Chapter 5: Symmetrization and First Concentration Results

Empirical Processes (MATH-522)

Myrto Limnios

April, 2025

1 Symmetric Measure and Properties

1.1 Definitions

Consider two independent i.i.d. samples $X_1, \dots X_n$ and $X'_1, \dots, X'_n$ defined on the same p.s. $(\Omega, \mathcal{A}, \mathbb{P})$ and valued in a measurable space $\mathcal{X} \subseteq \mathbb{R}^d$, $d \in \mathbb{N}^*$, of same probability distribution P. We define the *empirical measure* by

$$P_n := \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i} \ ,$$

where $\delta_X(A)$ is the Dirac measure of the event $\{X \in A\}$. This chapter focuses on providing a control of the supnorm of averages

$$Z_h = \frac{1}{n} \sum_{i=1}^{n} h(X_i) =: P_n h \ ,$$

indexed by a class of measurable functions $\mathcal{H}$, that is an empirical estimator of $Ph := \int_{\mathcal{X}} h(x) P(\mathrm{d}x) = \mathbb{E}h(X)$. Until now, we have supposed the process Z_h to be centered to avoid some technicalities. However, one would like to quantify the deviation of an estimator to its mean value $\mathbb{E}h(X)$ that is unknown in practice. We thus want to control in probability and expectation

$$\|P_n - P\|_{\mathcal{H}} = \sup_{h \in \mathcal{H}} \left| \frac{1}{n} \sum_{i=1}^{n} h(X_i) - \mathbb{E}h(X) \right| \ ,$$

using the **symmetrization method** that will allow to replace the *true mean* by an estimated value based on an independent i.i.d. sample $X'_1, \dots, X'_n$ and given by

$$Z'_h = (1/n) \sum_i h(X'_i) =: P'_n h \ ,$$

where

$$P'_n := \frac{1}{n} \sum_{i=1}^{n} \delta_{X'_i} \ .$$

Thus, we would like to approximate $\|P_n - P\|_{\mathcal{H}}$ by $\|P_n - P'_n\|_{\mathcal{H}}$.

Lemma 1.1. *Let a class $\mathcal{H}$ of measurable functions $h : \mathcal{X} \to \mathbb{R}$, and consider two independent and i.i.d. samples $X_1, \dots X_n$ and $X'_1, \dots, X'_n$ defined on the same p.s. $(\Omega, \mathcal{A}, \mathbb{P})$ and valued in a measurable space $\mathcal{X} \subseteq \mathbb{R}^d$, $d \in \mathbb{N}^*$, and of probability distribution P. Then,*

$$\mathbb{E}\|P_n - P\|_{\mathcal{H}} \leq \mathbb{E}\|P_n - P'_n\|_{\mathcal{H}} \ .$$

Proof. Notice that $Ph = \mathbb{E}[P'_n h] == \mathbb{E}[P'_n h | X_1, \dots X_n]$ for all $h \in \mathcal{H}$, by independence of the X's and X''s. Thus

$$\mathbb{E}\|P_n - P\|_{\mathcal{H}} = \mathbb{E}\left[\sup_{h\in\mathcal{H}} |P_n - P|\right] = \mathbb{E}\left[\sup_{h\in\mathcal{H}} |P_n - \mathbb{E}[P'_n]|\right] = \mathbb{E}\left[\sup_{h\in\mathcal{H}} |\mathbb{E}[P_n - P'_n | X_1, \dots X_n]|\right] .$$

Then, by Jensen's inequality

$$\mathbb{E}\|P_n - P\|_{\mathcal{H}} \leq \mathbb{E}\mathbb{E}\left[\sup_{h\in\mathcal{H}} |P_n - P'_n||X_1, \dots X_n\right] = \mathbb{E}\left[\sup_{h\in\mathcal{H}} |P_n - P'_n|\right]$$

by the law of total expectation that concludes the proof. □

This should remind you of train/test procedures in machine learning.

An important class of symmetrized processes is obtained by considering an auxiliary set of i.i.d. *Rademacher* r.v.s $\varepsilon_1, \dots, \varepsilon_n$, with $\mathbb{P}(\varepsilon = 1) = \mathbb{P}(\varepsilon = -1) = 1/2$.

Definition 1.2. Let an i.i.d. sequence of symmetric r.r.v. $\varepsilon_1, \dots, \varepsilon_n$ independent of $X_1, \dots X_n$. The *symmetrized empirical measure* based on the sample $X_1, \dots X_n$ is defined by the measurable map

$$P_n^0 : h \in \mathcal{H} \mapsto P_n^0 h = \frac{1}{n}\sum_{i=1}^{n} \varepsilon_i h(X_i) . \tag{1}$$

In particular, if the ε's are Rademacher, then P_n^0 defines the Rademacher empirical measure.

Reminder 1.1. *The intuition of considering such a process lies in the interpretation of the ε's as a vector of noise r.v.s. If the class $\mathcal{H}$ is too large, then we can always find a functions that is highly correlated with random noise, thus yielding in a large valued for the sup.*

Remark 1.2.

1. *In the following, we will only refer to P_n^0 as the Rademacher empirical measure, if not stated otherwise.*
2. *Again, we ignore measurability considerations for the suprema involved in the numerous formulations.*

We will now see why we will use for all the proofs results such as Lemma 1.1. to go from suprema of $P_n - P$ to those based on P_n^0: it is symmetric conditionally on the X's, so that the increment condition (sub-Gaussian) and zero mean for the chaining method are fulfilled, fundamental to next Chapter.

1.2 Symmetrization in expectation

Lemma 1.3. *Suppose conditions of Definition 1.2. Consider a nondecreasing and convex function $\Phi : \mathbb{R} \to \mathbb{R}$, then*

$$\mathbb{E}\Phi\left(\|P_n - P\|_{\mathcal{H}}\right) \leq \mathbb{E}\Phi\left(2\|P_n^0\|_{\mathcal{H}}\right) .$$

Proof. Let an independent sample $X'_1, \dots X'_n$, i.i.d. drawn from P, of symmetrized empirical measure

$$P_n^{'0} h = \frac{1}{n}\sum_{i=1}^{n} \varepsilon_i h(X'_i) .$$

Then, because $\|P_n - P'_n\|_{\mathcal{H}}$ has same distribution as $\|P_n^0 - P_n^{'0}\|_{\mathcal{H}}$ (notice this by conditioning on the X, X's and because the measures are based on the same ε's), by Lemma 1.1, and because Φ is nondecreasing

$$\mathbb{E}\Phi\left(\|P_n - P\|_{\mathcal{H}}\right) \leq \mathbb{E}\Phi\left(\|P_n - P'_n\|_{\mathcal{H}}\right) = \mathbb{E}\Phi\left(\|P_n^0 - P_n^{'0}\|_{\mathcal{H}}\right)$$

then, notice that by convexity using Jensen's inequality

$$\mathbb{E}\Phi\left(\|P_n^0 - P_n^{'0}\|_{\mathcal{H}}\right) \leq \mathbb{E}\Phi\left(\|P_n^0\|_{\mathcal{H}} + \|P_n^{'0}\|_{\mathcal{H}}\right) = \mathbb{E}\Phi\left(2\|P_n^0\|_{\mathcal{H}}\right) \ .$$

□

Remark 1.3. *Notice the similarity of with the results in Chapter 4 for generic Rademacher complexitis: we can view the symmetrization as a contraction property.*

We now generalize this result to any sequence of independent coordinate process $\{Z_{1,h}, \ldots, Z_{n,h}\}$ indexed by $h \in \mathcal{H}$, that is measurable, in the sense of marginal measurability w.r.t. i. If the coordinates are i.i.d., then the empirical measure is obtained by taking $Z_{i,h} = h(X_i) - Ph$. Importantly, it is interesting to *symmetrize* a process if one can then *de-symmetrize* it, that is to say, that the suprema of both measures are comparable.

Theorem 1.4. *Let a general independent centered process $\{Z_{1,h}, \ldots, Z_{n,h}\}_{h\in\mathcal{H}}$, and an independent i.i.d. sample of symmetric r.r.v. $\varepsilon_1, \ldots, \varepsilon_n$. Then, for any nondecreasing and convex function $\Phi : \mathbb{R} \to \mathbb{R}$,*

$$\mathbb{E}\Phi\left(\frac{1}{2}\left\|\sum_{i=1}^n \varepsilon_i Z_i\right\|_{\mathcal{H}}\right) \leq \mathbb{E}\Phi\left(\left\|\sum_{i=1}^n Z_i\right\|_{\mathcal{H}}\right) \leq \mathbb{E}\Phi\left(2\left\|\sum_{i=1}^n \varepsilon_i (Z_i - \mu_i)\right\|_{\mathcal{H}}\right) \ ,$$

where $\mu_i : \mathcal{H} \to \mathbb{R}$, for all $i \leq n$, are arbitrary functions.

Corollary 1.5. *Consider a set of i.i.d. r.v.s $X_1, \ldots X_n$, and an independent i.i.d. sequence of Rademacher r.v.s $\varepsilon_1, \ldots, \varepsilon_n$. Consider the class of measurable functions $\mathcal{H}$. Then, for any nondecreasing and convex function $\Phi : \mathbb{R} \to \mathbb{R}$,*

$$\mathbb{E}\Phi\left(\frac{1}{2}\sup_{h\in\mathcal{H}}\left|\sum_{i=1}^n \varepsilon_i (h(X_i) - Ph)\right|\right) \leq \mathbb{E}\Phi\left(\|P_n - P\|_{\mathcal{H}}\right) \leq \mathbb{E}\Phi\left(2\sup_{h\in\mathcal{H}}\left|\sum_{i=1}^n \varepsilon_i h(X_i)\right|\right) \ .$$

Proof of Corollary. The upperbound follows from Lemma 1.3. The lower bound is straightforward by using convexity and the definition of the supremum with the Rademacher r.v.s.. Precisely

$$\mathbb{E}\Phi\left(\frac{1}{2}\sup_{h\in\mathcal{H}}\left|\sum_{i=1}^n \varepsilon_i (h(X_i) - Ph)\right|\right) \leq \mathbb{E}\Phi\left(\frac{1}{2}\|P_n - P_n'\|_{\mathcal{H}}\right) \ .$$

Then, by adding and substracting the mean Ph inside the supnorm, the triangular inequality yields

$$\|P_n - P_n'\|_{\mathcal{H}} \leq \|P_n - P\|_{\mathcal{H}} + \|P_n' - P\|_{\mathcal{H}}$$

because Φ is nondecreasing

$$\Phi(\frac{1}{2}\|P_n - P_n'\|_{\mathcal{H}}) \leq \Phi(\frac{1}{2}\|P_n - P\|_{\mathcal{H}} + \frac{1}{2}\|P_n' - P\|_{\mathcal{H}})$$

and the convexity with the fact that the X and X's are identically distributed concludes the proof . □

Remark 1.4. *Corollary 1.5 will be used with specific choices of Φ, such as $\Phi : t \mapsto t$ and $\Phi : t \mapsto e^{\lambda t}$, for $\lambda > 0$.*

1.3 Symmetrization in probability

We state now a simple yet very useful lemma, that will be proved during the exercise session.

Lemma 1.6. *Let $\mathcal{H}$ be a class of measurable functions for any $h \in \mathcal{H}$, for any $u > 0$*

$$\mathbb{P}(|(P_n - P)(h)| > u/2) \leq \frac{1}{2} \ . \tag{2}$$

Then, for any $u > 0$

$$\mathbb{P}(\|P_n - P\|_{\mathcal{H}} > u) \leq 4\mathbb{P}\left(\|P_n^0\|_{\mathcal{H}} > \frac{u}{4}\right) \ .$$

Proof. Let an i.i.d. sample $X_1', \ldots, X_n'$ independent of the X's. Let $u > 0$. On the event $\{\|P_n - P\|_{\mathcal{H}} > u\}$ there exists a random function $h^* \in \mathcal{H}$ such that $|P_n h^* - P h^*| > u$. We first prove that

$$\mathbb{P}(\|P_n - P\|_{\mathcal{H}} > u) \leq 2\mathbb{P}\left(\|P_n - P_n'\|_{\mathcal{H}} > \frac{u}{2}\right) \ .$$

Notice that because $\|P_n - P_n'\|_{\mathcal{H}} \geq |P_n h^* - P_n' h^*|$ conditionally on the X's,

$$\begin{aligned}
\mathbb{P}(\|P_n - P_n'\|_{\mathcal{H}} > u/2) &= \mathbb{E}\mathbb{P}(\|P_n - P_n'\|_{\mathcal{H}} > u/2 | X_1, \ldots, X_n) \\
&\geq \mathbb{E}\mathbb{P}(|P_n h^* - P_n' h^*| > u/2 | X_1, \ldots, X_n) \\
&= \mathbb{P}(|P_n h^* - P_n' h^*| > u/2)
\end{aligned}$$

and by the triangular inequality

$$|P_n h^* - P_n' h^*| \leq |P_n h^* - P h^*| + |P_n' h^* - P h^*|$$

$$\begin{aligned}
\mathbb{P}(\|P_n - P_n'\|_{\mathcal{H}} > u/2) &\geq \mathbb{P}(|P_n h^* - P h^*| > u, \ |P_n' h^* - P h^*| < u/2) \\
&= \mathbb{E}\mathbb{P}(|P_n h^* - P h^*| > u, \ |P_n' h^* - P h^*| < u/2 | X_1, \ldots, X_n) \\
&= \mathbb{E}\left[1\{|P_n h^* - P h^*| > u\}\mathbb{P}(|P_n' h^* - P h^*| < u/2 | X_1, \ldots, X_n)\right]
\end{aligned}$$

Now using the assumption Eq. (2)

$$\mathbb{P}(\|P_n - P_n'\|_{\mathcal{H}} > u/2) \geq \frac{1}{2}\mathbb{E}\left[1\{|P_n h^* - P h^*| > u\}\right] = \frac{1}{2}\mathbb{P}(|P_n h^* - P h^*| > u) = \frac{1}{2}\mathbb{P}(\|P_n - P\|_{\mathcal{H}} > u) \ .$$

Then, the triangular inequality on the left-hand-side yields

$$\begin{aligned}
\mathbb{P}(\|P_n - P\|_{\mathcal{H}} > u) &\leq 2\mathbb{P}(\|P_n - P_n'\|_{\mathcal{H}} > u/2) = 2\mathbb{P}(\|P_n^0 - P_n'^0\|_{\mathcal{H}} > u/2) \\
&\leq 2\mathbb{P}(\|P_n^0\|_{\mathcal{H}} > u/4) + 2\mathbb{P}(\|P_n'^0\|_{\mathcal{H}} > u/4) = 4\mathbb{P}(\|P_n^0\|_{\mathcal{H}} > u/4) \ .
\end{aligned}$$

□

Remark 1.5. *If $\mathcal{H}$ is composed of functions uniformly bounded by 1, then the pointwise condition Eq. (2) is fulfilled as proved below.*
We next lower bound the tail probability using Chebychev's Inequality

$$\mathbb{P}(|P_n' h^* - P h^*| \leq u/2 | X_1, \ldots, X_n) \geq 1 - \frac{\mathrm{Var}(P_n' h^* | X_1, \ldots, X_n)}{u^2/4} \ .$$

The process $P_n' h^*$ conditionally on the X's, is invariant in permutations of the indices $i \leq n$, and recalling that the functions are uniformly bounded by 1, it fulfills the bounded differences condition, thus

$$\mathrm{Var}(P_n' h^* | X_1, \ldots, X_n) \leq \frac{\sum (1/n)^2}{4} = \frac{1}{4n} \ ,$$

and plugging it in the previous inequality yields

$$\mathbb{P}(|P_n' h^* - P h^*| \leq u/2 | X_1, \ldots, X_n) \geq 1 - \frac{1}{nu^2} \geq \frac{1}{2} \ .$$

because $nu^2 \geq 2$.

2 Concentration Inequalities

We prove that if the supremum of the Rademacher empirical process is of order $o(1)$, then the index class of measurable functions $\mathcal{H}$ is Glivenko-Cantelli. This Theorem illustrates how the Rademacher empirical process measures the size of a process. *Recall that the RC for a fixed function $h \in \mathcal{H}$ is intuitively the correlation between the vector $(h(X_1), \dots, h(X_n))$ and the 'noise' vector $(\varepsilon_1, \dots, \varepsilon_n)$.*

Theorem 2.1. *Suppose the class $\mathcal{H}$ to be composed of measurable functions $h : \mathcal{X} \to \mathbb{R}$ uniformly bounded by a finite constant $K > 0$, i.e, $\|h\|_\infty \leq K$. Then, for any $u > 0$,*

$$\mathbb{P}(\|P_n - P\|_{\mathcal{H}} \geq 2\mathbb{E}\|P_n^0\|_{\mathcal{H}} + u) \leq e^{-nu^2/(2K^2)} \ .$$

If it holds true, and that $\mathcal{R}_n(\mathcal{H}) = o(1)$, then $\mathcal{H}$ is P-Glivenko-Cantelli:

$$\|P_n - P\|_{\mathcal{H}} \overset{a.s.}{\rightsquigarrow} 0 \ .$$

Proof. We first prove the tail bound.

$$\|P_n - P\|_{\mathcal{H}} = \sup_{h \in \mathcal{H}} |P_n h - Ph| =: Z(X_1, \dots, X_n) \ .$$

We show how to use McDiarmid's inequality applied to Z. Notice that because the process Z is invariant on permutations on the coordinates, we check the bounded differences inequality only through one of the coordinates, e.g., X_1. Let X_1' and independent copy of X_1, then, for any $g \in \mathcal{H}$,

$$\begin{aligned} |\sum_{i=1}^n g(X_i) - Pg| - nZ(X_1', \dots, X_n) &= |g(X_1) + \sum_{i=2}^n g(X_i) - nPg| - \sup_{h \in \mathcal{H}} |h(X_1') + \sum_{i=2}^n h(X_i) - nPh| \\ &\leq |g(X_1) + \sum_{i=2}^n g(X_i) - nPg| - |g(X_1') + \sum_{i=2}^n g(X_i) - nPg| \leq |g(X_1) - g(X_1')| \leq 2K \ , \end{aligned}$$

by the triangular inequality and using the uniform boundness of the functions in $\mathcal{H}$. Then, taking the sup on the left-hand side proves that Z satisfies the bounded differences condition:

$$Z(X_1, \dots, X_n) - Z(X_1', \dots, X_n) \leq \frac{2K}{n} \ .$$

McDiarmid's inequality yields, for any $u \geq 0$,

$$\mathbb{P}(Z(X_1, \dots, X_n) - \mathbb{E}Z(X_1, \dots, X_n) \geq u) \leq e^{-nt^2/(2K^2)} \ .$$

That says with probability at least $1 - \delta$, for $\delta \in (0, 1)$,

$$Z(X_1, \dots, X_n) \leq \mathbb{E}Z(X_1, \dots, X_n) + K\sqrt{\frac{2\log(1/\delta)}{n}} \ .$$

We then use the symmetrization argument on the expectation, and in particular Lemma 1.3 that concludes the proof. □

Remark 2.1. *Notice that we can reformulate the concentration inequality as follows. Let $\delta \in (0, 1)$, then $\delta = e^{-nu^2/(2K^2)}$ iff. $u = K\sqrt{2\log(1/\delta)/n}$. Then with probability at least $1 - \delta$*

$$\|P_n - P\|_{\mathcal{H}} \leq 2\mathbb{E}\|P_n^0\|_{\mathcal{H}} + K\sqrt{\frac{2\log(1/\delta)}{n}} \ . \tag{3}$$

Some remarks are important. This holds true as soon as $\delta \in (0, 1)$, i.e., $n \geq K\sqrt{2\log(1/\delta)}$: the higher the probability of Eq. (3) holds true, the bigger the sample size should be. If $\mathbb{E}\|P_n^0\|_{\mathcal{H}}$ is at least of similar rate as $O_{\mathbb{P}}(n^{-1/2})$, then Eq. (3) recovers the rates of UCLT. This is especially interesting if it is sharp, i.e. constant K is the smallest as possible, and it is very not informative if $K \to \infty$, as it is equivalent to the asymptotic regime (because $n \to \infty$ then).

Notice that without the symmetrization argument, McDiarmid's inequality shows that, to control the probability of uniform deviation of Z, then one needs to control its mean value first. The next Chapter will focus on providing maximal inequalities for index classes fulfilling various assumptions.

We end with an important result when the class $\mathcal{H}$ is of VC-type, with finite dimension, then we can prove the Vapnik-Chervonenkis theorem, by upperbounding the Rademacher complexity by a function of the VC-dimension.

Theorem 2.2 (Vapnik-Chervonenkis Theorem)**.** *Let an i.i.d. sample $X_1, \ldots, X_n$, valued in $\mathbb{R}^d$, of probability distribution P. Then for any class of Borelian sets $\mathcal{C} \subset \mathbb{R}^d$, it holds true that for any $u > 0$,*

$$\mathbb{P}(\|P_n - P\|_{\mathcal{C}} > u) \leq 8 m_n(\mathcal{C}) e^{-nu^2/32} ,$$

where $m_n(\mathcal{C})$ is the shattering coefficient of the class $\mathcal{C}$.

Proof. Step 1: Symmetrization. To use Lemma 1.6, notice that we have a similar result than that of Remark 1.5. Precisely, notice that for any $C \in \mathcal{C}$,

$$P_n(C) - P(C) = \frac{1}{n} \sum_{i=1}^{n} (1_C(X_i) - P(C))$$

thus $nP_n(C)$ is a Binomial r.v. $\mathcal{B}(n, P(C))$ and thus has variance given by

$$\mathrm{Var}(P_n' h^* | X_1, \ldots, X_n) = \frac{P(C)(1 - P(C))}{n} \leq \frac{1}{4n} ,$$

because $\sup_{x \in [0,1]} x(1-x) = 1/4$. Thus we can apply Lemma 1.6, and we obtain for all $u > 0$,

$$\mathbb{P}(\|P_n - P\|_{\mathcal{C}} > u) \leq 4 \mathbb{P}\left(\|P_n^0\|_{\mathcal{C}} > \frac{u}{4} \right) . \tag{4}$$

Step 2: Approximation of the supnorm.

Notice that

$$\mathbb{P}\left(\|P_n^0\|_{\mathcal{C}} > \frac{u}{4} \right) = \mathbb{E}\mathbb{P}\left(\|P_n^0\|_{\mathcal{C}} > \frac{u}{4} \, | X_1, \ldots, X_n \right)$$

If we consider the points $X_1, \ldots, X_n$ to be fixed, then, the image vector $1_C(X_1), \ldots, 1_C(X_n)$ take $|\{\{X_1, \ldots, X_n\} \cap C, \ C \in \mathcal{C}\}|$ distinct values when letting $C \in \mathcal{C}$, that is bounded from above by the shattering coefficient given by $m_n(\mathcal{C})$. Thus, we can consider a finite subset $\mathcal{C}_0$ depending on the dataset, and of size at most $m_n(\mathcal{C})$ such that

$$\begin{aligned}
\mathbb{P}\left(\|P_n^0\|_{\mathcal{C}} > \frac{u}{4} | X_1, \ldots, X_n \right) &= \mathbb{P}\left(\|P_n^0\|_{\mathcal{C}} > \frac{u}{4} \, | X_1, \ldots, X_n \right) \\
&\leq \mathbb{P}\left(\exists C \in \mathcal{C}_0, \ |P_n^0(C)| > \frac{u}{4} \, | X_1, \ldots, X_n \right) \\
&\leq \sum_{C \in \mathcal{C}_0} \mathbb{P}\left(|P_n^0(C)| > \frac{u}{4} \, | X_1, \ldots, X_n \right) \\
&\leq m_n(\mathcal{C}) \sup_{C \in \mathcal{C}} \mathbb{P}\left(|P_n^0(C)| > \frac{u}{4} \, | X_1, \ldots, X_n \right) .
\end{aligned}$$

Thus

$$\mathbb{P}(\|P_n - P\|_{\mathcal{C}} > u) \leq 4 m_n(\mathcal{C}) \mathbb{E}\left[\sup_{C \in \mathcal{C}} \mathbb{P}\left(|P_n^0(C)| > \frac{u}{4} \, | X_1, \ldots, X_n \right) \right] \tag{5}$$

Step 3: Concentration inequality. Conditionally on the X's, the average $\sum_i \varepsilon_i 1_C(X_i)$ is only a sum of n independent r.v.s, centered, and valued in $[-1,1]$. Thus, Hoeffding's inequality applies

$$\mathbb{P}\left(|P_n^0(C)| > \frac{u}{4} \,|X_1,\ldots,X_n\right) = \mathbb{P}\left(|n\sum_i \varepsilon_i 1_C(X_i)| > \frac{nu}{4} \,|X_1,\ldots,X_n\right) \leq 2e^{-nu^2/32} \ .$$

and thus

$$\mathbb{P}(\|P_n - P\|_{\mathcal{C}} > u) \leq 8m_n(\mathcal{C})e^{-nu^2/32} \ . \tag{6}$$

□

We will see in the next Chapter how to obtain similar bound for general uncountable classes of functions.