note that $(1 - \mathbf{1}_{A_1}) \cdot \mathbf{1}_{A_2} = \mathbf{1}_{A_2 \setminus A_1}$. Furthermore A_1, $A_2 \setminus A_1$ and A form a measurable partition of the underlying measurable space. Therefore,

$$|f - g|^2 = |f - \varphi_1|^2 \cdot \mathbf{1}_{A_1} + |f - \varphi_2|^2 \cdot (1 - \mathbf{1}_{A_1}) \cdot \mathbf{1}_{A_2} + f^2 \cdot \mathbf{1}_A.$$

Therefore by $QA = 0$,

$$Q\left[|f - g|^2\right] = Q\left[|f - \varphi_1|^2 \cdot \mathbf{1}_{A_1}\right] + Q\left[|f - \varphi_2|^2 \cdot (1 - \mathbf{1}_{A_1}) \cdot \mathbf{1}_{A_2}\right].$$

By definition of A_ℓ, $\mathbf{1}_{A_\ell} \leq 2h_\ell \cdot \mathbf{1}_{A_\ell} \leq 2h_\ell$ (since $h_\ell \geq 0$ Q-a.e.). Since $0 \leq 1 - \mathbf{1}_{A_1} \leq 1$ everywhere,

$$Q\left[|f - g|^2\right] \leq 2\sum_{\ell=1}^{2} Q\left[h_\ell |f - \varphi_\ell|^2\right] = 2\sum_{\ell=1}^{2} Q_\ell\left[|f - \varphi_\ell|^2\right] \leq 4\delta^2.$$

□

Lemma C.6. *Let* Q_1, Q_2 *be measures and* $\mathcal{F} \subseteq \mathscr{L}_2(Q_1) \cap \mathscr{L}_2(Q_2)$*. For any* $\delta > 0$*,*

$$N\left(\sqrt{2}\delta, \mathscr{L}_2\left((Q_1 + Q_2)/2\right), \mathcal{F}\right) \leq N\left(\delta, \mathscr{L}_2(Q_1), \mathcal{F}\right) \cdot N\left(\delta, \mathscr{L}_2(Q_2), \mathcal{F}\right).$$

Proof of Lemma C.6. Let $N_\ell = N(\delta, \mathscr{L}_2(Q_\ell), \mathcal{F})$ and let $\{\varphi_{\ell,j}\}_{j=1}^{N_\ell} \subseteq \mathcal{F}$ be a δ-cover for $\mathcal{F}$ in $\mathscr{L}_2(Q_\ell)$. Now follow Lemma C.5: define $Q = Q_1 + Q_2$, $h_\ell = \mathrm{d}Q_\ell/\mathrm{d}Q$ and $A_\ell = \{h_\ell \geq 1/2\}$. Furthermore, define approximating functions $g_{ij} = \varphi_{1,i} \cdot \mathbf{1}_{A_1} + \varphi_{2,j} \cdot (1 - \mathbf{1}_{A_1}) \cdot \mathbf{1}_{A_2}$. The set $\{g_{ij}\}_{i,j=1}^{N_1,N_2}$ is of size $N_1 \cdot N_2$. Fix $f \in \mathcal{F}$, and i, j such that $Q_1\left[|f - \varphi_{1,i}|^2\right] \leq \delta^2$ and $Q_2\left[|f - \varphi_{2,j}|^2\right] \leq \delta^2$. By

Lemma C.5,

$$Q\left[|f-g_{ij}|^2\right] \le 4\delta^2, \quad \text{whence} \quad \left(\frac{Q_1+Q_2}{2}\right)\left[|f-g_{ij}|^2\right] \le 2\delta^2$$

$$\text{so that} \quad \left\{\left(\frac{Q_1+Q_2}{2}\right)\left[|f-g_{ij}|^2\right]\right\}^{1/2} \le \sqrt{2}\delta.$$

Hence, $N\left(\sqrt{2}\delta, \mathscr{L}_2\left(\left(Q_1+Q_2\right)/2\right), \mathcal{F}\right) \le N\left(\delta, \mathscr{L}_2\left(Q_1\right), \mathcal{F}\right) \cdot N\left(\delta, \mathscr{L}_2\left(Q_2\right), \mathcal{F}\right)$ as desired. □

Lemma C.7. *For $x, y \ge 0$, $\min\{1, x \cdot y\} \le \min\{1, x\} + \min\{1, y\}$.*

Proof of Lemma C.7. Break into three cases.

Case 1: $x, y \le 1$. Then $\min\{1, x \cdot y\} = x \cdot y \le x \le x + y = \min\{1, x\} + \min\{1, y\}$.

Case 2: $x, y > 1$. Then $\min\{1, x \cdot y\} = 1 < 2 = \min\{1, x\} + \min\{1 + y\}$.

Case 3: $x \le 1 < y$. Then $\min\{1, x \cdot y\} \le 1 \le x + 1 = \min\{1, x\} + \min\{1, y\}$. □