



**HAL**  
open science

# PAC-Bayes with Unbounded Losses through Supermartingales

Maxime Haddouche, Benjamin Guedj

► **To cite this version:**

Maxime Haddouche, Benjamin Guedj. PAC-Bayes with Unbounded Losses through Supermartingales. 2022. hal-03815101v1

**HAL Id: hal-03815101**

**<https://inria.hal.science/hal-03815101v1>**

Preprint submitted on 14 Oct 2022 (v1), last revised 27 Apr 2023 (v3)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# PAC-Bayes with Unbounded Losses through Supermartingales

**Maxime Haddouche**

*Inria*

MAXIME.HADDOUCHE@INRIA.FR

**Benjamin Guedj**

*Inria & University College London*

BENJAMIN.GUEDJ@INRIA.FR

## Abstract

While PAC-Bayes is now an established learning framework for bounded losses, its extension to the case of unbounded losses (as simple as the squared loss on an unbounded space) remains largely uncharted and has attracted a growing interest in recent years. We contribute to this line of work by developing an extension of Markov’s inequality for supermartingales, which we use to establish a novel PAC-Bayesian generalisation bound holding for unbounded losses. We show that this bound extends, unifies and even improves on existing PAC-Bayesian bounds.

**Keywords:** PAC-Bayes; supermartingales; unbounded losses; generalisation bounds.

## 1. Introduction

PAC-Bayes learning is a branch of statistical learning theory aiming to produce (tight) generalisation guarantees for learning algorithms, and as a byproduct leads to designing new efficient learning procedures by minimising such guarantees. Generalisation bounds are helpful for understanding how a learning algorithm may perform on future similar batches of data. Since its emergence in the late 1990s, PAC-Bayes theory (see the seminal works of [Shawe-Taylor and Williamson, 1997](#); [McAllester, 1998, 1999, 2003](#); [Catoni, 2003, 2007](#)) has revealed powerful enough to explain the generalisation abilities of learning algorithms which output distributions over the predictors space (a particular case being when the output is a Dirac mass on a single predictor) from which our predictors of interest are designed. We refer to the recent surveys from [Guedj \(2019\)](#); [Alquier \(2021\)](#) for an overview on PAC-Bayes.

At first, PAC-Bayes theory was mainly focused on classification tasks (see e.g. [Seeger, 2002](#) or [Langford and Schapire, 2005](#); [Catoni, 2007](#)) but has quickly been generalised to any bounded loss function in regression (see e.g., [Maurer, 2004](#); [Germain et al., 2009, 2016](#)). PAC-Bayes learning covers a broad scope of domains and tools, from information theory ([Kakade et al., 2008](#); [Wu and Seldin, 2022](#)) to statistical learning ([Catoni, 2003, 2007](#)), convex optimisation ([Thiemann et al., 2017](#)), Bernstein-type concentration inequalities ([Tolstikhin and Seldin, 2013](#); [Mhammedi et al., 2019](#)), margins ([Biggs and Guedj, 2022a](#); [Biggs et al., 2022](#)) and martingales ([Seldin et al., 2011, 2012a,b](#)), to name but a few.

From a practical perspective, the above led to generalisations guarantees for PAC-Bayes-inspired neural networks (NN): a promising line of work initiated by [Dziugaite and Roy \(2017\)](#) and pursued further by [Letarte et al. \(2019\)](#); [Rivasplata et al. \(2019\)](#); [Pérez-Ortiz et al. \(2021\)](#); [Biggs and Guedj \(2021\)](#); [Perez-Ortiz et al. \(2021a,b\)](#); [Biggs and Guedj \(2022b\)](#), among others, established NN architectures enjoying tight generalisation guarantees, turning PAC-Bayes into a powerful tool to handle complex neural structures (e.g., [Chérif-Abdellatif et al., 2022](#)).

These encouraging breakthroughs gave rise to several initiatives to extend PAC-Bayes beyond the bounded losses assumption which is limiting in practice. Indeed, the goal is to make PAC-Bayes

able to provide efficiency guarantees to any learning algorithm attached to a loss function. For instance, consider a NN trained to solve regression problems without constraints on the training domain. Several works (Alquier and Guedj, 2018; Holland, 2019; Haddouche et al., 2021) already proposed routes to overcome the boundedness constraint. Our work stands in the continuation of this spirit while developing and exploiting a novel technical toolbox.

To better highlight the novelty of our approach, we first present the two classical building blocks of PAC-Bayes.

## 1.1. Understanding PAC-Bayes: a celebrated route of proof

### 1.1.1. TWO ESSENTIAL BUILDING BLOCKS

In PAC-Bayes, we typically assume access to a non-negative loss function  $\ell(h, z)$  taking as argument a predictor  $h \in \mathcal{H}$  and data  $z \in \mathcal{Z}$  (think of  $z$  as a pair input-output  $(x, y)$  for supervised learning problems, or as a single datum  $x$  in unsupervised learning). We also assume access to a  $m$ -sized sample  $S = (z_1, \dots, z_m) \in \mathcal{Z}^m$  of data on which we will learn a meaningful posterior distribution  $Q$  on  $\mathcal{H}$ , from a prior  $P$  (or reference measure – see e.g., Guedj, 2019 for a discussion on the terminology of probability distributions in PAC-Bayes).

To do so, PAC-Bayesian proofs are built upon two cornerstones. The first one is the change of measure inequality (Csiszár, 1975; Donsker and Varadhan, 1975; Dupuis and Ellis, 2011 – see also Banerjee, 2006; Guedj, 2019 for a proof).

**Lemma 1 (Change of measure inequality)** *For any measurable function  $\psi : \mathcal{H} \rightarrow \mathbb{R}$  and any distributions  $Q, P$  on  $\mathcal{H}$ :*

$$\mathbb{E}_{h \sim Q}[\psi(h)] \leq \text{KL}(Q, P) + \log(\mathbb{E}_{h \sim P}[\exp(\psi(h))])$$

with KL denoting the Kullback-Leibler divergence.

The change of measure inequality is then applied to a certain function  $f_m : \mathcal{Z}^m \times \mathcal{H} \rightarrow \mathbb{R}$  of the data and a candidate predictor: for all posteriors  $Q$ ,

$$\mathbb{E}_{h \sim Q}[f_m(S, h)] \leq \text{KL}(Q, P) + \log(\mathbb{E}_{h \sim P}[\exp(f_m(S, h))]) \quad (1)$$

To deal with the random variable  $X(S) := \mathbb{E}_{h \sim P}[\exp(f_m(S, h))]$ , our second building block is Markov's inequality:

**Lemma 2** *For any non-negative integrable random variable  $X$ , the following holds for any  $a > 0$ :*

$$\mathbb{P}(X > a) \leq \frac{\mathbb{E}[X]}{a}.$$

For  $\delta \in ]0, 1[$ , we apply Markov's inequality on  $X(S)$  with  $a = \mathbb{E}_S[X(S)]/\delta$ . Taking the complementary event gives that for any  $m$ , with probability at least  $1 - \delta$  over the sample  $S$ ,  $X(S) \leq \mathbb{E}_S[X(S)]/\delta$ , thus:

$$\mathbb{E}_{h \sim Q}[f_m(S, h)] \leq \text{KL}(Q, P) + \log(1/\delta) + \log(\mathbb{E}_{h \sim P} \mathbb{E}_S[\exp(f_m(S, h))]).$$

## 1.1.2. WHAT NEXT?

There are then several routes to conclude and obtain PAC-Bayesian generalisation bounds, all being tied to specific choices of  $f$  and the assumptions on the dataset  $S$ . For instance, the celebrated result of [McAllester \(1999\)](#), tightened by [Maurer \(2004\)](#), exploits in particular, a data-free prior, an iid assumption on  $S$  and a loss function bounded by 1 (or without loss of generality, any known positive constant). Most of the existing results stand with those assumptions (see *e.g.*, [Catoni, 2007](#); [Germain et al., 2009](#); [Guedj and Alquier, 2013](#); [Tolstikhin and Seldin, 2013](#); [Guedj and Robbiano, 2018](#); [Mhammedi et al., 2019](#); [Wu and Seldin, 2022](#)). However, many works tried to mitigate at least one of these three assumptions:

1. **Data-free priors:** with an alternative set of techniques, [Catoni \(2007\)](#) obtained bounds with localised (*i.e.* data-dependent) priors. More recently, [Lever et al. \(2010\)](#); [Parrado-Hernández et al. \(2012\)](#); [Lever et al. \(2013\)](#); [Oneto et al. \(2016\)](#); [Dziugaite and Roy \(2017\)](#); [Mhammedi et al. \(2019\)](#); [Rivasplata et al. \(2020\)](#) also obtained PAC-Bayes bound with data-dependent priors.
2. **The iid assumption on  $S$ :** the work of [Fard and Pineau \(2010\)](#) established links between reinforcement learning and PAC-Bayes theory. This naturally led to the study of PAC-Bayesian bound for martingales instead of iid data ([Seldin et al., 2011, 2012a,b](#)).
3. **Bounded loss:** The works of [Alquier and Guedj \(2018\)](#), followed by [Holland \(2019\)](#) as well as [Kuzborskij and Szepesvári \(2019\)](#); [Haddouche et al. \(2021\)](#) extended PAC-Bayes to unbounded losses under several technical assumptions.

Several questions legitimately arise.

**Can we avoid these three assumptions simultaneously?** The answer is yes: for instance the work of [Rivasplata et al. \(2020\)](#) proposed a preliminary PAC-Bayes bound holding with none of the three assumptions listed above. Building on their theorem, [Haddouche and Guedj \(2022\)](#) only exploited a bounded loss assumption to derive a PAC-Bayesian framework for online learning, requiring no assumption on data and allowing data (history in their context)-dependent priors.

**Can we obtain PAC-Bayes bounds without the change of measure inequality?** Yes, for instance [Alquier and Guedj \(2018\)](#) proposed PAC-Bayes bounds involving  $f$ -divergences and exploiting Holder’s inequality instead of Lemma 1. More recently, [Picard-Weibel and Guedj \(2022\)](#) developed a broader discussion about generalising the change of measure inequality for a wide range of  $f$ -divergences. We note also that [Germain et al. \(2009\)](#) proposed a version of the classical route of proof stated above avoiding the use of the change of measure inequality. This comes at the cost of additional technical assumptions (see [Haddouche et al., 2021](#), Theorem 1) for a statement of the theorem in a proper measure-theoretic framework).

**Can we avoid Markov’s inequality?** We mentioned above that several works avoided the change of measure inequality to obtain PAC-Bayesian bounds, but can we do the same with Markov’s inequality? The answer is yes but this is a rare breed. To the best of our knowledge, only two papers are explicitly not using Markov’s inequality: [Kakade et al. \(2008\)](#) obtained a PAC-Bayes bound using results on Rademacher complexity based on the McDiarmid concentration inequality, and [Kuzborskij and Szepesvári \(2019\)](#) exploited a concentration inequality from [De la Peña et al. \(2009\)](#), up to a technical assumptions to obtain results for unbounded losses.

## 1.2. Originality of our approach

Avoiding Markov's inequality appears challenging in PAC-Bayes. We successfully manage to remove it from the picture by using a generalisation of Markov's inequality for supermartingale: Ville's inequality (as noticed by [Doob 1939](#)).

**Lemma 3 (Ville's maximal inequality for supermartingales)** *Let  $(\mathcal{F}_t)$  be a filtration and  $(Z_t)$  a non-negative super-martingale satisfying  $Z_0 = 1$  a.s. If  $Z_t$  is adapted to  $\mathcal{F}_t$  and  $\mathbb{E}[Z_t | \mathcal{F}_{t-1}] \leq Z_{t-1}$  a.s.,  $t \geq 1$ , then, for any  $0 < \delta < 1$ , it holds*

$$\mathbb{P}(\exists T \geq 1 : Z_T > \delta^{-1}) \leq \delta.$$

**Proof** We apply the optional stopping theorem ([Durrett, 2019](#), Thm 4.8.4) with Markov's inequality defining the stopping time  $i = \inf\{t > 1 : Z_t > \delta^{-1}\}$  so that

$$\mathbb{P}(\exists t \geq 1 : Z_t > \delta^{-1}) = \mathbb{P}(Z_i > \delta^{-1}) \leq \mathbb{E}[Z_i] \delta \leq \mathbb{E}[Z_0] \delta \leq \delta.$$

■

A major interest of Ville's result is that it holds for a countable sequence of random variables simultaneously. This point is new in PAC-Bayes as it will allow us to obtain bounds holding for a countable (not necessarily finite) dataset  $S = (z_i)_{i \geq 1}$ .

**On which supermartingale do we apply Ville's bound ?** To fully exploit Lemma 3, we now take a countable dataset  $S = (z_i)_{i \geq 1} \in \mathcal{Z}^m$ . Recall that, because we use the change of measure inequality, we have to deal with the following exponential random variable appearing in Eq. (1) for any  $m \geq 1$ :

$$Z_m := \mathbb{E}_{h \sim P}[\exp(f_m(S, h))].$$

Our goal is to choose a sequence of functions  $f_m$  such that  $(Z_m)_m$  is a supermartingale. A way to do so comes from [Bercu and Touati \(2008\)](#).

**Lemma 4** *Let  $(M_m)$  be a locally square-integrable martingale with respect to the filtration  $(\mathcal{F}_m)$ . For all  $\eta \in \mathbb{R}$  and  $m \geq 0$ , one has:*

$$\mathbb{E} \left[ \exp \left( \eta \Delta M_m - \frac{\eta^2}{2} (\Delta[M]_m + \Delta \langle M \rangle_m) \right) \mid \mathcal{F}_{m-1} \right] \leq 1,$$

where  $\Delta M_m = M_m - M_{m-1}$ ,  $\Delta[M]_m = \Delta M_m^2$  and  $\Delta \langle M \rangle_m = \mathbb{E}[\Delta M_m^2 | \mathcal{F}_{m-1}]$ . We let

$$V_m(\eta) = \exp \left( \eta M_m - \frac{\eta^2}{2} ([M]_m + \langle M \rangle_m) \right).$$

Then, for all  $\eta \in \mathbb{R}$ ,  $(V_m(\eta))$  is a positive supermartingale with  $\mathbb{E}[V_m(\eta)] \leq 1$ .

In the sequel, we make suggestions to design a supermartingale (i.e., to choose a relevant  $f_m$  for any  $m$ ) without further assumption in the context of PAC-Bayes for martingales.

### 1.3. Contributions and outline

We provide a novel PAC-Bayesian bound holding for data-free priors and unbounded losses. From this, we recover PAC-Bayes bounds for unbounded losses and iid data as a significant particular case. We also instantiate our new bound in various settings of the existing literature and discuss its merits. We show that our approach, not only generalises, but also unifies (and sometimes improves on) various existing bounds. Sec. 2 contains our main contribution: a PAC-Bayes bound for unbounded martingales. Then in Sec. 3 we instantiate our results onto a proper learning theory framework with the assumption of iid data. We also discuss the implications of our result. Finally, we draw in Sec. 4 a comparison between our new results and those of [Seldin et al. \(2012a\)](#); [Alquier et al. \(2016\)](#); [Haddouche et al. \(2021\)](#). We show that adapting our bounds to the assumptions made in those papers allows to recover similar or improved bounds. Appendix A gathers more details on PAC-Bayes and we defer to Appendix B the proofs of results established in Secs. 3 and 4.

## 2. A PAC-Bayesian bound for unbounded martingales

A line of work led by [Seldin et al. \(2011, 2012a,b\)](#) provided PAC-Bayes bounds for almost surely bounded martingales. We provably extend the remits of their result to the case of unbounded losses.

**Framework** Our framework is close to the one of [Seldin et al. \(2012a\)](#): we assume having access to a countable dataset  $S = (z_i)_{i \geq 1} \in$  with no restriction on the distribution of  $S$  (in particular the  $z_i$  can depend on each others). We denote for any  $m$ ,  $S_m := (z_i)_{i=1..m}$  the restriction of  $S$  to its  $m$  first points.  $(\mathcal{F}_i)_{i \geq 0}$  is a filtration adapted to  $S$ . We denote for any  $i \in \mathbb{N}$   $\mathbb{E}_{i-1}[\cdot] := \mathbb{E}[\cdot \mid \mathcal{F}_{i-1}]$ . We also precise the space  $\mathcal{H}$  to be an index (or a hypothesis) space, possibly uncountably infinite. Let  $\{X_1(S_1, h), X_2(S_2, h), \dots : h \in \mathcal{H}\}$  be martingale difference sequences, meaning that for any  $m \geq 1, h \in \mathcal{H}$ ,  $\mathbb{E}_{m-1}[X_m(S_m, h)] = 0$ .

For any  $h \in \mathcal{H}$ , let  $M_m(h) = \sum_{i=1}^m X_i(S_i, h)$  be martingales corresponding to the martingale difference sequences and we define, as in [Bercu and Touati \(2008\)](#), the following

$$[M]_m(h) := \sum_{i=1}^m X_i(S_i, h)^2,$$

$$\langle M \rangle_m(h) = \sum_{i=1}^m \mathbb{E}_{i-1}[X_i(S_i, h)^2].$$

For a distribution  $Q$  over  $\mathcal{H}$  define weighted averages of the martingales with respect to  $Q$  as  $M_m(Q) = \mathbb{E}_{h \sim Q}[M_m(h)]$  (similar definitions hold for  $[M]_m(Q), \langle M \rangle_m(Q)$ ).

**Theorem 5** *For any data-free prior  $P \in \mathcal{M}_1^+(\mathcal{H})$ , any  $\lambda > 0$ , any collection of martingales  $(M_m(h))_{m \geq 1}$  indexed by  $h \in \mathcal{H}$ , the following holds with probability  $1 - \delta$  over the sample  $S = (z_i)_{i \in \mathbb{N}}$ , for all  $m \in \mathbb{N}/\{0\}$ ,  $Q \in \mathcal{M}_1^+(\mathcal{H})$ :*

$$|M_m(Q)| \leq \frac{\text{KL}(Q, P) + \log(2/\delta)}{\lambda} + \frac{\lambda}{2} ([M]_m(Q) + \langle M \rangle_m(Q)).$$

This theorem involves several terms. The change of measure inequality introduces the KL divergence term, the approximation term  $\log(2/\delta)$  comes from Ville's inequality (instead of Markov

in classical PAC-Bayes). Finally, the terms  $[M]_m(Q)$ ,  $\langle M \rangle_m(Q)$  come from our choice of supermartingale as suggested by [Bercu and Touati \(2008\)](#). The term  $[M]_m(Q)$  can be interpreted as an empirical variance term while  $\langle M \rangle_m(Q)$  is its theoretical counterpart. Note that  $\langle M \rangle_m(Q)$  also appears in [Seldin et al. \(2012a, Theorem 1\)](#).

We recall that this general result stands with no assumption on the martingale difference sequence  $(X_i)_{i \geq 1}$  and holds uniformly on all  $m \geq 1$ . Those two points are, to the best of our knowledge, new within the PAC-Bayes literature. We discuss in [Secs. 3 and 4](#) more concrete instantiations.

**Proof** We fix  $\eta \in \mathbb{R}$  and we consider the function  $f_m$  to be  $\forall(S, h)$ :

$$\begin{aligned} f_m(S, h) &:= \eta M_m(h) - \frac{\eta^2}{2} ([M]_m(h) + \langle M \rangle_m(h)) \\ &= \sum_{i=1}^m \eta \Delta M_i(h) - \frac{\eta^2}{2} (\Delta [M]_i(h) + \Delta \langle M \rangle_i(h)), \end{aligned}$$

where  $\Delta M_i(h) = X_i(S_i, h)$ ,  $\Delta [M]_i(h) = X_i(S_i, h)^2$ ,  $\Delta \langle M \rangle_i(h) = \mathbb{E}_{i-1} [X_i(S_i, h)^2]$ . For the sake of clarity, we dropped the dependency in  $S$  of  $M_m$ . Note that, given the definition of  $M_m$ ,  $M_m(h)$  is  $\mathcal{F}_m$  measurable for any fixed  $h$ .

Let  $P$  a fixed data-free prior, we first apply the change of measure inequality to obtain  $\forall m \in \mathbb{N}, \forall Q \in \mathcal{M}_1^+(\mathcal{H})$ :

$$\mathbb{E}_{h \sim Q} [f_m(S, h)] \leq \text{KL}(Q, P) + \log \left( \underbrace{\mathbb{E}_{h \sim P} [\exp(f_m(S, h))]}_{:= Z_m} \right),$$

with the convention  $f_0 = 0$ . We now have to show that  $(Z_m)_m$  is a supermartingale with  $Z_0 = 1$ . To do so remark that for any  $m$ , because  $P$  is data free one has the following result.

**Lemma 6** *For any data-free prior  $P$ , any  $\sigma$ -algebra  $\mathcal{F}$  belonging to the filtration  $(\mathcal{F}_i)_{i \geq 0}$ , any nonnegative function  $f$  taking as argument the sample  $S$  and a predictor  $h$ , one has almost surely:*

$$\mathbb{E} [\mathbb{E}_{h \sim P} [f(S, h)] \mid \mathcal{F}] = \mathbb{E}_{h \sim P} [\mathbb{E} [f(S, h) \mid \mathcal{F}]]$$

**Proof** Let  $A$  be a  $\mathcal{F}$ -measurable event. We want to show that

$$\mathbb{E} [\mathbb{E}_{h \sim P} [f(S, h)] \mathbb{1}_A] = \mathbb{E} [\mathbb{E}_{h \sim P} [\mathbb{E} [f(S, h) \mid \mathcal{F}]] \mathbb{1}_A],$$

where the first expectation in each term is taken over  $S$ . Note that it is possible to take this expectation thanks to the Kolomogorov's extension theorem (see *e.g.* [Tao, 2011](#), Thm 2.4.4) which ensure the existence of a probability space for the discrete-time stochastic process  $S = (z_i)_{i \geq 1}$ .

Thus, this is enough to conclude that

$$\mathbb{E} [\mathbb{E}_{h \sim P} [f(S, h)] \mid \mathcal{F}] = \mathbb{E}_{h \sim P} [\mathbb{E} [f(S, h) \mid \mathcal{F}]],$$

by definition of the conditional expectation. To do so, notice that because  $f(S, h) \mathbb{1}_A$  is a nonnegative function, and that  $P$  is data-free, we can apply the classical Fubini-Tonelli theorem.

$$\mathbb{E} [\mathbb{E}_{h \sim P} [f(S, h)] \mathbb{1}_A] = \mathbb{E}_{h \sim P} [\mathbb{E} [f(S, h) \mathbb{1}_A]]$$

One now conditions by  $\mathcal{F}$  and use the fact that  $\mathbb{1}_A$  is  $\mathcal{F}$ -measurable:

$$= \mathbb{E}_{h \sim P} [\mathbb{E} [\mathbb{E} [f(S, h) \mid \mathcal{F}] \mathbb{1}_A]]$$

We finally re-apply Fubini-Tonelli to re-intervert the expectations:

$$= \mathbb{E} [\mathbb{E}_{h \sim P} [\mathbb{E} [f(S, h) \mid \mathcal{F}] \mathbb{1}_A]].$$

This concludes the proof of Lemma 6. ■

We then use Lemma 6 with  $f = \exp(f_m)$  and  $\mathcal{F} = \mathcal{F}_{m-1}$  to obtain:

$$\begin{aligned} \mathbb{E}_{m-1}[Z_m] &= \mathbb{E}_{h \sim P} [\mathbb{E}_{m-1}[(\exp(f_m(S, h)))] \\ &= \mathbb{E}_{h \sim P} \left[ \exp(f_{m-1}(S, h)) \mathbb{E}_{m-1} \left[ \exp(\eta \Delta M_m(h) - \frac{\eta^2}{2} (\Delta[M]_m(h) + \Delta \langle M \rangle_m(h))) \right] \right], \end{aligned}$$

with  $f_{m-1}(S, h) = \sum_{i=1}^{m-1} \eta (\Delta M_i(h)) - \frac{\eta^2}{2} (\Delta[M]_i(h) + \Delta \langle M \rangle_i(h))$ . Using Lemma 4 ensures that for any  $h$ ,

$$\mathbb{E}_{m-1}[\exp(\eta \Delta M_m(h) - \frac{\eta^2}{2} (\Delta[M]_m(h) + \Delta \langle M \rangle_m(h)))] \leq 1,$$

thus we have

$$\mathbb{E}_{m-1}[Z_m] \leq \mathbb{E}_{h \sim P} [\exp(f_{m-1}(S, h))] = Z_{m-1}.$$

Thus  $(Z_m)_m$  is a nonnegative supermartingale with  $Z_0 = 1$ . We can use Ville's inequality (Lemma 3) which states that

$$\mathbb{P}_S (\exists m \geq 1 : Z_m > \delta^{-1}) \leq \delta.$$

Thus, with probability  $1 - \delta$  over  $S$ , for all  $m \in \mathbb{N}$ ,  $Z_m \leq 1/\delta$ . We then have the following intermediary result. For all  $P$  a data-free prior,  $\eta \in \mathbb{R}$ , with probability  $1 - \delta$  over  $S$ , for all  $m > 0$ ,  $Q \in \mathcal{M}_1^+(\mathcal{H})$

$$\eta M_m(Q) \leq \text{KL}(Q, P) + \log(1/\delta) + \frac{\eta^2}{2} ([M]_m(Q) + \langle M \rangle_m(Q)), \quad (2)$$

recalling that  $M_m(Q) = \mathbb{E}_{h \sim Q}[M_m(h)]$ , and that similar definitons hold for  $[M]_m(Q)$ ,  $\langle M \rangle_m(Q)$ . Thus, applying the bound with  $\eta = \pm \lambda$  ( $\lambda > 0$ ) and taking an union bound gives, with probability  $1 - \delta$  over  $S$ , for any  $m \in \mathbb{N}$ ,  $Q \in \mathcal{M}_1^+(\mathcal{H})$

$$\lambda |M_m(Q)| \leq \text{KL}(Q, P) + \log(2/\delta) + \frac{\lambda^2}{2} ([M]_m(Q) + \langle M \rangle_m(Q)).$$

Dividing by  $\lambda$  concludes the proof. ■

### 3. An important particular case: learning theory with iid data

In this section, we instantiate Thm. 5 onto a learning theory framework with iid data.



**Framework** We consider a *learning problem* specified by a tuple  $(\mathcal{H}, \mathcal{Z}, \ell)$  consisting of a set  $\mathcal{H}$  of predictors, the data space  $\mathcal{Z}$ , and a loss function  $\ell : \mathcal{H} \times \mathcal{Z} \rightarrow \mathbb{R}^+$ . We consider a countable dataset  $S = (z_i)_{i \geq 1} \in \mathcal{Z}^{\mathbb{N}}$  and assume that sequence is iid following the distribution  $\mu$ . We also denote by  $\mathcal{M}_1^+(\mathcal{H})$  is the set of probabilities on  $\mathcal{H}$ .

**Definitions** The *generalisation error*  $R$  of a predictor  $h \in \mathcal{H}$  is  $\forall h, R(h) = \mathbb{E}_{z \sim \mu}[\ell(h, z)]$ , the *empirical error* of  $h$  is  $\forall h, R_m(h) = \frac{1}{m} \sum_{i=1}^m \ell(h, z_i)$  and finally the *quadratic generalisation error*  $V$  of  $h$  is  $\forall h, Quad(h) = \mathbb{E}_{z \sim \mu}[\ell(h, z)^2]$ . We also denote by *generalisation gap* for any  $h$  the quantity  $|R(h) - R_m(h)|$ .

We now state the main result of this section – its proof is deferred to Appendix B.

**Theorem 7** For any data-free prior  $P \in \mathcal{M}_1^+(\mathcal{H})$ , any  $\lambda > 0$  the following holds with probability  $1 - \delta$  over the sample  $S = (z_i)_{i \in \mathbb{N}}$ , for all  $m \in \mathbb{N}/\{0\}$ ,  $Q \in \mathcal{M}_1^+(\mathcal{H})$

$$\mathbb{E}_{h \sim Q}[R(h)] \leq \mathbb{E}_{h \sim Q} \left[ R_m(h) + \frac{\lambda}{2m} \sum_{i=1}^m \ell(h, z_i)^2 \right] + \frac{\text{KL}(Q, P) + \log(2/\delta)}{\lambda m} + \frac{\lambda}{2} \mathbb{E}_{h \sim Q}[Quad(h)].$$

A major novelty in this theorem is that the bound holds *simultaneously on all*  $m > 0$  while the natural shape of nearly all previous PAC-Bayesian bounds is to hold with high probability for a single  $m$  – this is due to the use of Ville’s inequality

Furthermore, note that this bound is empirical (all terms can be computer or approximated) at the exception of the term  $\mathbb{E}_{h \sim Q}[Quad(h)]$ . Hopefully, this quantity yet theoretical, can be controlled by a wise choice of  $\lambda$ . For instance,  $\lambda_m = 1/\sqrt{m}$  ensures a convergence rate of  $O(1/\sqrt{m})$  for the expected generalisation gap. However, doing so reduces the range of our result: as  $\lambda_m$  depends on  $m$ , our bound would then only hold for this  $m$  with high probability (and not with high probability for all  $m$ ).

Also, note that if we consider the particular case of the quadratic loss  $\ell(h, z) = (h - z)^2$ , if we assume that our posteriors  $Q$  have both bounded means and variance, then  $\lambda \mathbb{E}_{h \sim Q}[Quad(h)]$  is bounded almost surely. Thus, in this particular case, we recover a fully empirical (computable) bound for this learning problem involving an unbounded loss at the cost of reasonable assumptions on our class of posterior distributions (bounded means and variance).

Last but not least, this bound suggests a new optimisation objective for unbounded losses which is for any  $m > 0$

$$\operatorname{argmin}_Q \mathbb{E}_{h \sim Q} \left[ \frac{1}{m} \sum_{i=1}^m \left( \ell(h, z_i) + \frac{\lambda}{2} \ell(h, z_i)^2 \right) \right] + \frac{\text{KL}(Q, P)}{\lambda m}.$$

This objective differs from the classical objective of [Catoni \(2007, Thm 1.2.6\)](#) on the additional quadratic term  $\frac{\lambda}{2} \ell(h, z_i)^2$ . This objective stresses the role of the parameter  $\lambda$ . Indeed, in Catoni’s bound, it was seen as an arbitrary factor designed to weight the implicit tradeoff between the KL term and the efficiency on training data. Now this tradeoff is made explicit with the appearance of the quadratic loss onto the learning objective. Note that because  $\mathbb{E}_{h \sim Q}[Quad(h)]$  is theoretical, we did not incorporate it in our optimisation objective.

## 4. Extensions of previous results

Here we gather several corollaries of our main result in order to show how our Thm. 7 extends the validity of some classical results in the literature. More precisely we show that our result extends (up to numerical factors) the PAC-Bayes Bernstein inequality of Seldin et al. (2012a). Then, going back to the bounded case, we generalise a result from Catoni (2007) reformulated in Alquier et al. (2016) and we also show how our work strictly improves on the bound of Haddouche et al. (2021).

### 4.1. Extension of the PAC-Bayes Bernstein inequality

Here we rename two terms for consistency with Theorem 2.1 of Seldin et al. (2012a) (see Thm. 12). For a martingale  $M_m(h) = \sum_{i=1}^m X_i(S_i, h)$ , we define, at time  $m$ , *empirical cumulative variance* to be  $\hat{V}_m(h) = [M]_m(h) = \sum_{i=1}^m X_i(S_i, h)^2$  and the *cumulative variance* as  $V_m(h) = \langle M \rangle_m(h) = \sum_{i=1}^m \mathbb{E}_{i-1}[X_i(S_i, h)^2]$ .

We provide below a corollary containing two bounds: the first one being a straightforward corollary of Thm. 5, the second being valid for bounded martingales and formally close to Theorem 2.1 of Seldin et al. (2012a).

**Corollary 8** *Let  $\{P_1, P_2, \dots\}$  be a sequence of data-free prior distributions over  $\mathcal{H}$ . Let  $(\lambda_i)_{i \geq 1}$  be a sequence of positive numbers. Then the following holds with probability  $1 - \delta$  over  $S = (z_i)_{i \geq 1}$ : for any tuple  $(m, \lambda_k, P_k)$  with  $m, k \geq 1$ , any posterior  $Q$  over  $\mathcal{H}$ ,*

$$|M_m(Q)| \leq \frac{KL(Q, P_k) + 2 \log(k+1) + \log(2/\delta)}{\lambda_k} + \frac{\lambda_k}{2} (\hat{V}_m(Q) + V_m(Q)), \quad (3)$$

with  $\hat{V}_m(Q) = \mathbb{E}_{h \sim Q}[\hat{V}_m(h)]$ ,  $V_m(Q) = \mathbb{E}_{h \sim Q}[V_m(h)]$ . Furthermore, if we assume that for any  $i$ , there exists  $C_i > 0$  such that  $|X_i(S_i, h)| \leq C_i$  for all  $S_i, h$  then we have the following corollary: with probability  $1 - \delta$  over  $S$ , for any tuple  $(m, \lambda_m, P_m)$   $m \geq 1$ , any posterior  $Q$ ,

$$|M_m(Q)| \leq \frac{KL(Q, P_m) + 2 \log(m+1) + \log(2/\delta)}{\lambda_m} + \lambda_m \sum_{i=1}^m C_i^2. \quad (4)$$

The proof is deferred to Appendix B. Note that Eq. (3) holds uniformly on all tuples  $\{(\lambda_k, P_k, m) \mid k \geq 1, m \geq 1\}$  while Eq. (4), as well as Theorem 2.1 of Seldin et al. (2012a) holds uniformly on the tuples  $\{(\lambda_m, P_m, m) \mid m \geq 1\}$  which is a strictly smaller collection. Hence our approach gives guarantees for a larger event with the same confidence level.

Furthermore, Theorem 2.1 of Seldin et al. (2012a) involves the cumulative variance  $V_m(Q)$  (and not its empirical counterpart). Because this term is theoretical, we bound it in Thm. 12 by  $\sum_{i=1}^m C_i^2$  which is supposedly empirical. In this context, Eq. (4), recovers nearly exactly the bound of Seldin et al. (2012a) with the transformation of a factor  $(e - 2)$  into 1. Notice also that Eq. (4) stands with no assumption on the range of the  $\lambda_i$ , which is not the case in Thm. 12.

Finally, we stress two fundamental differences between our work and the one of Seldin et al. (2012a). First, we replace Markov's inequality by Ville's inequality; second, we exploited the exponential inequality of Lemma 4 instead of the Bernstein inequality. These allow for results for unbounded martingales for all  $m$  simultaneously.

## 4.2. Extensions of learning theory results

### 4.2.1. A GENERAL RESULT FOR BOUNDED LOSSES

We use definitions from Sec. 3 and provide a corollary of our main result when the loss is bounded by a positive constant  $K > 0$ . We assume our data are iid.

**Corollary 9** *For any data-free prior  $P \in \mathcal{M}_1^+(\mathcal{H})$ , any  $\lambda > 0$  the following holds with probability  $1 - \delta$  over the sample  $S = (z_i)_{i \in \mathbb{N}}$ , for all  $m \in \mathbb{N}/\{0\}$ ,  $Q \in \mathcal{M}_1^+(\mathcal{H})$*

$$|\mathbb{E}_{h \sim Q}[R(h)] - \mathbb{E}_{h \sim Q}[R_m(h)]| \leq \frac{\text{KL}(Q, P) + \log(2/\delta)}{\lambda m} + \lambda K^2.$$

We also have the local bound: for any  $m \geq 1$ , with probability  $1 - \delta$  over  $S$ , for all  $Q \in \mathcal{M}_1^+(\mathcal{H})$

$$\mathbb{E}_{h \sim Q}[R(h)] \leq \mathbb{E}_{h \sim Q}[R_m(h)] + \frac{\text{KL}(Q, P) + \log(2/\delta)}{\lambda} + \frac{\lambda K^2}{m}.$$

The proof is deferred to Appendix B. Remark that the second bound of Cor. 9 is exactly the Catoni bound stated in Alquier et al. (2016) (see Thm. 13 in Appendix A) up to a numerical factor of 2.

The first bound is, to our knowledge, the first PAC-Bayesian bound for bounded losses holding uniformly (for a given parameter  $\lambda$ ) on the choice of  $Q, m$  and thus extends the scope of Catoni's bound which holds for a single  $m$  with high probability. Indeed, if we want for instance Thm. 13 to hold for any  $i \in \{1..m\}$ , we then have to take an union bound on  $m$  events which turns the term  $\log(1/\delta)$  into  $\log(m/\delta)$  (but with the benefit of holding for  $m$  parameters  $\lambda_1, \dots, \lambda_m$ ). This point is common to the most classical PAC-Bayesian bounds (as those of McAllester, 1998, 1999; Maurer, 2004; Catoni, 2007; Tolstikhin and Seldin, 2013) and impeach us to have a bound uniformly on all  $m \in \mathbb{N}/\{0\}$  as  $\log(m)$  goes to infinity asymptotically.

### 4.2.2. AN EXTENSION OF HADDOUCHE ET AL. (2021)

We now focus on the work of Haddouche et al. (2021) which provides general PAC-Bayesian bounds for unbounded losses. Their theorems hold for iid data and under the so-called *HYPE* (for HYPothesis-dependent rangE) condition. It states that a loss function  $\ell$  is *HYPE*( $K$ ) compliant if there exists a function  $K : \mathcal{H} \rightarrow \mathbb{R}^+$  (supposedly accessible) such that  $\forall z \in \mathcal{Z}, \ell(h, z) \leq K(h)$ . We provide Cor. 10 to compare ourselves with their main result (stated in Thm. 14 for convenience).

**Corollary 10** *For any data-free prior  $P \in \mathcal{M}_1^+(\mathcal{H})$ , any loss function  $\ell$  being *HYPE*( $K$ ) compliant, any  $\alpha \in [0, 1], m \geq 1$ , the following holds with probability  $1 - \delta$  over the sample  $S = (z_i)_{i \in \mathbb{N}}$ , for all  $Q \in \mathcal{M}_1^+(\mathcal{H})$*

$$\begin{aligned} \mathbb{E}_{h \sim Q}[R(h)] \leq & \mathbb{E}_{h \sim Q} \left[ \frac{1}{m} \sum_{i=1}^m \left( \ell(h, z_i) + \frac{1}{2m^{1-\alpha}} \ell(h, z_i)^2 \right) \right] + \frac{\text{KL}(Q, P) + \log(1/\delta)}{m^\alpha} \\ & + \frac{1}{2m^{1-\alpha}} \mathbb{E}_{h \sim Q}[K^2(h)]. \end{aligned}$$

**Proof** The proof is a straightforward application of Thm. 7 by fixing  $m \geq 1$  choosing  $\lambda = m^{\alpha-1}$  (thus we localise Thm. 7 to a single  $m$ ), and bounding  $\text{Quad}(h)$  by  $K^2(h)$ .  $\blacksquare$

The main improvement of our bound over Thm. 14 is that we do not have to assume the convergence of an exponential moment to obtain a non-trivial bound. Indeed, we transformed the (implicit) assumption  $\mathbb{E}_{h \sim P} \left[ \exp \left( \frac{K(h)^2}{2m^{1-2\alpha}} \right) \right] < +\infty$  onto  $\mathbb{E}_{h \sim Q} [K(h)^2] < +\infty$ , which is significantly less restrictive. Furthermore, Thm. 14 holds for a single choice of  $m$  while ours still holds uniformly over all integers  $m > 0$ .

Cor. 10 also sheds new light on the *HYPE* condition. Indeed, in Haddouche et al. (2021),  $K$  only intervenes in an exponential moment involving the prior  $P$ , while ours considers a second-order moment on  $K$  implying the posterior  $Q$ . The difference is major as  $\mathbb{E}_{h \sim Q} [K(h)^2]$  can be controlled by a wise choice of posterior. Thus it can be incorporated in our optimisation route, acting now as an optimisation constraint instead of an environment constraint.

## 5. Application to the multi-armed bandit problem

We exploit our main result in the context of the multi-armed bandit problem – we adopt the framework of Seldin et al. (2012a).

**Framework.** Let  $\mathcal{A}$  be a set of actions of size  $|\mathcal{A}| = K < +\infty$  and  $a \in \mathcal{A}$  be an action. At each round  $i$ , the environment furnishes a reward function  $R_i : \mathcal{A} \rightarrow \mathbb{R}$  which associate a reward  $R_i(a)$  to the arm  $a$ . Assuming the  $R_i$ s are iid, we denote for any  $a$ , the *expected reward for action a* to be  $R(a) = \mathbb{E}_{R_1} [R_1(a)]$ . At each round  $i$ , the player executes an action  $A_i$  according to a policy  $\pi_i$ . We then set the filtration  $(\mathcal{F}_i)_{i \geq 1}$  to be  $\mathcal{F}_i = \sigma(\{\pi_j, A_j, R_j \mid 1 \leq j \leq i\})$ .

**Assumptions.** We suppose here that  $(R_i)_{i \geq 1}$  is an iid sequence and that at each time  $i$ ,  $A_i$  and  $R_i$  are independent and that  $\pi_i$  is  $\mathcal{F}_{i-1}$  measurable. This means that the player is not aware of the rewards each round and performs its current move with regards to the past.

We also add two technical assumptions. First, the order two moment of the expected reward is uniformly bounded:  $\sup_{a \in \mathcal{A}} \mathbb{E}_{R_1} [R_1(a)^2] \leq C$ . This assumption is strictly less restrictive than the boundedness assumption made in Seldin et al. (2012a). Similarly to this work, we also assume that there exists a sequence  $(\varepsilon_i)_{i \geq 1}$  such that  $\inf_{a \in \mathcal{A}} \pi_i(a) \geq \varepsilon_i$ . We say that  $(\pi_i)$  are *bounded from below* by  $(\varepsilon_i)$ .

**Definitions.** For  $i \geq 1$  and  $a \in \{1, \dots, K\}$ , define a set of random variables  $(R_i^a)_{i \geq 1}$  (*the importance weighted samples*, Sutton and Barto, 2018)

$$R_i^a := \begin{cases} \frac{1}{\pi_i(a)} R_i, & \text{if } A_i = a, \\ 0, & \text{otherwise.} \end{cases}$$

We define for any time  $m$ :  $\hat{R}_m(a) = \frac{1}{m} \sum_{i=1}^m R_i^a$ . Observe that for all  $i$ ,  $\mathbb{E} [R_i^a \mid \mathcal{F}_{i-1}] = R(a)$  and  $\mathbb{E} [\hat{R}_m(a)] = R(a)$ . Let  $a^*$  be the "best" action (the action with the highest expected reward, if there are multiple "best" actions pick any of them). Define the *expected and empirical per-round regrets* as

$$\Delta(a) = R(a^*) - R(a), \quad \hat{\Delta}_m(a) = \hat{R}_m(a^*) - \hat{R}_m(a).$$

Observe that  $m \left( \hat{\Delta}_m(a) - \Delta(a) \right)$  forms a martingale. Let

$$V_m(a) = \sum_{i=1}^m \mathbb{E} \left[ \left( R_i^{a^*} - R_i^a - [R(a^*) - R(a)] \right)^2 \mid \mathcal{F}_{i-1} \right]$$

be the cumulative variance of this martingale and

$$\hat{V}_m(a) = \sum_{i=1}^m \left( R_i^{a^*} - R_i^a - [R(a^*) - R(a)] \right)^2$$

its empirical counterpart. We denote for any distribution  $Q$  over  $\mathcal{A}$ ,  $\Delta(Q) = \mathbb{E}_{a \sim Q}[\Delta(a)]$ ,  $V_m(Q) = \mathbb{E}_{a \sim Q}[V_m(a)]$ , similar definitions hold for  $\hat{\Delta}_m(Q)$ ,  $\hat{V}_m(Q)$ . We can now state the main result of this section – its proof is deferred to Appendix B.

**Theorem 11** *For any  $m \geq 1$ , any history-dependent policy sequence  $(\pi_i)_{i \geq 1}$  bounded from below by  $(\varepsilon_i)_{i \geq 1}$ , we have with probability  $1 - \delta$ , for all posterior  $Q$*

$$\left| \Delta(Q) - \hat{\Delta}_m(Q) \right| \leq 2 \sqrt{\frac{\left(1 + \frac{2K}{\delta}\right) (\log(K) + \log(4/\delta))}{m\varepsilon_m}}.$$

To the best of our knowledge, this result is the first PAC-Bayesian guarantees for multi-armed bandits with unbounded rewards. The proposed bound is as tight as Theorem 2.3 of Seldin et al. (2012a), up to a factor  $(e - 2)$  transformed into  $\left(1 + \frac{4K}{\delta}\right)$  within the square root. Note that our result comes at the price of the localisation: Theorem 2.3 of Seldin et al. (2012a) proposes a bound holding uniformly for all time  $m$  while our approach only holds for a single time  $m$ .

We believe there is room for improvement in Thm. 11. Indeed, the current approach is naive as it consists in bounding crudely with high probability the empirical variance. Maybe more subtle arguments or additional assumptions could tighten the bound, or extend its validity to several times  $m$  simultaneously. We leave this as an open question for future work.

## 6. Conclusion

We showed that it is possible to generalise the classical PAC-Bayes toolbox to unbounded martingales (resp. learning problem with unbounded losses), the solely implicit assumption being the existence of second order moments on the martingale difference sequence (resp. on the loss function) which is reasonable as many PAC-Bayes bound lies on assumptions on exponential moments (e.g. the subgaussian assumption) to work. We also proved that our main theorem can be seen as a general basis allowing to recover several PAC-Bayesian bounds. We believe this demonstrates the use of the supermartingale techniques is a fruitful approach to establish general PAC-Bayes bounds in the batch setting. A natural next step is to adapt our novel techniques to the online PAC-Bayes framework of Haddouche and Guedj (2022) (which only consider bounded losses) as martingales naturally appear in their work. We expect this to lead to the first online PAC-Bayes bound for unbounded losses with non-iid data.

## References

- Pierre Alquier. User-friendly introduction to PAC-Bayes bounds, 2021. URL <https://arxiv.org/abs/2110.11216>.
- Pierre Alquier and Benjamin Guedj. Simpler PAC-Bayesian bounds for hostile data. *Machine Learning*, 107(5):887–902, 2018. ISSN 1573-0565. URL <http://dx.doi.org/10.1007/s10994-017-5690-0>.

- Pierre Alquier, James Ridgway, and Nicolas Chopin. On the properties of variational approximations of gibbs posteriors. *Journal of Machine Learning Research*, 17(236):1–41, 2016. URL <http://jmlr.org/papers/v17/15-290.html>.
- Arindam Banerjee. On Bayesian Bounds. In *Proceedings of the 23rd international conference on Machine learning*, pages 81–88, 2006.
- Bernard Bercu and Abderrahmen Touati. Exponential inequalities for self-normalized martingales with applications. *The Annals of Applied Probability*, 18(5):1848–1869, 2008.
- Felix Biggs and Benjamin Guedj. Differentiable PAC-Bayes objectives with partially aggregated neural networks. *Entropy*, 23(10):1280, 2021.
- Felix Biggs and Benjamin Guedj. On margins and derandomisation in PAC-Bayes. In Gustau Camps-Valls, Francisco J. R. Ruiz, and Isabel Valera, editors, *Proceedings of The 25th International Conference on Artificial Intelligence and Statistics [AISTATS]*, volume 151 of *Proceedings of Machine Learning Research*, pages 3709–3731. PMLR, 28–30 Mar 2022a. URL <https://proceedings.mlr.press/v151/biggs22a.html>.
- Felix Biggs and Benjamin Guedj. Non-vacuous Generalisation Bounds for shallow neural networks. In *ICML*, 2022b.
- Felix Biggs, Valentina Zantedeschi, and Benjamin Guedj. On margins and generalisation for voting classifiers. In *NeurIPS*, 2022. URL <https://arxiv.org/abs/2206.04607>.
- O Catoni. PAC-Bayesian Supervised Classification: The Thermodynamics of Statistical Learning. Institute of Mathematical Statistics Lecture Notes—Monograph Series 56. *IMS, Beachwood, OH. MR2483528*, 5544465, 2007.
- Olivier Catoni. A PAC-Bayesian approach to adaptive classification. *preprint*, 840, 2003.
- Badr-Eddine Chérif-Abdellatif, Yuyang Shi, Arnaud Doucet, and Benjamin Guedj. On PAC-Bayesian reconstruction guarantees for VAEs. In Gustau Camps-Valls, Francisco J. R. Ruiz, and Isabel Valera, editors, *Proceedings of The 25th International Conference on Artificial Intelligence and Statistics [AISTATS]*, volume 151 of *Proceedings of Machine Learning Research*, pages 3066–3079. PMLR, 28–30 Mar 2022. URL <https://proceedings.mlr.press/v151/cherief-abdellatif22a.html>.
- Imre Csiszár. I-divergence geometry of probability distributions and minimization problems. *The annals of probability*, pages 146–158, 1975.
- Victor H De la Peña, Tze Leung Lai, and Qi-Man Shao. *Self-normalized processes: Limit theory and Statistical Applications*, volume 204. Springer, 2009.
- Monroe D Donsker and SR Srinivasa Varadhan. Asymptotic evaluation of certain Markov process expectations for large time, I. *Communications on Pure and Applied Mathematics*, 28(1):1–47, 1975.
- JL Doob. Jean Ville, Étude Critique de la Notion de Collectif. *Bulletin of the American mathematical society*, 45(11):824–824, 1939.

Paul Dupuis and Richard S Ellis. *A Weak Convergence Approach to the Theory of Large Deviations*. Wiley-Interscience, 2011.

Rick Durrett. *Probability: theory and examples*, volume 49. Cambridge university press, 2019.

Gintare Karolina Dziugaite and Daniel M Roy. Computing Nonvacuous Generalization Bounds for Deep (stochastic) Neural Networks with many more parameters than training data. *arXiv preprint arXiv:1703.11008*, 2017.

M Fard and Joelle Pineau. PAC-Bayesian Model Selection for Reinforcement Learning. *Advances in Neural Information Processing Systems (NeurIPS)*, 23, 2010.

Pascal Germain, Alexandre Lacasse, François Laviolette, and Mario Marchand. PAC-Bayesian Learning of Linear Classifiers. In *Proceedings of the 26th Annual International Conference on Machine Learning, ICML '09*, page 353–360, New York, NY, USA, 2009. Association for Computing Machinery. ISBN 9781605585161. URL <https://doi.org/10.1145/1553374.1553419>.

Pascal Germain, Francis Bach, Alexandre Lacoste, and Simon Lacoste-Julien. PAC-Bayesian theory meets Bayesian inference. *Advances in Neural Information Processing Systems (NeurIPS)*, 29, 2016.

Benjamin Guedj. A Primer on PAC-Bayesian Learning. In *Proceedings of the second congress of the French Mathematical Society*, 2019.

Benjamin Guedj and Pierre Alquier. PAC-Bayesian estimation and prediction in sparse additive models. *Electron. J. Statist.*, 7:264–291, 2013. doi: 10.1214/13-EJS771. URL <https://doi.org/10.1214/13-EJS771>.

Benjamin Guedj and Sylvain Robbiano. PAC-Bayesian high dimensional bipartite ranking. *Journal of Statistical Planning and Inference*, 196:70 – 86, 2018. ISSN 0378-3758. doi: <https://doi.org/10.1016/j.jspi.2017.10.010>. URL <http://www.sciencedirect.com/science/article/pii/S0378375817301945>.

Maxime Haddouche and Benjamin Guedj. Online PAC-Bayes Learning. *To be published in Advances in Neural Information Processing Systems (NeurIPS) 2022*, 2022.

Maxime Haddouche, Benjamin Guedj, Omar Rivasplata, and John Shawe-Taylor. PAC-Bayes unleashed: generalisation bounds with unbounded losses. *Entropy*, 23(10):1330, 2021.

Matthew Holland. PAC-Bayes under potentially heavy tails. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems (NeurIPS) 32*, pages 2715–2724. Curran Associates, Inc., 2019. URL <http://papers.nips.cc/paper/8539-pac-bayes-under-potentially-heavy-tails.pdf>.

Sham M Kakade, Karthik Sridharan, and Ambuj Tewari. On the complexity of linear prediction: Risk bounds, margin bounds, and regularization. *Advances in Neural Information Processing Systems (NeurIPS)*, 21, 2008.

- Ilja Kuzborskij and Csaba Szepesvári. Efron-Stein PAC-Bayesian Inequalities. *arXiv preprint arXiv:1909.01931*, 2019.
- John Langford and Robert Schapire. Tutorial on practical prediction theory for classification. *Journal of machine learning research*, 6(3), 2005.
- Gaël Letarte, Pascal Germain, Benjamin Guedj, and François Laviolette. Dichotomize and generalize: PAC-Bayesian binary activated deep neural networks. *Advances in Neural Information Processing Systems (NeurIPS)*, 32, 2019.
- Guy Lever, François Laviolette, and John Shawe-Taylor. Distribution-dependent PAC-Bayes priors. In *International Conference on Algorithmic Learning Theory*, pages 119–133. Springer, 2010.
- Guy Lever, François Laviolette, and John Shawe-Taylor. Tighter PAC-Bayes bounds through distribution-dependent priors. *Theoretical Computer Science*, 473:4–28, 2013.
- Andreas Maurer. A note on the PAC Bayesian theorem. *arXiv preprint cs/0411099*, 2004.
- David A McAllester. Some PAC-Bayesian theorems. In *Proceedings of the eleventh annual conference on Computational Learning Theory*, pages 230–234. ACM, 1998.
- David A McAllester. PAC-Bayesian model averaging. In *Proceedings of the twelfth annual conference on Computational Learning Theory*, pages 164–170. ACM, 1999.
- David A McAllester. PAC-Bayesian stochastic model selection. *Machine Learning*, 51(1):5–21, 2003.
- Zakaria Mhammedi, Peter Grünwald, and Benjamin Guedj. PAC-Bayes Un-Expected Bernstein Inequality. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems (NeurIPS)* 32, pages 12202–12213. Curran Associates, Inc., 2019. URL <http://papers.nips.cc/paper/9387-pac-bayes-un-expected-bernstein-inequality.pdf>.
- Luca Oneto, Davide Anguita, and Sandro Ridella. PAC-Bayesian analysis of distribution dependent priors: Tighter risk bounds and stability analysis. *Pattern Recognition Letters*, 80:200–207, 2016.
- Emilio Parrado-Hernández, Amiran Ambroladze, John Shawe-Taylor, and Shiliang Sun. PAC-Bayes bounds with data dependent priors. *The Journal of Machine Learning Research*, 13(1): 3507–3531, 2012.
- Maria Perez-Ortiz, Omar Rivasplata, Benjamin Guedj, Matthew Gleeson, Jingyu Zhang, John Shawe-Taylor, Mirosław Bober, and Josef Kittler. Learning PAC-Bayes Priors for Probabilistic Neural Networks. 2021a. URL <https://arxiv.org/abs/2109.10304>.
- Maria Perez-Ortiz, Omar Rivasplata, Emilio Parrado-Hernandez, Benjamin Guedj, and John Shawe-Taylor. Progress in self-certified neural networks. In *NeurIPS 2021 workshop Bayesian Deep Learning [BDL]*, 2021b. URL <http://bayesiandeeplearning.org/2021/papers/38.pdf>.
- Maria Pérez-Ortiz, Omar Rivasplata, John Shawe-Taylor, and Csaba Szepesvári. Tighter Risk Certificates for Neural Networks. *Journal of Machine Learning Research*, 22, 2021.



- Antoine Picard-Weibel and Benjamin Guedj. On change of measure inequalities for  $f$ -divergences. *arXiv preprint arXiv:2202.05568*, 2022.
- Omar Rivasplata, Vikram M Tankasali, and Csaba Szepesvári. PAC-Bayes with Backprop. *arXiv preprint arXiv:1908.07380*, 2019.
- Omar Rivasplata, Ilya Kuzborskij, Csaba Szepesvári, and John Shawe-Taylor. PAC-Bayes analysis beyond the usual bounds. *Advances in Neural Information Processing Systems (NeurIPS)*, 33: 16833–16845, 2020.
- Matthias Seeger. PAC-Bayesian Generalization Error Bounds for Gaussian Process Classification. *Journal of Machine Learning Research*, 3, 08 2002.
- Yevgeny Seldin, François Laviolette, John Shawe-Taylor, Jan Peters, and Peter Auer. PAC-Bayesian Analysis of Martingales and Multiarmed Bandits. *arXiv preprint arXiv:1105.2416*, 2011.
- Yevgeny Seldin, Nicolò Cesa-Bianchi, Peter Auer, François Laviolette, and John Shawe-Taylor. PAC-Bayes-Bernstein Inequality for Martingales and its Application to Multiarmed Bandits. In Dorota Glowacka, Louis Dorard, and John Shawe-Taylor, editors, *Proceedings of the Workshop on Online Trading of Exploration and Exploitation 2*, volume 26 of *Proceedings of Machine Learning Research*, pages 98–111, Bellevue, Washington, USA, 02 Jul 2012a. PMLR. URL <https://proceedings.mlr.press/v26/seldin12a.html>.
- Yevgeny Seldin, François Laviolette, Nicolo Cesa-Bianchi, John Shawe-Taylor, and Peter Auer. PAC-Bayesian Inequalities for Martingales. *IEEE Transactions on Information Theory*, 58(12): 7086–7093, 2012b.
- J. Shawe-Taylor and R. C. Williamson. A PAC analysis of a Bayes estimator. In *Proceedings of the 10th annual conference on Computational Learning Theory*, pages 2–9. ACM, 1997.
- Richard S Sutton and Andrew G Barto. *Reinforcement Learning: An introduction*. MIT press, 2018.
- Terence Tao. *An introduction to measure theory*, volume 126. American Mathematical Society Providence, 2011.
- Niklas Thiemann, Christian Igel, Olivier Wintenberger, and Yevgeny Seldin. A strongly quasiconvex PAC-Bayesian bound. In *International Conference on Algorithmic Learning Theory*, pages 466–492. PMLR, 2017.
- Ilya O Tolstikhin and Yevgeny Seldin. PAC-Bayes-Empirical-Bernstein Inequality. In C.J. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems (NeurIPS)*, volume 26. Curran Associates, Inc., 2013. URL <https://proceedings.neurips.cc/paper/2013/file/a97da629b098b75c294dffdc3e46390>
- Yi-Shan Wu and Yevgeny Seldin. Split-kl and pac-bayes-split-kl inequalities. *To be published in Advances in Neural Information Processing Systems (NeurIPS) 2022*, 2022.

## Appendix A. Some PAC-Bayesian background

We present below an immediate corollary of [Seldin et al. \(2012a, Thm 2.1\)](#) where we upper bounded the cumulative by an empirical quantity (the sum of squared upper bound of the martingale difference sequence).

**Theorem 12 (immediate corollary of [Seldin et al., 2012a, Thm 2.1](#))** *Let  $\{C_1, C_2, \dots\}$  be an increasing sequence set in advance, such that  $|X_i(S_i, h)| \leq C_i$  for all  $S_i, h$  with probability 1. Let  $\{P_1, P_2, \dots\}$  be a sequence of data-free prior distributions over  $\mathcal{H}$ . Let  $(\lambda_i)_{i \geq 1}$  be a sequence of positive numbers such that*

$$\lambda_m \leq \frac{1}{C_m}.$$

*Then with probability  $1 - \delta$  over  $S = (z_i)_{i \geq 1}$ , for all  $m \geq 1$ , any posterior  $Q$  over  $\mathcal{H}$ ,*

$$|M_m(Q)| \leq \frac{KL(Q||P_m) + 2\log(m+1) + \log \frac{2}{\delta}}{\lambda_m} + (e-2)\lambda_m \sum_{i=1}^m C_i^2,$$

*where  $V_m(Q)$  is defined in [Sec. 4.1](#).*

Below, we use the definitions introduced in [Sec. 3](#). We study here a particular case of [Alquier et al. \(2016\)](#) for bounded losses which are especially subgaussian thanks to Hoeffding's lemma.

**Theorem 13 (Adapted from [Alquier et al., 2016, Theorem 4.1](#))** *Let  $m > 0, S = (z_1, \dots, z_m)$  be an iid sample from the same law  $\mu$ . For any data-free prior  $P$ , for any loss function  $\ell$  bounded by  $K$ , any  $\lambda > 0, \delta \in ]0; 1[$ , one has with probability  $1 - \delta$  for any posterior  $Q \in \mathcal{M}_1(\mathcal{H})$*

$$\mathbb{E}_{h \sim Q}[R(h)] \leq \mathbb{E}_{h \sim Q}[R_m(h)] + \frac{KL(Q||P) + \log(1/\delta)}{\lambda} + \frac{\lambda K^2}{2m}.$$

**Theorem 14 ([Haddouche et al., 2021, Theorem 3](#))** *Let the loss  $\ell$  be HYPE( $K$ ) compliant. For any  $P \in \mathcal{M}_1^+(\mathcal{H})$  with no data dependency, for any  $\alpha \in \mathbb{R}$  and for any  $\delta \in [0, 1]$ , we have with probability at least  $1 - \delta$  over size- $m$  samples  $S$ , for any  $Q$*

$$\mathbb{E}_{h \sim Q}[R(h)] \leq \mathbb{E}_{h \sim Q}[R_m(h)] + \frac{KL(Q||P) + \log(\frac{1}{\delta})}{m^\alpha} + \frac{1}{m^\alpha} \log \left( \mathbb{E}_{h \sim P} \left[ \exp \left( \frac{K(h)^2}{2m^{1-2\alpha}} \right) \right] \right).$$

## Appendix B. Proofs

### B.1. Proof of [Thm. 7](#)

**Proof** Let  $P$  a fixed data-free prior, set  $(\mathcal{F}_i)_{i \geq 0}$  such that for all  $i$ ,  $z_i$  is  $\mathcal{F}_i$  measurable. We also set for any fixed  $h \in \mathcal{H}$ ,  $M_m(h) := \sum_{i=1}^m \ell(h, z_i) - R(h)$ . Note that because data are iid, for any fixed  $h$ , the sequence  $(M_m(h))_m$  is indeed a martingale. We set for any  $m \geq 1, h \in \mathcal{H}$

$$[M]_m(h) = \sum_{i=1}^m (\ell(h, z_i) - R(h))^2$$

and

$$\langle M \rangle_m(h) = \sum_{i=1}^m \mathbb{E}_{i-1}[(\ell(h, z_i) - R(h))^2] = \sum_{i=1}^m \mathbb{E}_{z \sim \mu}[(\ell(h, z) - R(h))^2].$$

The last equality holds because data is assumed iid. Thus, we can apply Thm. 5 to obtain with probability  $1 - \delta$

$$|M_m(Q)| \leq \frac{\text{KL}(Q, P) + \log(2/\delta)}{\lambda} + \frac{\lambda}{2} ([M]_m(Q)^2 + \langle M \rangle_m(Q)^2).$$

Now, we notice that  $|M_m(Q)| = m|\mathbb{E}_{h \sim Q}[R(h) - R_m(h)]|$  and that for any  $m, h$ , because  $\ell$  is nonnegative

$$\begin{aligned} [M]_m(h) + \langle M \rangle_m(h) &= \sum_{i=1}^m (\ell(h, z_i) - R(h))^2 + \mathbb{E}_{z \sim \mu} [(\ell(h, z) - R(h))^2] \\ &\leq \sum_{i=1}^m \ell(h, z_i)^2 + R(h)^2 + \mathbb{E}_{z \sim \mu} [\ell(h, z)^2] - R(h)^2. \end{aligned}$$

Thus integrating over  $h$  gives:

$$[M]_m(Q) + \langle M \rangle_m(Q) \leq \sum_{i=1}^m \mathbb{E}_{h \sim Q} [\ell(h, z_i)^2] + m \mathbb{E}_{h \sim Q} [\text{Quad}(h)].$$

Then dividing by  $m$  and applying the last inequality gives

$$\begin{aligned} &\mathbb{E}_{h \sim Q} [R(h)] \\ &\leq \mathbb{E}_{h \sim Q} \left[ \frac{1}{m} \sum_{i=1}^m \left( \ell(h, z_i) + \frac{\lambda}{2} \ell(h, z_i)^2 \right) \right] + \frac{\text{KL}(Q, P) + \log(2/\delta)}{\lambda m} + \frac{\lambda}{2} \mathbb{E}_{h \sim Q} [\text{Quad}(h)]. \end{aligned}$$

This concludes the proof. ■

## B.2. Proof of Cor. 8

**Proof** Fix  $\delta > 0$ . For any pair  $(\lambda_k, P_k), k \geq 1$ , we apply Thm. 5 with

$$\delta_k := \frac{\delta}{k(k+1)} \geq \frac{\delta}{(k+1)^2}.$$

Notice that we have  $\sum_{k=1}^{+\infty} \delta_k = \delta$ . We then have with probability  $1 - \delta_k$  over  $S$ , for any  $m \geq 1$ , any posterior  $Q$ ,

$$|M_m(Q)| \leq \frac{\text{KL}(Q, P_k) + 2 \log(k+1) + \log(2/\delta)}{\lambda_k} + \frac{\lambda_k}{2} (\hat{V}_m(Q) + V_m(Q)).$$

Taking an union bound on all those event, gives the final result, valid with probability  $1 - \delta$  over the sample  $S$ , for any any tuple  $(m, \lambda_k, P_k)$  with  $m, k \geq 1$ , any posterior  $Q$  over  $\mathcal{H}$ . This gives Eq. (3).

To obtain Eq. (4), we restrict the range of Eq. (3) to the tuples  $(m, \lambda_m, P_m), m \geq 1$  (the restricted set of tuples where  $k = m$ ) and we bound both  $\hat{V}_m(Q), V_m(Q)$  by  $\sum_{i=1}^m C_i^2$  to conclude. ■

### B.3. Proof of Cor. 9

**Proof** For the first bound we start from the intermediary result Eq. (2) of Thm. 5. Using the same martingale as in Thm. 7 gives, for any  $\eta \in \mathbb{R}$ , holding with probability  $1 - \delta$  for any  $m > 0$ ,  $Q \in \mathcal{M}_1^+(\mathcal{H})$

$$\begin{aligned} & \eta \left( \sum_{i=1}^m \mathbb{E}_{h \sim Q}[\ell(h, z_i)] - m \mathbb{E}_{h \sim Q}[R(h)] \right) \\ & \leq \text{KL}(Q, P) + \log(1/\delta) + \frac{\eta^2}{2} \sum_{i=1}^m \mathbb{E}_{h \sim Q}[\Delta[M]_i(h) + \Delta\langle M \rangle_i(h)]. \end{aligned}$$

Taking  $\eta = \pm\lambda$  with  $\lambda > 0$  gives

$$\begin{aligned} & \lambda m |\mathbb{E}_{h \sim Q}[R(h) - R_m(h)]| \\ & \leq \text{KL}(Q, P) + \log(1/\delta) + \frac{\lambda^2}{2} \sum_{i=1}^m \mathbb{E}_{h \sim Q}[\Delta[M]_i(h) + \Delta\langle M \rangle_i(h)]. \quad (5) \end{aligned}$$

Finally, divide by  $\lambda m$  and bound  $\Delta[M]_i(h) + \Delta\langle M \rangle_i(h)$  by  $2K^2$  to conclude.

For the second bound, we start from Eq. (5) again and for a fixed  $m$ , we now apply our result with  $\lambda' = \lambda/m$ . We then have for any  $m$ , with probability  $1 - \delta$ , for any  $Q$

$$\lambda |\mathbb{E}_{h \sim Q}[R(h) - R_m(h)]| \leq \text{KL}(Q, P) + \log(1/\delta) + \frac{\lambda^2}{2m^2} \sum_{i=1}^m \mathbb{E}_{h \sim Q}[\Delta[M]_i(h) + \Delta\langle M \rangle_i(h)]$$

Finally, dividing by  $\lambda$ , bounding  $\Delta[M]_i(h) + \Delta\langle M \rangle_i(h)$  by  $2K^2$  and rearranging the terms concludes the proof.  $\blacksquare$

### B.4. Proof of Thm. 11

**Proof** Let  $(\lambda_m)_{i \geq 1}$  be a countable sequence of positive scalars. As precised earlier  $M_m(a) := m \left( \hat{\Delta}_m(a) - \Delta(a) \right)$  is a martingale. We then apply Thm. 5 with the uniform prior ( $\forall a, P(a) = \frac{1}{K}$ ) and  $\lambda = \lambda_m$  (depending possibly on  $m$ ): with probability  $1 - \delta/2$ , for any tuple  $(m, \lambda_m)$  with  $m \geq 1$ , any posterior  $Q$ ,

$$|M_m(Q)| \leq \frac{\text{KL}(Q, P) + 2 + \log(4/\delta)}{\lambda_m} + \frac{\lambda_m}{2} \left( \hat{V}_m(Q) + V_m(Q) \right).$$

Notice that for any  $Q$ ,  $\text{KL}(Q, P) \leq \log(K)$  by concavity of the log. We now fix an horizon  $M > 0$ , we then have in particular, with probability  $1 - \delta/2$ : for any posterior  $Q$ ,

$$|M_m(Q)| \leq \frac{\log(K) + 2 \log(k+1) + \log(4/\delta)}{\lambda_k} + \frac{\lambda_m}{2} \left( \hat{V}_m(Q) + V_m(Q) \right).$$

We now have to deal with  $V_k(Q), \hat{V}_k(Q)$  for all  $k \leq m$ . To do so, we propose the two following lemmas.

**Lemma 15** For all  $m \geq 1$ ,  $a \in \mathcal{A}$ ,  $V_m(a) \leq \frac{2Cm}{\varepsilon_m}$ . Then, we have for any  $m, Q$ ,  $V_m(Q) \leq \frac{2Cm}{\varepsilon_m}$ .

**Proof** We have

$$\begin{aligned}
 V_t(a) &= \sum_{i=1}^m \mathbb{E} \left[ \left( \left[ R_i^{a^*} - R_i^a \right] - \Delta(a) \right)^2 \mid \mathcal{F}_{i-1} \right] \\
 &= \sum_{i=1}^m \mathbb{E} \left[ \left( R_i^{a^*} - R_i^a \right)^2 \mid \mathcal{F}_{i-1} \right] - m\Delta(a)^2 \\
 &\leq \sum_{i=1}^m \mathbb{E} \left[ \left( R_i^{a^*} - R_i^a \right)^2 \mid \mathcal{F}_{i-1} \right] \\
 &= \sum_{i=1}^m \mathbb{E} \left[ \mathbb{E}_{A_i \sim \pi_i} \mathbb{E}_{R_i} \left[ \frac{1}{\pi_i(a^*)^2} R_i(a^*)^2 \mathbf{1}(A_i = a^*) + \frac{1}{\pi_i(a)^2} R_i(a)^2 \mathbf{1}(A_i = a) \mid \mathcal{F}_{i-1} \right] \right].
 \end{aligned}$$

The last line holding because  $R_i$  is independent of  $\mathcal{F}_{i-1}$ ,  $A_i$  is independent of  $R_i$  and  $\pi$  is  $\mathcal{F}_{i-1}$  measurable. We now use that for all  $i, a$ ,  $\mathbb{E}_{R_i}[R_i(a)^2] \leq C$

$$\begin{aligned}
 &= \sum_{i=1}^m \mathbb{E} \left[ \mathbb{E}_{A_i \sim \pi_i} \left[ \frac{1}{\pi_i(a^*)^2} C \mathbf{1}(A_i = a^*) + \frac{1}{\pi_i(a)^2} C \mathbf{1}(A_i = a) \right] \mid \mathcal{F}_{i-1} \right] \\
 &= \sum_{i=1}^m C \left( \frac{\pi_i(a)}{\pi_i(a)^2} + \frac{\pi_i(a^*)}{\pi_i(a^*)^2} \right) \\
 &= \sum_{i=1}^m C \left( \frac{1}{\pi_i(a)} + \frac{1}{\pi_i(a^*)} \right) \\
 &\leq \frac{2Cm}{\varepsilon_m}.
 \end{aligned}$$

■

**Lemma 16** Let  $m \geq 1$ , with probability  $1 - \delta/2$ , for any posterior  $Q$ , we have

$$\hat{V}_m(Q) \leq \frac{4CKm}{\varepsilon_m \delta}.$$

**Proof** Let  $Q$  a distribution over  $\mathcal{A}$ . Recall that

$$\begin{aligned}
 \hat{V}_m(Q) &= \sum_{i=1}^m \left( R_i^{a^*} - R_i^a - [R(a^*) - R(a)] \right)^2 \\
 &= \sum_{a \in \mathcal{A}} Q(a) \hat{V}_m(a)
 \end{aligned}$$

Notice that for any  $a$ ,  $(\hat{S}_m^a)_m$  is a nonnegative random variable. We then apply Markov's inequality for any  $a$ , with probability  $1 - \delta/2K$

$$\hat{V}_m(a) \leq \frac{2K \mathbb{E}[\hat{V}_m(a)]}{\delta}$$

Noticing that  $\mathbb{E}[\hat{V}_m(a)] = \mathbb{E}[V_m(a)]$ , we can apply Thm. 15 to conclude that

$$\mathbb{E}[\hat{V}_m(a)] \leq \frac{2Cm}{\varepsilon_m}.$$

Finally, taking an union bound on those events for all  $a \in \mathcal{A}$  gives us, with probability  $1 - \delta/2$ , for any posterior  $Q$

$$\begin{aligned} V_m(Q) &\leq \sum_{a \in \mathcal{A}} Q(a) \hat{V}_m(a) \\ &\leq \sum_{a \in \mathcal{A}} Q(a) \frac{4CKm}{\varepsilon_m \delta} \\ &= \frac{4CKm}{\varepsilon_m \delta}. \end{aligned}$$

This concludes the proof. ■

To conclude, we apply Thms. 15 and 16 to get that with probability  $1 - \delta$ , for any posterior  $Q$

$$|M_m(Q)| \leq \frac{\text{KL}(Q, P) + \log(4/\delta)}{\lambda_m} + \frac{Cm\lambda_m}{\varepsilon_m} \left(1 + \frac{2K}{\delta}\right).$$

Dividing by  $m$  and taking

$$\lambda_m = \sqrt{\frac{(\log(K) + \log(4/\delta)) \varepsilon_m}{Cm \left(1 + \frac{2K}{\delta}\right)}}$$

concludes the proof. ■