Concentration inequalities for functions of independent random variables

Antoine Marchina

NNT : 2017SACLV068

THÈSE DE DOCTORAT

de

l'Université Paris-Saclay

École doctorale de mathématiques Hadamard (EDMH, ED 574)

Établissement d'inscription : Université de Versailles Saint-Quentin-en-Yvelines

Laboratoire d'accueil : Laboratoire de mathématiques de Versailles, UMR 8100 CNRS

Spécialité de doctorat : Mathématiques appliquées

Antoine MARCHINA

Inégalités de concentration pour des fonctions de variables aléatoires indépendantes

Date de soutenance : 8 décembre 2017

Après avis des rapporteurs :	Radosław ADAMCZAK	(University of Warsaw)
	Yannick BARAUD	(Université de Nice Sophia Antipolis)

Jury de soutenance :

Yannick BARAUD	(Université de Nice Sophia Antipolis)	Rapporteur
Bernard BERCU	(Université de Bordeaux)	Président du jury
Catherine DONATI-MARTIN	(Université de Versailles)	Examinatrice
Claire LACOUR	(Université Paris-Sud)	Examinatrice
Emmanuel RIO	(Université de Versailles)	Directeur de thèse
Paul-Marie SAMSON	(Université Paris-Est Marne-La-Vallée)	Examinateur

Remerciements

Je ne suis pas sûr que commencer une thèse par les mots suivants soit une bonne idée, mais ça ne fait rien, la plupart de mes proches le savent : je ne suis pas très à l'aise pour manier les mots et je suis plutôt concis dans ce que j'écris. J'aimerais pouvoir écrire des remerciements à la hauteur de ce que je souhaite témoigner aux différentes personnes qui m'ont amené jusqu'ici, mais on ne change pas du jour au lendemain (c'est d'autant plus vrai à l'heure où j'écris cette page). J'espère que vous ne m'en tiendrez pas rigueur et que vous savez tous ce qu'au fond je souhaite vous dire.

Tout d'abord je tiens à adresser mes plus chaleureux remerciements à mon directeur de thèse Emmanuel Rio, pour la confiance qu'il m'a accordée en me proposant de m'encadrer pour le mémoire de master puis la thèse, pour sa grande disponibilité, pour sa patience, pour toutes nos discussions, et pour tous les précieux conseils qu'il m'a prodigués tout au long de ce parcours. J'ai beaucoup apprécié la manière dont il a encadré ma thèse et je serai ravi de pouvoir continuer à travailler avec Emmanuel.

Je suis très honoré que Radosław Adamczak et Yannick Baraud aient accepté de rapporter cette thèse. Je les remercie pour leur relecture attentive et pour l'intérêt qu'ils ont porté à mon travail. Leurs nombreux commentaires ont permis d'améliorer très largement ce manuscrit.

Je remercie également Yannick Baraud, Bernard Bercu, Catherine Donati-Martin, Claire Lacour et Paul-Marie Samson pour avoir accepté de faire partie de mon jury. En particulier, je tiens à remercier Claire Lacour et Paul-Marie Samson chacun pour la gentillesse qu'ils m'ont témoignée en prenant du temps pour que l'on discute de mes travaux et d'idées possibles pour la suite.

J'affectionne tout particulièrement l'université de Versailles, puisqu'en plus de ma thèse, j'y ai passé trois superbes années de licence et mon parcours jusque là doit beaucoup à cette période.

Je suis très heureux d'avoir ainsi retrouvé comme collègues mes anciens enseignants et d'en compter maintenant certains parmi mes amis. Je remercie infiniment tous les membres du département de mathématiques pour tous les bons souvenirs que j'ai accumulés pendant ces six années et demie, pour tout ce que j'ai pu apprendre auprès d'eux, du coupage d'epsilon en deux aux inégalités de concentration, mais également pour toutes nos discussions au RU, autour d'un café, dans les couloirs,... Au risque d'en oublier certains (je m'en excuse profondément) j'ai une pensée pour Agnès, Alain, Bernhard avec qui j'ai beaucoup apprécié échanger lorsque j'étais son chargé de TD, Brigitte, Catherine, Christine, Christophe, Emmanuel dont j'appréciais déjà la rigueur pendant le cours de théorie de la mesure, Julien, Laurent, Luc, Nicolas dont je me remémore ses cours captivants où on met les mains dans le cambouis, Oleksiy, Philippe, Pierre, Stéphane, Tahar, Vahagn, Vincent. Merci aussi à Laure, Liliane et Nadège pour toutes leur aide, leur efficacité et leur gentillesse.

J'aimerais remercier plus particulièrement Jean Pian pour ses cours toujours très vivants et pendant lesquels j'ai compris qu'un bon coup de pied pouvait résoudre beaucoup de problèmes mathématiques. Il s'est toujours montré encourageant et a toujours pris le temps pour me guider dans mes choix. Merci aussi à Alexis Devulder qui n'a eu cesse de m'encourager pendant ma dernière année de licence et de me pousser à toujours faire mieux. Je le remercie également de m'avoir fait découvrir le milieu de la recherche en acceptant de m'encadrer dans un (malheureusement trop court) stage au laboratoire. Je les remercie tous deux du fond du coeur.

Je n'oublie pas non plus mes années à Orsay où mon goût pour les probabilités et statistiques s'est largement développé. Je remercie l'ensemble de mes professeurs du laboratoire d'Orsay pour cela.

Merci aussi aux membres du laboratoire de Marne-la-Vallée qui m'ont réservé un excellent accueil en tant qu'ATER cette année.

J'ai maintenant une pensée pour les thésards, ATER et post-docs de Versailles qui ont contribué à maintenir une très bonne ambiance tout au long de ces trois ans et demie. J'ai la chance de dire qu'ils sont pour la plupart bien plus que des collègues, mille mercis pour le soutien, les pauses cafés, les quizz au bombardier, les soirées... Merci à Camilla, ma co-bureau d'en face depuis le début, et la meilleure pâtissière du séminaire des jeunes. Je ne compte plus nos longues discussions, nos fous rires, nos « il faut que je te dise un truc », et allez, je te le dis au moins une fois : oui, tu parles bien français ! Merci à Maxime, c'est dommage que tu n'arrives dans notre bureau qu'au moment où je pars ! Je me souviendrai que tu es aussi habile que moi lorsqu'il s'agit d'écrire un mot sur une carte ! Merci à Ilaria, pour sa bonne humeur, sa gentillesse et évidemment son caractère sarde. Merci à « oula... » Patricio Patricio, peu de fois présent à Versailles mais à chaque fois la journée était mémorable ! Merci à Florian et Sébastien qui ont, à eux deux et en un an, ramené beaucoup de vie et de dynamisne dans le laboratoire ! Merci à Thomas de nos longues discussions, on ne voyait pas le temps passé ! Merci à Colin qui ramène de la jeunesse et des excellents cookies ! J'ai une pensée aussi aux autres doctorants que j'ai eu moins le temps de connaître en dehors mais qui ont toujours amené de la bonne humeur : Hélène, Salim, Sybille ainsi le très sérieux déjà docteur Félix !

Merci à mes amis de plus longues dates à qui je dois beaucoup et qui me soutiennent (et supportent !) depuis plusieurs années : Mathias, Marie, Amaury et Laura, Marc, Dimitri, Olivier, Constance, Benjamin et Marie et leur petite Éloïse.

Je tiens à dire un immense merci à ma famille et ma belle-famille, qui ont toujours été là pour moi, surtout dans les moments importants, et sans qui rien de tout cela aurait été possible. Je leur suis reconnaissant au delà des mots pour tout ce qu'ils m'offrent.

Anne-Charlotte, merci d'être là, toujours. Tout cela n'aurait pas beaucoup d'importance sans toi.

À la mémoire de mon père,

Table des matières

1 Introduction **1**
- 1.1 Introduction générale 1
- 1.2 Inégalités de comparaison 3
  - 1.2.1 Description de la méthode 3
  - 1.2.2 Repères bibliographiques 5
- 1.3 Présentation des travaux de thèse 9
  - 1.3.1 Première partie 9
  - 1.3.2 Seconde partie 11

I Concentration inequalities for separately convex functions **21**

2 Concentration inequalities for separately convex functions **23**
- 2.1 Introduction 23
- 2.2 Main results 26
- 2.3 Concentration inequalities from comparison inequalities in $\mathcal{H}_+^\alpha$ 27
- 2.4 New comparison inequalities 28
- 2.5 Lipschitz functions of independent random vectors 31
  - 2.5.1 Moment inequality 31
  - 2.5.2 A deviation inequality for the bounded case 33
- 2.6 Weighted empirical distribution functions 34
- 2.7 Suprema of randomized empirical processes 36
  - 2.7.1 Case $G = 0$ 38
  - 2.7.2 Case $G \neq 0$ 42
- 2.8 Chaos of order two 46

2.9 Proofs of the results of Sections 2.2 and 2.4 50
2.9.1 Proofs of Section 2.4 50
2.9.2 Proofs of Section 2.2 53

II Concentration inequalities for suprema of empirical processes 57

3 Concentration inequalities for suprema of unbounded empirical processes 59
3.1 Introduction 59
3.2 Definitions and notations 63
3.3 Statement of results 64
3.3.1 Bound of generalized moments of $Z - \mathbb{E}[Z]$ 66
3.3.2 Proofs of the main results 67
3.4 Application to power-type tail 72

4 About the rate function in the concentration inequalities for suprema of bounded empirical processes 79
4.1 Introduction 79
4.2 Notations and preliminary results 81
4.3 Main results 83
4.4 About the rate function $\ell^*_{\mathscr{F}}$ 85
4.4.1 Comments on Large Deviation Principle 85
4.4.2 The case of nondecreasing 1-Lipschitz functions 88
4.5 Proofs 88
4.5.1 Proofs of Section 4.3 88
4.5.2 Proofs of Section 4.4 93

5 Comparison inequalities for suprema of bounded empirical processes 97
5.1 Introduction 97
5.2 Results 99
5.2.1 Two-sided boundedness condition 99
5.2.2 One-sided boundedness condition 100
5.3 Proofs 101

6 An exponential inequality for suprema of empirical processes with heavy-tails on the left 105
6.1 Introduction 105
6.2 Result 106

Chapitre 1

Introduction

1.1 Introduction générale

Cette thèse s'intéresse à l'étude de la concentration autour de la moyenne des fonctions de variables aléatoires indépendantes. Précisément, soit une suite $X_1, \ldots, X_n$ de variables aléatoires indépendantes à valeurs dans un espace mesurable $(\mathcal{X}, \mathcal{F})$ et soit F une fonction mesurable de $\mathcal{X}^n$ dans $\mathbb{R}$. On définit la variable aléatoire

$$Z := F(X_1, \ldots, X_n). \tag{1.1.1}$$

L'objectif est d'obtenir de bonnes majorations des quantités $\mathbb{P}(Z - \mathbb{E}[Z] \geq x)$ et $\mathbb{P}(Z - \mathbb{E}[Z] \leq -x)$ pour tout réel positif x.

La somme $X_1 + \ldots + X_n$ a été le premier cas étudié (voir Bernstein [15], Bennett [4], Hoeffding [25]). Afin d'obtenir des inégalités de concentration pour des fonctions plus générales, de puissants outils ont été développés à partir des années 70. Citons, entre autres, la méthode de martingale (voir Azuma [2], Yurinskii [63], Maurey [35], Milman et Schechtman [38] et les études de McDiarmid [36, 37]), les inégalités de transport initiées par Marton [31, 32, 33], les inégalités isopérimétriques et la méthode d'induction de Talagrand [58, 59, 60], ainsi que la méthode entropique, basée sur des inégalités logarithmiques, introduite par Ledoux [29].

L'origine de ces travaux de thèse concernait l'analyse de la concentration des suprema de processus empiriques, c'est-à-dire :

$$Z := \sup\left\{ \sum_{k=1}^{n} (f(X_k) - \mathbb{E}[f(X_k)]) : f \in \mathscr{F} \right\}, \tag{1.1.2}$$

où $\mathscr{F}$ est une classe de fonctions mesurables de $\mathcal{X}$ dans $\mathbb{R}$. Les méthodes de Talagrand et de Ledoux, mentionnées ci-dessus, se sont révélées être efficaces et très performantes pour l'obtention d'inégalités de concentration pour Z lorsque les fonctions de la classe $\mathscr{F}$ sont bornées. Cependant, dès que l'hypothèse de bornitude est relaxée, les résultats existants ne sont plus aussi bons qu'espérés et les constantes apparaissant dans les inégalités ne semblent pas optimales. L'objectif initial de ces travaux est alors d'améliorer les résultats dans ce cas ci, en adoptant une autre méthode.

D'autre part, dans plusieurs travaux, Pinelis et Bentkus ont obtenu des inégalités fines de comparaison (sur des moments généralisés et sur les queues de distribution), valables notamment pour des accroissements de martingales non nécessairement bornés. Cela a donc motivé l'adoption de la démarche suivante dans toute la thèse : l'association de techniques de martingales avec des inégalités de comparaison. Précisément, posons $\mathcal{F}_0 := \{\varnothing, \Omega\}$ et pour tout $k = 1, \ldots, n$, $\mathcal{F}_k := \sigma(X_1, \ldots, X_k)$. Alors $Z - \mathbb{E}[Z]$ admet la décomposition en martingale suivante (connu comme la décomposition de Doob) :

$$Z - \mathbb{E}[Z] = \sum_{k=1}^{n} \Delta_k \quad \text{où } \Delta_k := \mathbb{E}[Z \mid \mathcal{F}_k] - \mathbb{E}[Z \mid \mathcal{F}_{k-1}]. \tag{1.1.3}$$

L'étape suivante consiste essentiellement à obtenir une inégalité de comparaison du type

$$\mathbb{E}[\varphi(\Delta_k) \mid \mathcal{F}_{k-1}] \leq \mathbb{E}[\varphi(T_k)], \tag{1.1.4}$$

pour toute fonction réelle φ d'une classe de fonctions assez riche (afin de pouvoir optimiser), où T_k est une variable aléatoire de loi connue. On en déduit alors une inégalité de comparaison pour $\mathbb{E}[\varphi(Z - \mathbb{E}[Z])]$ puis pour $\mathbb{P}(Z - \mathbb{E}[Z] \geq x)$ par l'inégalité de Markov et en optimisant sur la classe de fonctions ou en utilisant des résultats (avec constantes explicites) de Pinelis.

Dans la suite de ce chapitre introductif, nous présentons les résultats de comparaison sur lesquels s'appuie cette thèse ainsi qu'un descriptif des résultats existants. Enfin dans une dernière section, nous décrivons les travaux réalisés durant la thèse et l'organisation du manuscrit.

Pour une présentation complète des phénomènes de concentration, nous renvoyons le lecteur aux ouvrages de Dubhashi et Panconesi [20], Boucheron, Lugosi et Massart [17], Bercu, Delyon et Rio [14] et Ledoux [30].

1.2 Inégalités de comparaison

1.2.1 Description de la méthode

Dans toute cette section, X et ξ sont deux variables aléatoires réelles intégrables. On appelle inégalité de comparaison de moments généralisés, toute inégalité :

$$\mathbb{E}[\varphi(X)] \leq \mathbb{E}[\varphi(\xi)], \tag{1.2.1}$$

vraie pour toute fonction φ d'une certaine classe de fonctions réelles $\mathcal{H}$. Si de plus $\mathcal{H}$ est inclus dans l'ensemble des fonctions croissantes positives, par l'inégalité de Markov, on a pour tout $x \in \mathbb{R}$

$$\mathbb{P}(X \geq x) \leq \inf_{\varphi \in \mathcal{H}} \frac{\mathbb{E}[\varphi(X)]}{\varphi(x)} \leq \inf_{\varphi \in \mathcal{H}} \frac{\mathbb{E}[\varphi(\xi)]}{\varphi(x)}. \tag{1.2.2}$$

Ceci nous donne alors une méthode générale pour obtenir une inégalité de concentration pour X : on cherche une variable aléatoire ξ et une classe de fonctions $\mathcal{H}$ vérifiant une inégalité de comparaison (1.2.1), telles que l'on puisse calculer (ou majorer) le membre de droite de (1.2.2).

Lorsque $\mathcal{H}$ est la classe des fonctions exponentielles croissantes, c'est-à-dire $\mathcal{E} := \{x \mapsto e^{tx}, t > 0\}$, on reconnaît la méthode de Cramér-Chernoff. Les inégalités exponentielles classiques, par exemple, pour les sommes de variables aléatoires indépendantes ou plus généralement pour les martingales, sont basées sur cette méthode (voir les chapitres 2 et 3 de [14]). Considérons, par exemple, le résultat suivant démontré par Hoeffding [25, Théorème 2] :

Théorème 1.2.1. *Soit $a_1, \dots, a_n$ des réels positifs vérifiant $\sum_{k=1}^n a_k^2 = 1$. Soit $\varepsilon_1, \dots, \varepsilon_n$ une suite de variables aléatoires indépendantes de loi de Rademacher. Soit Z une variable aléatoire Gaussienne standard et $R_n := \sum_{k=1}^n a_k \varepsilon_k$. Alors pour tout $x \geq 0$,*

$$\mathbb{P}(R_n \geq x) \leq \inf_{t>0} e^{-tx}\, \mathbb{E}[e^{tR_n}] \leq \inf_{t>0} e^{-tx}\, \mathbb{E}[e^{tZ}] = e^{-x^2/2}. \tag{1.2.3}$$

Maintenant, en comparant le membre de droite de (1.2.3) avec le comportement asymptotique connu $\mathbb{P}(Z \geq x) \sim (x\sqrt{2\pi})^{-1}\, e^{-x^2/2}$ quand x tend vers l'infini, il « manque » un facteur d'ordre $1/x$ pour des réels positifs x grands. La cause apparente de cette perte est que la classe des fonctions exponentielles $\mathcal{E}$ n'est pas assez riche. Pour remédier ce problème, on définit une famille de classes de fonctions plus riches : pour tout réel $\alpha > 0$, on pose

$$\mathcal{H}_+^\alpha := \Big\{\varphi : \varphi(u) = \int_{-\infty}^{\infty} (u-t)_+^\alpha\, \mu(dt) \text{ pour une mesure borélienne } \mu \geq 0 \text{ sur } \mathbb{R} \text{ et pour tout } u \in \mathbb{R}\Big\}.$$

Pinelis (voir [41, Proposition 1 (*ii*)] et [43, Proposition 1.1]) a montré les propriétés suivantes concernant les classes de fonctions $\mathcal{H}_+^\alpha$:

Proposition 1.2.2. *(i)* $\mathbb{E}[\varphi(X)] \leq \mathbb{E}[\varphi(\xi)]$ *pour tout* $\varphi \in \mathcal{H}_+^\alpha$ *si et seulement si* $\mathbb{E}[(X-t)_+^\alpha] \leq \mathbb{E}[(\xi-t)_+^\alpha]$ *pour tout* $t \in \mathbb{R}$.

(ii) $0 < \beta \leq \alpha$ *implique* $\mathcal{H}_+^\alpha \subset \mathcal{H}_+^\beta$.

(iii) Soit $\alpha \in \mathbb{N}^*$. *Alors* $f \in \mathcal{H}_+^\alpha$ *si et seulement si* f *est dérivable* $(\alpha-1)$-*fois,* $f^{(\alpha-1)}$ *est convexe sur* $\mathbb{R}$ *et* $\lim_{x\to-\infty} f^{(j)}(x) = 0$ *pour tout* $j = 0, \ldots, \alpha-1$.

(iv) Pour tout $t \in \mathbb{R}$, *tout* $\beta \geq \alpha$ *et tout* $\lambda > 0$, *les fonctions* $x \mapsto (x-t)_+^\beta$ *et* $x \mapsto e^{\lambda(x-t)}$ *appartiennent à* $\mathcal{H}_+^\alpha$.

Au vu de (1.2.2), ces classes sont intéressantes si on peut calculer l'infimum

$$\inf_{\varphi \in \mathcal{H}_+^\alpha} \{(\varphi(x))^{-1}\, \mathbb{E}[\varphi(X)]\}. \tag{1.2.4}$$

La méthode a été donnée par Pinelis [40, 41] ; le point clé est le théorème suivant [41, Théorème 4] (voir aussi [40, Théorème 3.11]).

Théorème 1.2.3. *Soit* $\alpha > 0$. *Soit* ξ *une variable aléatoire réelle telle que la fonction* $x \mapsto \mathbb{P}(\xi \geq x)$ *est log-concave sur* $\mathbb{R}$. *Alors*

$$\inf_{\varphi \in \mathcal{H}_+^\alpha} \frac{\mathbb{E}[\varphi(\xi)]}{\varphi(x)} = P_\alpha(\xi\,;x) := \inf_{t<x} \frac{\mathbb{E}[(\xi-t)_+^\alpha]}{(x-t)^\alpha} \tag{1.2.5}$$

$$\leq c_{\alpha,0}\,\mathbb{P}(\xi \geq x), \tag{1.2.6}$$

où la constante $c_{\alpha,0} := \Gamma(\alpha+1)(e/\alpha)^\alpha$ *est la meilleure possible.*

On peut montrer que la borne exponentielle classique est meilleure que (1.2.6) pour des réels x petits. En revanche, dans les applications statistiques où l'intérêt est porté sur les grandes valeurs de x, la borne (1.2.6) est nettement meilleure que la borne exponentielle.

Remarque 1.2.4 (À propos de la log-concavité)**.** *Pour toute fonction* f *de* $\mathbb{R}$ *dans* $\mathbb{R}_+$, *on peut définir son enveloppe log-concave notée* f^{LC}, *comme la fonction minimale log-concave telle que* $f \leq f^{LC}$. *On définit* $x \mapsto \mathbb{P}^{LC}(\xi \geq x)$ *l'enveloppe log-concave de la fonction* $x \mapsto \mathbb{P}(\xi \geq x)$. *On peut alors enlever l'hypothèse de log-concavité requise dans le théorème ci-dessus en remplaçant* $\mathbb{P}(\xi \geq x)$ *dans* (1.2.6) *par* $\mathbb{P}^{LC}(\xi \geq x)$. *Cependant l'optimalité de la constante* $c_{\alpha,0}$ *n'est plus assurée.*

Dans [48], Pinelis a mené une étude approfondie de $P_\alpha(\xi; x)$. On peut également trouver une description du calcul pour des α dans $\{1, 2, 3\}$ et pour des familles de lois spécifiques (exponentielle, uniforme, normale, binomiale et Poissonnienne) dans Bentkus, Kalosha et van Zuijlen [13].

Étant donné que ces résultats de comparaison de moments généralisés sont intéressants en eux-mêmes, mentionnons quelques extensions à des classes plus riches que $\mathcal{H}_+^\alpha$. Commençons par établir le lien existant entre $\mathcal{H}_+^1$ et la classe des fonctions convexes. On renvoie le lecteur à Bentkus [10, Proposition 3] pour une preuve.

Proposition 1.2.5. *Les propositions suivantes sont équivalentes :*

(i) $\mathbb{E}[X] = \mathbb{E}[\xi]$ *et* $\mathbb{E}[\varphi(X)] \leq \mathbb{E}[\varphi(\xi)]$ *pour tout* $\varphi \in \mathcal{H}_+^1$.

(ii) $\mathbb{E}[\varphi(X)] \leq \mathbb{E}[\varphi(\xi)]$ *pour toute fonction convexe* φ.

Enfin, Pinelis [49, Corollaire 5.8] a obtenu un énoncé général permettant de passer d'une inégalité de comparaison dans $\mathcal{H}_+^\alpha$ avec $\alpha \in \mathbb{N}^*$, à une inégalité de comparaison dans une classe légérement plus grande sous des conditions de moments.

Proposition 1.2.6. *Soit* $k \leq n$ *deux entiers naturels et définissons la classe de fonctions réelles*

$$\mathcal{G}^{k:n} := \{\varphi : \varphi \in \mathcal{C}^{n-1} \text{ et } \varphi^{(j)} \text{ est croissante pour tout } j = k-1, \ldots, n\},$$

où $\varphi^{(n)}$ *désigne la dérivée à droite de la fonction convexe* $f^{(n-1)}$*. Alors*

$$\mathbb{E}[\varphi(X)] \leq \mathbb{E}[\varphi(\xi)] \quad \text{pour tout } \varphi \in \mathcal{G}^{k:n},$$

si et seulement si les conditions suivantes sont satisfaites :

(i) $\mathbb{E}[\varphi(X)] \leq \mathbb{E}[\varphi(\xi)] \quad$ *pour tout* $\varphi \in \mathcal{H}_+^n$,

(ii) $\mathbb{E}[X^j] = \mathbb{E}[\xi^j] \quad$ *pour tout* $j = 1, \ldots, k-1$,

(iii) $\mathbb{E}[X^k] \leq \mathbb{E}[\xi^k]$.

1.2.2 Repères bibliographiques

Dans ce manuscrit, notre intérêt se porte sur les inégalités de comparaison de moments généralisés (associés à une classe de fonctions convexes) et de queues de distribution qui en découlent par la méthode décrite. Ainsi, dans cette sous-section, nous dressons un état des résultats obtenus dans cette direction. D'autres méthodes ont été utilisées afin d'obtenir directement des inégalités de comparaison sur les queues de distribution. Un exemple important est l'étude de la constante c dans la généralisation du Théorème 1.2.1 :

$\mathbb{P}(R_n \geq x) \leq c\,\mathbb{P}(Z \geq x)$ où Z est une variable aléatoire Gaussienne standard. Mentionnons, entre autres, les travaux de Bobkov, Götze et Houdré [16], Bentkus [5, 6, 8], Pinelis [45] et Bentkus et Dzindzalieta [12] dans lequel les auteurs obtiennent la meilleure constante possible.

Le concept d'inégalité de comparaison convexe (c'est-à-dire pour la classe des fonctions convexes) a été introduit en 1963 par Hoeffding [25, Section 6]. Dans ce papier, il remarque qu'avec un résultat de comparaison convexe entre X et ξ et un contrôle des moments exponentiels de ξ, alors on obtient de ce fait une inégalité de déviation pour X. En exemple, il montre que la somme de n variables aléatoires issues d'un tirage sans remise au sein d'une population finie est plus concentrée pour la classe des fonctions convexes que la somme de n variables aléatoires issues d'un tirage avec remise au sein de cette même population. Il montre également le résultat suivant (voir les inégalités (4.1) et (4.2) de ce même papier) :

Lemme 1.2.7. *Soit a, b deux réels positifs et X une variable aléatoire tels que $-a \leq X \leq b$. Soit θ une variable aléatoire prenant uniquement les valeurs $-a$ et b et telle que $\mathbb{E}[\theta] = \mathbb{E}[X]$. On a alors*

$$\mathbb{E}[\varphi(X)] \leq \mathbb{E}[\varphi(\theta)] \quad \textit{pour toute fonction convexe } \varphi. \tag{1.2.7}$$

Hoeffding avait déjà étudié les phénomènes de comparaison convexe en 1956 [24, Théorème 3]. Il a montré que la somme de n variables aléatoires indépendantes de Bernoulli de paramètre p_k est plus concentrée au sens des fonctions strictement convexes qu'une variable aléatoire binomiale de paramètres n et la moyenne arithmétique des p_k. Ce résultat a été étendu aux fonctions convexes en 1975 par Gleser [23, Corollary 2.1].

En 1970, Eaton [21] généralise le Théorème 1.2.1 à une classe de fonctions contenant les fonctions réelles $x \mapsto (|x| - t)_+^3$ pour tout $t > 0$ (voir Pinelis [39, Proposition A.1.]). Puis, en 1974, Eaton [22] (voir son Corollaire 1 et la remarque qui suit) a conjecturé que cette dernière inégalité impliquait l'inégalité de déviation suivante :

$$\mathbb{P}(R_n \geq x) \leq c_{3,0}(\sqrt{2\pi})^{-1}e^{-x^2/2} \quad \text{pour tout } x > \sqrt{2}. \tag{1.2.8}$$

En 1994, Pinelis [39, Théorème 2.2] a prouvé le raffinement suivant de la conjecture d'Eaton :

$$\mathbb{P}(R_n \geq x) \leq c_{3,0}\,\mathbb{P}(Z \geq x) \quad \text{pour tout } x \in \mathbb{R}, \tag{1.2.9}$$

où $c_{3,0} = 2e^3/9$ (définie dans le Théorème 1.2.3). Comme mentionné dans la sous-section précédente, en 1998-99, Pinelis [40, 41] a réalisé que les inégalités ci-dessus provenait de la log-concavité de $x \mapsto \mathbb{P}(Z \geq x)$ et a alors développé la méthode générale précédemment décrite pour extraire une inégalité de comparaison sur les queues de distribution à partir d'une inégalité de comparaison dans $\mathcal{H}_+^\alpha$. Ces travaux de Pinelis ont alors motivé la recherche d'obtention d'inégalités de comparaison dans $\mathcal{H}_+^\alpha$.

En 2004, Bentkus [7] s'intéresse aux martingales dont les accroissements sont au moins majorés. Il obtient le résultat suivant :

Théorème 1.2.8. *Soit $s_1^2, \ldots, s_n^2$ des réels positifs. Soit $M_n = \sum_{k=1}^n X_k$ une martingale par rapport à une filtration $(\mathcal{F}_k)$ telle que $M_0 = 0$ et*

$$X_k \leq 1, \quad \textit{et } \mathbb{E}[X_k^2 \mid \mathcal{F}_{k-1}] \leq s_k^2 \ \textit{p.s.} \tag{1.2.10}$$

Alors

$$\mathbb{E}[\varphi(M_n)] \leq \mathbb{E}[\varphi(T_n)] \quad \textit{pour tout } \varphi \in \mathcal{H}_+^2,$$

où $T_n = \theta_1 + \ldots + \theta_n$ est une somme de variables aléatoires indépendantes telle que θ_k est une variable aléatoire centrée prenant uniquement les deux valeurs $-s_k^2$ et 1.

Ce résultat repose sur la généralisation suivante du Lemme 1.2.7 de Hoeffding pour les variables aléatoires réelles majorées et à variance majorée :

Lemme 1.2.9. *Soit $\sigma^2 > 0$ et $b > 0$. Soit X une variable aléatoire telle que $X \leq b$ et $\mathbb{E}[X^2] \leq \sigma^2$. Soit θ une variable aléatoire prenant uniquement les valeurs $-\sigma^2/b$ et b telle que $\mathbb{E}[\theta] = 0$. On a alors $\mathbb{E}[\theta^2] = \sigma^2$ et*

$$\mathbb{E}[\varphi(X)] \leq \mathbb{E}[\varphi(\theta)] \quad \textit{pour toute fonction } \varphi \in \mathcal{H}_+^2.$$

De plus, en utilisant un argument de Schur-concavité, similaire à celui de Eaton [22], sous la condition $s_1^2 + \ldots + s_n^2 = ns^2 =$ constante, il montre que $\mathbb{E}[\varphi(T_n)]$ est maximale lorsque les θ_k sont identiquement distribuées. Il note également que ce même lemme permet de traiter le cas $X_k \leq b_k$ (où $b_k \geq 0$) et $\mathbb{E}[X_k^2 \mid \mathcal{F}_{k-1}] \leq s_k^2$) de manière similaire, mais le résultat est alors dans $\mathcal{H}_+^3$. Si on remplace (1.2.10) par $-p_k \leq X_k \leq 1 - p_k$, le Lemme 1.2.7 et le résutat mentionné auparavant de Hoeffding [24] permet aussi d'obtenir un résultat comparable dans $\mathcal{H}_+^1$. Citons dès à présent le résultat analogue suivant, obtenu par Pinelis en 2014 [47, Théorème 2.1] (voir aussi la prépublication [46], plus complète).

Théorème 1.2.10. *Sous les conditions du Théorème 1.2.8, si de plus*

$$\mathbb{E}[X_{k+}^3] \leq \beta_k, \quad \textit{et } \beta := \sum_{k=1}^n \beta_k \leq \sum_{k=1}^n s_k^2 := s^2,$$

alors

$$\mathbb{E}[\varphi(M_n)] \leq \mathbb{E}[\varphi(\Gamma_{n(s^2-\beta)} + \tilde{\Pi}_{n\beta})] \quad \textit{pour tout } \varphi \in \mathcal{H}_+^3,$$

où Γ_a *est une variable aléatoire de loi Gaussienne centrée et de variance* a *et* $\tilde{\Pi}_a$ *est une variable aléatoire de Poisson recentrée de paramètre* a.

La comparaison directe entre le résultat de Bentkus et celui de Pinelis n'est pas simple. On renvoie le lecteur à [46] pour une analyse de ces bornes.

Pinelis [42] a généralisé le Théorème 1.2.8 aux surmartingales sous les mêmes hypothèses sur les accroissements. De plus comme $x \mapsto \mathbb{P}(T_n \geq x)$ n'est pas log-concave, il montre comment améliorer la borne $\mathbb{P}^\circ(T_n \geq x)$ que fournit le Théorème 1.2.3 et la remarque qui suit. La même année, dans [43], Pinelis étudie les surmartingales dont les accroissements sont bornés (possiblement asymétriquement) et les compare dans la classe $\mathcal{H}_+^5$ à une loi Gaussienne standard améliorant les résultats de Bentkus [8] (obtenus par une méthode plus directe). De plus, (voir aussi [8] qui est antérieur), il remarque que l'association de techniques de martingales avec les inégalités de comparaison obtenues permet de traiter le cas des fonctions séparément Lipschitz :

Théorème 1.2.11. *Soit* $(E_1, d_1), \ldots, (E_n, d_n)$ *une suite d'espaces métriques de diamètres positifs* $\Delta_1, \ldots, \Delta_n$. *Soit* $E^n := E_1 \times \ldots \times E_n$. *Soit* F *une fonction Borélienne de* E^n *dans* $\mathbb{R}$ *1-Lipschitz :*

$$|F(x_1, \ldots, x_n) - F(y_1, \ldots, y_n)| \leq \sum_{k=1}^{n} d_k(x_k, y_k).$$

Soit $\sigma_n^2 := \Delta_1^2 + \ldots + \Delta_n^2$ *le diamètre de McDiarmid. Soit* $X_1, \ldots, X_n$ *une suite de variables aléatoires indépendantes à valeurs dans* E^n *et soit*

$$Z := F(X_1, \ldots X_n).$$

Alors

$$\mathbb{E}[\varphi(Z - \mathbb{E}[Z])] \leq \mathbb{E}[\varphi((\sigma_n/2)Z)] \quad \textit{pour tout } \varphi \in \mathcal{H}_+^5, \tag{a}$$

où Z *est une variable aléatoire Gaussienne standard. En conséquence,*

$$\mathbb{P}(Z - \mathbb{E}[Z] \geq \sigma_n x) \leq c_{5,0}\,\mathbb{P}(Z \geq 2x). \tag{b}$$

Ensuite, Pinelis [44] remplace la condition de bornitude asymétrique sur les X_k par un nouveau type de condition de bornitude asymétrique (ne portant pas directement sur les X_k). Il obtient une comparaison dans $\mathcal{H}_+^3$ d'une surmartingale avec une somme de variables aléatoires indépendantes et identiquement distribuées de loi à deux points et donne des applications. Ce résultat permet également de généraliser le théorème de Bentkus [7, Théorème 1.3].

Enfin, dans [50], Pinelis s'intéresse à la déviation à gauche de sommes de variables aléatoires positives satisfaisant une certaine condition par rapport à une filtration croissante (le cas indépendant en fait partie) et obtient des résultats de comparaison similaires aux résultats précédents dans les classes « réfléchies » de $\mathcal{H}_+^\alpha$ (c'est-à-dire l'ensemble des fonctions $x \mapsto \varphi(-x)$ où $\varphi \in \mathcal{H}_+^\alpha$).

Dans la plupart des travaux ayant pour objectif d'améliorer ou d'étendre les résultats de Hoeffding [25] (y compris les résultats mentionnés ci-dessus), peu concernent le cas des variables aléatoires non bornées. En 2008-10, Bentkus [9, 10, 11] a developpé une méthode afin de relaxer ces hypothèses de bornitude. L'idée générale est de remplacer les bornes déterministes par des dominations stochastiques par des variables aléatoires dont on connaît la loi. Bentkus a montré (dans divers cas) que l'on pouvait définir une variable aléatoire ξ à partir des variables aléatoires dominantes telle que $\mathbb{E}[\varphi(X)] \leq \mathbb{E}[\varphi(\xi)]$ pour φ dans $\mathcal{H}_+^\alpha$ avec $\alpha \in \{1, 2, 3\}$ selon les cas considérés. En particulier, il obtient :

Lemme 1.2.12. *Soit X, Y et Z trois variables aléatoires réelles telles que pour tout $x \in \mathbb{R}$,*

$$\mathbb{P}(Y \geq x) \leq \mathbb{P}(X \geq x) \leq \mathbb{P}(Z \geq x).$$

Alors

$$\mathbb{E}[\varphi(X)] \leq \mathbb{E}[\varphi(\xi)] \quad \text{pour tout fonction convexe } \varphi,$$

où ξ est une variable aléatoire de loi entièrement définie par celles de Y et Z et telle que $\mathbb{E}[\xi] = \mathbb{E}[X]$. (Une formule explicite de la loi de ξ est donnée).

1.3 Présentation des travaux de thèse

1.3.1 Première partie

Dans la première partie, nous nous intéressons à $Z := F(X_1, \ldots, X_n)$ où $X_1, \ldots, X_n$ sont des variables aléatoires à valeurs dans un espace vectoriel $(E, \|.\|)$, F est une fonction de E^n dans $\mathbb{R}$ et nous ne supposons pas les variables aléatoires $\|X_k\|$ bornées. Nous appliquons la décomposition en martingale (1.1.3) en remarquant de plus, que pour toute variable aléatoire A_{k-1}, $\mathcal{F}_{k-1}$-mesurable,

$$\Delta_k = Z_k - A_{k-1} - \mathbb{E}[Z_k - A_{k-1} \mid \mathcal{F}_{k-1}] \quad \text{où } Z_k := \mathbb{E}[Z \mid \mathcal{F}_k]. \qquad (1.3.1)$$

Cett décomposition avait déjà été faite par Pinelis et Sakhanenko [51] pour la norme de la somme. Pour remplacer l'absence de condition de bornitude déterministe sur les X_k, on suppose que l'on a une condition stochastique $-\xi_k \preceq Z_k - A_{k-1} \preceq \xi_k$ pour des variables aléatoires positives ξ_k. Maintenant, afin d'obtenir une inégalité du type (1.1.4), on se restreint d'abord au cas d'une variable aléatoire réelle vérifiant les mêmes hypothèses que les accroissements. Nous avons dans un premier temps, et à l'origine sans connaissance des travaux de Bentkus [9, 10, 11], ré-obtenu le Lemme 1.2.12 (voir Lemme 2.4.3) et nous avons également prouvé une extension dans le cas de domination symétrique (voir Lemme 2.4.6). On remarque alors que pour appliquer ce dernier lemme à Δ_k (conditionnellement à $\mathcal{F}_{k-1}$), une condition est naturellement satisfaite si l'on suppose que F est séparément convexe. On obtient ainsi le résultat suivant qui est le résultat principal de cette première partie (voir Théorème 2.2.1) :

Théorème 1

Soit $X_1, \dots, X_n$ une suite de variables aléatoires centrées à valeurs dans un espace vectoriel $(E, \|.\|)$, $F : E^n \to \mathbb{R}$ une fonction mesurable séparément convexe et Z définie par (1.1.1)*. Soit*

$$Z^{(k)} := F(X_1, \dots, X_{k-1}, 0, X_{k+1}, \dots, X_n).$$

Supposons que pour tout $k = 1, \dots, n$, il existe des variables aléatoires T_k et W_k, positives, de carré intégrables, telles que $\mathbb{E}_k[T_k]$ et $\mathbb{E}_k[W_k]$ sont indépendantes de $\mathcal{F}_{k-1}$, et

$$-T_k \leq Z - Z^{(k)} \leq W_k \quad p.s.$$

Soit $\xi_1, \dots, \xi_n$ une suite de variables aléatoires positives vérifiant, pour tout $t \in \mathbb{R}$,

$$\max\Big(\mathbb{E}[(T_k - t)_+],\ \mathbb{E}[(W_k - t)_+]\Big) \leq \mathbb{E}[(\xi_k - t)_+].$$

Alors

$$\mathbb{E}[\varphi(Z - \mathbb{E}[Z])] \leq \mathbb{E}\Big[\varphi\Big(\sum_{k=1}^{n} \varepsilon_k Q_{\xi_k}(U_k/2)\Big)\Big] \quad \textit{pour tout } \varphi \in \mathcal{H}_+^2,$$

où $Q_Y(u)$ désigne le $(1-u)$-quantile pour toute variable aléatoire réelle Y, $\varepsilon_1, \dots, \varepsilon_n$ sont des variables aléatoires indépendantes de Rademacher,

$U_1, \ldots, U_n$ sont des variables aléatoires indépendantes de loi uniforme sur $[0,1]$, et ces deux familles sont indépendantes.

La suite du chapitre est consacrée à plusieurs applications de ce résultat : les fonctions F séparément convexes et séparément Lipschitz, les fonctions de répartition empiriques à poids, les suprema de processus empiriques randomisés ainsi que les chaos d'ordre deux.

1.3.2 Seconde partie

La seconde partie de la thèse est uniquement centrée sur les suprema de processus empiriques définis par (1.1.2) et associés à une suite de variables aléatoires indépendantes identiquement distribuées selon une loi P.

Dans le Chapitre 3, nous cherchons à obtenir des inégalités de concentration lorsque $\mathscr{F}$ est une classe de fonctions non nécessairement bornées possédant une enveloppe de carré intégrable Φ, c'est à dire,

$$|f| \leq \Phi \text{ pour tout } f \in \mathscr{F}, \text{ et } P(\Phi^2) < \infty.$$

Notons tout d'abord la raison pour laquelle appliquer la méthode « directe » de la première partie n'est pas souhaitable ici. Prenons l'exemple traité dans le Chapitre 2, des suprema de processus empiriques randomisés, c'est-à-dire :

$$Z = \sup_{f \in \mathscr{F}} \sum_{k=1}^{n} Y_k f(X_k),$$

où $X_1, \ldots, X_n$ est une suite de variables aléatoires indépendantes à valeurs dans un espace mesurable $(\mathcal{X}, \mathcal{F})$, $Y_1, \ldots, Y_n$ est une suite de variables aléatoires réelles symétriques telles que les deux suites sont indépendantes et $\mathscr{F}$ est une classe de fonctions mesurables de $\mathcal{X}$ dans $\mathbb{R}$ possédant une enveloppe de carré intégrable. Les bornes obtenues sur les moments de $Z - \mathbb{E}[Z]$ par le Théorème 1 font uniquement intervenir la fonction enveloppe de $\mathscr{F}$ et permettent de dériver des résultats satisfaisants dans plusieurs cas (voir la section 7 du Chapitre 2). Cependant, pour les suprema de processus empiriques, l'objectif est d'obtenir les extensions exactes des résultats connus sur les processus Gaussiens (qui sont la limite en loi des processus empiriques associés à une classe Donsker) ou des résultats connus sur les sommes de variables aléatoires indépendantes (qui correspondent au cas $\mathscr{F}$ réduit à une seule fonction). Ainsi, on voudrait classiquement avoir dans les bornes de déviations les paramètres $\sigma^2 = \sup_{f \in \mathcal{F}} \frac{1}{n} \sum_{k=1}^{n} \mathbb{E}[f^2(X_k)]$ (« wimpy variance ») et $\mathbb{E}[Z]$. Un exemple important est l'inégalité de type Bennett suivante obtenue par Bousquet [18], améliorant un résultat antérieur de Rio [54] :

Théorème 1.3.1. *Soit $X_1, \ldots, X_n$ une suite de variables aléatoires indépendantes à valeurs dans un espace mesurable $(\mathcal{X}, \mathcal{F})$ de distribution commune P. Soit $\mathscr{F}$ une classe de fonctions mesurable de $\mathcal{X}$ dans $\mathbb{R}$ telle que $P(f) = 0$ et $f \leq 1$ pour tout $f \in \mathscr{F}$. Soit Z définie par* (1.1.2) *et posons $v_n := n\sigma^2 + 2\,\mathbb{E}[Z]$. Soit h la fonction définie, pour tout $u \geq -1$, par $h(u) := (1+u)\log(1+u) - u$. Alors pour tout $t \geq 0$,*

$$\mathbb{P}(Z - \mathbb{E}[Z] \geq t) \leq \exp\Big(- v_n\, h\Big(\frac{t}{v_n}\Big)\Big).$$

Ce type de résultat avait été obtenu initialement par Talagrand [59] au moyen d'inégalités isopérimétriques pour les mesures produits. Ensuite, Ledoux [29] a montré que ces inégalités pouvaient être obtenues à partir de la méthode entropique. Ces travaux ont été le point de départ d'une série de papiers dont l'objectif est d'améliorer les constantes dans ces inégalités et de relaxer les hypothèses : Massart [34], Rio [53, 54, 55], Klein[26], [18], Klein et Rio [27], Boucheron & al. [17]. Une approche par les inégalités de transports a également été développée par Samson [57].

Ainsi, dans le Chapitre 3, nous analysons plus finement les accroissements Δ_k. Nos résultats font intervenir les paramètres σ^2, la fonction enveloppe de $\mathscr{F}$ et les espérances

$$E_k := \mathbb{E} \sup_{f \in \mathscr{F}} k^{-1} \sum_{j=1}^{k} f(X_j). \tag{1.3.2}$$

Les résultats existants dans le cas non borné font intervenir en plus de σ^2 et de la fonction enveloppe :

— soit la weak variance $\Sigma^2 := \mathbb{E}[\sup_{f\in\mathscr{F}} \sum_{k=1}^{n} f^2(X_k)]$ qui est, contrairement au cas borné, difficile à comparer avec σ^2 (voir la Section 15 dans Boucheron & al. [17]),

— soit ne font pas intervenir d'autres paramètres, mais alors concernent la déviation de Z autour de $(1+\eta)\mathbb{E}[Z]$ avec $\eta > 0$ et les constantes sont non explicites (voir Adamczak [1], van de Geer et Lederer [61, 28]).

Décrivons maintenant notre approche : la structure de processus empiriques nous permet de réécrire la décomposition du Chapitre 2 sous la forme

$$\Delta_k = \xi_k + r_k - \mathbb{E}[r_k \mid \mathcal{F}_{k-1}],$$

où ξ_k est une variable aléatoire de moyenne nulle conditionnellement à $\mathcal{F}_{k-1}$ et r_k est une variable aléatoire positive. Nous étudions alors séparément les deux martingales $\sum \xi_k$ et $\sum(r_k - \mathbb{E}[r_k \mid \mathcal{F}_{k-1}])$ à l'aide d'inégalités de type Fuk-Nagaev obtenues par Courbot [19] et Rio [56]. La première est la partie principale et son étude est directe. L'analyse de la deuxième, partie corrective, se fait à l'aide d'inégalités de comparaison (analogue au Lemme 1.2.12) et grâce au lemme suivant qui repose sur une propriété d'échangeabilité des variables (voir Lemme 3.3.10) :

Lemme 2

On a $0 \leq \mathbb{E}_{k-1}[r_k] \leq E_{n-k+1}$ *p.s.*

Ce lemme est en fait l'outil fondamental de toute cette seconde partie. On obtient alors les résultats suivants (voir Théorèmes 3.3.2 et 3.3.3) :

Théorème 3

Soit $X_1, \ldots, X_n$ *une suite de variables aléatoires à valeurs dans un espace mesurable* $(\mathcal{X}, \mathcal{F})$ *de distribution commune* P. *Soit* $\mathscr{F}$ *une classe de fonctions mesurables de* $\mathcal{X}$ *dans* $\mathbb{R}$ *telle que* $P(f) = 0$ *pour tout* $f \in \mathscr{F}$. *Soit* Φ *une fonction enveloppe de* $\mathscr{F}$ *de carré intégrable. Soit* Z *définie par* (1.1.2). *Soit* $x > 0$. *Pour tout* $s > 0$, *on a :*

$$\mathbb{P}((Z - \mathbb{E}[Z])_+ \geq x) \leq \left(1 + \frac{x^2}{sn\sigma^2}\right)^{-s/2} + \left(1 + \frac{x^2}{sV_n}\right)^{-s/2} + 2n\,\mathbb{P}\Big(\Phi(X_1) \geq \frac{x}{2s}\Big).$$

Si $\Phi(X_1)$ *possède un moment faible d'ordre* $\ell > 2$, *pour tout* $u \in]0, 1[$,

$$\mathbb{P}\Big(Z - \mathbb{E}[Z] > \sqrt{2\log(1/u)}\,\big(\sigma\sqrt{n} + \sqrt{V_n}\,\big) + 3\, n^{1/\ell}\mu_\ell\, \Lambda_\ell^+(\Phi(X_1))u^{-1/\ell}\Big) \leq u.$$

$\Lambda_\ell^+(Y)$ désigne le moment faible d'ordre ℓ pour toute variable aléatoire réelle Y, $V_n = \sum_{k=1}^n \mathbb{E}[\xi_k^2]$ où pour tout $k = 1, \ldots, n$, ξ_k est une variable aléatoire dont la loi est définie par Φ et E_k, et telle que V_n/n converge vers 0 lorsque n tends vers 0 (on donne une expression explicite de V_n). On donne également une application aux suprema de processus empiriques satisfaisant une condition de queues de distribution lourdes.

Ensuite, dans le Chapitre 4, nous remarquons que la démarche précédente peut être utilisée pour obtenir des résultats dans la bande de grandes déviations dans le cas des classes de fonctions uniformément bornées, étendant ainsi les résultats obtenus par Rio [53] pour les suprema de processus empiriques indexés par une classe d'ensembles. Ce dernier résultat est le seul à notre connaissance dans cette direction. Le résultat principal est le suivant (voir Théorème 4.3.1) :

Théorème 4

Soit $X_1, \ldots, X_n$ une suite de variables aléatoires indépendantes à valeurs dans un espace mesurable $(\mathcal{X}, \mathcal{F})$ de distribution commune P. Soit $\mathscr{F}$ une classe de fonctions mesurables de $\mathcal{X}$ dans $[-1, 1]$ telle que $P(f) = 0$ pour tout $f \in \mathscr{F}$. Soit Z définie par (1.1.2). Pour tout $f \in \mathscr{F}$, soit ℓ_f et $\ell_{\mathscr{F}}$ les fonctions définies par

$$\ell_f(t) := \log P(e^{tf}) \text{ et } \ell_{\mathscr{F}}(t) := \sup_{f \in \mathscr{F}} \ell_f(t) \quad \text{pour tout } t \geq 0. \tag{1.3.3}$$

Soit $\bar{E} := n^{-1}(E_1 + \ldots + E_n)$ et $v_n := (\bar{E}/2)(1 - (\bar{E}/2))$. Soit $\theta^{(n)}$ une variable aléatoire centrée prenant uniquement les valeurs $-v_n$ et 1. On dénote par ℓ_{v_n} la transformée log-Laplace de $\theta^{(n)}$. Alors pour tout $x \geq 0$,

$$\mathbb{P}(Z - \mathbb{E}[Z] > n(\ell_{\mathscr{F}}^{*-1}(x) + 2\,\ell_{v_n}^{*-1}(x))) \leq e^{-nx},$$

où ℓ_X^* désigne la transformée de Young de ℓ_X pour toute variable aléatoire réelle X.

Ce résultat est à rapprocher du principe de grandes déviations obtenu par Wu [62] où il montre que si les fonctions de $\mathscr{F}$ sont à valeurs dans $[0, 1]$, alors la fonction de taux est $\inf_{f \in \mathscr{F}} \ell_f^*$. La preuve repose sur les mêmes idées que le résultat du Chapitre 3, en utilisant cette fois le Théorème 1.2.8 de Bentkus pour contrôler la partie corrective. Aussi, un résultat récent de Baraud [3] pour la majoration de $\mathbb{E}[Z]$ lorsque $\mathscr{F}$ est une VC-subgraph classe, nous permet de fournir une borne avec constantes explicites pour $\ell_{v_n}^{*-1}(x)$.

Dans le Chapitre 5, on exploite les Théorèmes 1.2.8 et 1.2.10 (respectivement de Bentkus et Pinelis) lorsque les fonctions de $\mathscr{F}$ sont à valeurs dans $[-a, 1]$ (pour un $a > 0$) ou dans $]-\infty, 1]$. On compare ainsi dans les classes $\mathcal{H}_+^\alpha$ avec $\alpha = 2$ ou 3 selon les cas, $Z - \mathbb{E}[Z]$ avec une somme de variables aléatoires indépendantes de loi à deux points ou avec une somme d'une loi de Poisson et d'une loi Gaussienne. En particulier, appliqué aux suprema de processus empiriques indexés par des classes d'ensembles, nos résultats étendent un résultat de Rio [53] pour les ensembles de petite mesure sous P. La preuve repose sur une vérification stricte des hypothèses des théorèmes mentionnés, toujours à l'aide du Lemme 2.

Enfin, dans le Chapitre 6, on s'intéresse au cas où les fonctions de $\mathscr{F}$ sont majorées par 1 mais n'ont pas de variance. Le seul résultat à notre connaissance dans cette direction est donné par Rio [52]. Nous avons déjà remarqué précédemment, que si les fonctions possèdent une variance, la démarche adoptée dans le Chapitre 2 n'est pas favorable (puisque l'on n'a alors pas le bon terme de variance dans les bornes de déviations). En revanche, cette méthode s'applique quand même lorsque la variance est infinie. C'est ce qui est exploité dans ce Chapitre. Aussi, afin de fournir des constantes explicites, on suppose que, sous P, les fonctions ont une queue de distribution du type $t \mapsto t^p$ pour $1 < p < 2$ sur la gauche. Le résultat principal peut alors être résumé ainsi (voir Théorème 6.2.1) :

Théorème 5

Soit $1 < p < 2$. Soit $X_1, \ldots, X_n$ une suite de variables aléatoires à valeurs dans un espace mesurable $(\mathcal{X}, \mathcal{F})$ de distribution commune P. Soit $\mathscr{F}$ une classe de fonctions mesurables de $\mathcal{X}$ dans $]-\infty, 1]$ telle que pour tout $f \in \mathscr{F}$, $P(f) = 0$ et

$$\mathbb{P}(f(X_1) \leq -t) \leq \min\{t^{-p}, 1\}.$$

Soit Z définie par (1.1.2). Alors pour x suffisamment petit,

$$\mathbb{P}(Z - \mathbb{E}[Z] \geq n^{1/p}x) \leq \exp\Big(- K_p\, x^q \Big(1 + O\Big(x^{2/(p-1)} n^{-1/q}\Big)\Big)\Big), \quad (1.3.4)$$

où $q = p/(p-1)$ est l'exposant conjugué de Hölder de p et K_p est une constante dépendant uniquement de p.

On remarque que ce résultat est analogue au résultat connu suivant lorsque l'on suppose que les fonctions ont une variance finie :

$$\mathbb{P}(Z - \mathbb{E}[Z] \geq \sqrt{n}x) \leq \exp\Big(- \frac{x^2}{2v}\Big(1 + O\Big(xn^{-1/2}\Big)\Big)\Big), \quad (1.3.5)$$

où $v := \sigma^2 + n^{-1}\mathbb{E}[Z]$.

References

[1] R. Adamczak. A tail inequality for suprema of unbounded empirical processes with applications to Markov chains. *Electron. J. Probab.*, 13 :no. 34, 1000–1034, 2008.

[2] K. Azuma. Weighted sums of certain dependent random variables. *Tôhoku Math. J. (2)*, 19 :357–367, 1967.

[3] Y. Baraud. Bounding the expectation of the supremum of an empirical process over a (weak) VC-major class. *Electron. J. Stat.*, 10(2) :1709–1728, 2016.

[4] G. Bennett. Probability inequalities for the sum of independent random variables. *Journal of the American Statistical Association*, 57(297) :33–45, 1962.

[5] V. Bentkus. An inequality for tail probabilities of martingales with bounded differences. *Liet. Mat. Rink.*, 42(3) :323–331, 2002.

[6] V. Bentkus. An inequality for tail probabilities of martingales with differences bounded from one side. *Journal of Theoretical Probability*, 16(1) :161–173, 2003.

[7] V. Bentkus. On Hoeffding's inequalities. *Ann. Probab.*, 32(2) :1650–1673, 04 2004.

[8] V. Bentkus. On measure concentration for separately Lipschitz functions in product spaces. *Israel J. Math.*, 158 :1–17, 2007.

[9] V. Bentkus. An extension of the Hoeffding inequality to unbounded random variables. *Lithuanian Mathematical Journal*, 48(2) :137–157, 2008.

[10] V. Bentkus. Addendum to "An extension of an inequality of Hoeffding to unbounded random variables" : The non-i.i.d. case. *Lithuanian Mathematical Journal*, 48(3) :237–255, 2008.

[11] V. Bentkus. Bounds for the stop loss premium for unbounded risks under the variance constraints. Preprint on https ://www.math.uni-bielefeld.de/sfb701/preprints/view/423, 2010.

[12] V. Bentkus and D. Dzindzalieta. A tight gaussian bound for weighted sums of rademacher random variables. *Bernoulli*, 21(2) :1231–1237, 05 2015.

[13] V. Bentkus, N. Kalosha, and M. Zuijlen. On domination of tail probabilities of (super)martingales : Explicit bounds. *Lithuanian Mathematical Journal*, 46(1) :1–43, 2006.

[14] B. Bercu, B. Delyon, and E. Rio. *Concentration Inequalities for Sums and Martingales.* SpringerBriefs in Mathematics. Springer, 2015.

[15] S. N. Bernstein. The theory of probabilities, 1946.

[16] S.G. Bobkov, F. Götze, and C. Houdré. On Gaussian and Bernoulli covariance representations. *Bernoulli*, 7(3) :439–451, 06 2001.

[17] S. Boucheron, G. Lugosi, and P. Massart. *Concentration Inequalities : A Nonasymptotic Theory of Independence*. OUP Oxford, 2013.

[18] O. Bousquet. Concentration inequalities for sub-additive functions using the entropy method. In *Stochastic inequalities and applications*, volume 56 of *Progr. Probab.*, pages 213–247. Birkhäuser, Basel, 2003.

[19] B. Courbot. Rates of convergence in the functional CLT for martingales. *Academie des Sciences Paris Comptes Rendus Serie Sciences Mathematiques*, 328 :509–513, March 1999.

[20] D.P. Dubhashi and A. Panconesi. *Concentration of measure for the analysis of randomized algorithms*. Cambridge University Press, 2009.

[21] M.L. Eaton. A note on symmetric bernoulli random variables. *Ann. Math. Statist.*, 41(4) :1223–1226, 08 1970.

[22] M.L. Eaton. A probability inequality for linear combinations of bounded random variables. *Ann. Statist.*, 2(3) :609–614, 05 1974.

[23] L.J. Gleser. On the distribution of the number of successes in independent trials. *Ann. Probab.*, 3(1) :182–188, 02 1975.

[24] W. Hoeffding. On the distribution of the number of successes in independent trials. *Ann. Math. Statist.*, 27(3) :713–721, 09 1956.

[25] W. Hoeffding. Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.*, 58 :13–30, 1963.

[26] T. Klein. Une inégalité de concentration à gauche pour les processus empiriques. *C. R. Math. Acad. Sci. Paris*, 334(6) :501–504, 2002.

[27] T. Klein and E. Rio. Concentration around the mean for maxima of empirical processes. *Ann. Probab.*, 33(3) :1060–1077, 05 2005.

[28] J. Lederer and S. van de Geer. New concentration inequalities for suprema of empirical processes. *Bernoulli*, 20(4) :2020–2038, 11 2014.

[29] M. Ledoux. On Talagrand's deviation inequalities for product measures. *ESAIM : Probability and Statistics*, 1 :63–87, 1997.

[30] M. Ledoux. *The concentration of measure phenomenon*, volume 89 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2001.

[31] K. Marton. A simple proof of the blowing-up lemma. *IEEE Trans. Inform. Theory*, 32(3) :445–446, 1986.

[32] K. Marton. Bounding $\overline{d}$-distance by informational divergence : a method to prove measure concentration. *Ann. Probab.*, 24(2) :857–866, 1996.

[33] K. Marton. A measure concentration inequality for contracting Markov chains. *Geom. Funct. Anal.*, 6(3) :556–571, 1996.

[34] P. Massart. About the constants in Talagrand's concentration inequalities for empirical processes. *Ann. Probab.*, 28(2) :863–884, 2000.

[35] B. Maurey. Construction de suites symétriques. *C. R. Acad. Sci. Paris Sér. A-B*, 288(14) :A679–A681, 1979.

[36] C. McDiarmid. On the method of bounded differences. *Surveys in combinatorics*, 141(1) :148–188, 1989.

[37] C. McDiarmid. Concentration. In *Probabilistic methods for algorithmic discrete mathematics*, pages 195–248. Springer, 1998.

[38] V.D. Milman and G. Schechtman. *Asymptotic theory of finite dimensional normed spaces*, volume 1200 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 1986. With an appendix by M. Gromov.

[39] I. Pinelis. Extremal probabilistic problems and Hotelling's T^2 test under a symmetry condition. *Ann. Statist.*, 22(1) :357–368, 1994.

[40] I. Pinelis. Optimal tail comparison based on comparison of moments. In *High dimensional probability*, pages 297–314. Springer, 1998.

[41] I. Pinelis. Fractional sums and integrals of r-concave tails and applications to comparison probability inequalities. *Advances in stochastic inequalities (Atlanta, GA, 1997)*, 234 :149–168, 1999.

[42] I. Pinelis. Binomial upper bounds on generalized moments and tail probabilities of (super) martingales with differences bounded from above. In *High dimensional probability*, pages 33–52. Institute of Mathematical Statistics, 2006.

[43] I. Pinelis. On normal domination of (super)martingales. *Electron. J. Probab.*, 11 :no. 39, 1049–1070, 2006.

[44] I. Pinelis. Exact inequalities for sums of asymmetric random variables, with applications. *Probability Theory and Related Fields*, 139(3) :605–635, 2007.

[45] I. Pinelis. Toward the best constant factor for the Rademacher-Gaussian tail comparison. *ESAIM : Probability and Statistics*, 11 :412–426, 2007.

[46] I. Pinelis. On the Bennett–Hoeffding inequality. Preprint. Available at arXiv:0902.4058 [math.PR], 2009.

[47] I. Pinelis. On the Bennett–Hoeffding inequality. *Ann. Inst. H. Poincaré Probab. Statist.*, 50(1) :15–27, 02 2014.

[48] I. Pinelis. An optimal three-way stable and monotonic spectrum of bounds on quantiles : a spectrum of coherent measures of financial risk and economic inequality. *Risks*, 2(3) :349–392, 2014.

[49] I. Pinelis. Convex cones of generalized multiply monotone functions and the dual cones. *Banach J. Math. Anal.*, 10(4) :864–897, 10 2016.

[50] I. Pinelis. Optimal binomial, Poisson, and normal left-tail domination for sums of nonnegative random variables. *Electron. J. Probab.*, 21 :19 pp., 2016.

[51] I.F. Pinelis and A.I. Sakhanenko. Remarks on inequalities for large deviation probabilities. *Theory Probab. Appl.*, 30 :143–148, 1986.

[52] E. Rio. Inégalités exponentielles pour les processus empiriques. *Comptes Rendus de l'Académie des Sciences - Series I - Mathematics*, 330(7) :597 – 600, 2000.

[53] E. Rio. Inégalités de concentration pour les processus empiriques de classes de parties. *Probab. Theory Related Fields*, 119(2) :163–175, 2001.

[54] E. Rio. Une inégalité de Bennett pour les maxima de processus empiriques. *Ann. Inst. H. Poincaré Probab. Statist.*, 38(6) :1053–1057, 2002. En l'honneur de J. Bretagnolle, D. Dacunha-Castelle, I. Ibragimov.

[55] E. Rio. Sur la fonction de taux dans les inégalités de Talagrand pour les processus empiriques. *C. R. Math. Acad. Sci. Paris*, 350(5-6) :303–305, 2012.

[56] E. Rio. About the constants in the Fuk-Nagaev inequalities. *Electron. Commun. Probab.*, 22 :12 pp., 2017.

[57] P.-M. Samson. Infimum-convolution description of concentration properties of product probability measures, with applications. *Annales de l'Institut Henri Poincare (B) Probability and Statistics*, 43(3) :321 – 338, 2007.

[58] M. Talagrand. Concentration of measure and isoperimetric inequalities in product spaces. *Publications Mathematiques de l'IHES*, 81(1) :73–205, 1995.

[59] M. Talagrand. New concentration inequalities in product spaces. *Invent. Math.*, 126(3) :505–563, 1996.

[60] M. Talagrand. A new look at independence. *The Annals of probability*, pages 1–34, 1996.

[61] S. van de Geer and J. Lederer. The Bernstein-Orlicz norm and deviation inequalities. *Probab. Theory Related Fields*, 157(1-2) :225–250, 2013.

[62] L. Wu. Large deviations, moderate deviations and LIL for empirical processes. *Ann. Probab.*, 22(1) :17–27, 1994.

[63] V.V. Yurinskii. Exponential bounds for large deviations. *Theory of Probability & Its Applications*, 19(1) :154–155, 1974.

Part I

Concentration inequalities for separately convex functions

Chapter 2

Concentration inequalities for separately convex functions

We provide new comparison inequalities for separately convex functions of independent random variables. Our method is based on the decomposition in Doob martingale. However we only impose that the martingale increments are stochastically bounded. For this purpose, building on the results of Bentkus ([4], [5], [6]), we establish comparison inequalities for random variables stochastically dominated from below and from above. We illustrate our main results by showing how they can be used to derive deviation or moment inequalities for functions which are both separately convex and separately Lipschitz, for weighted empirical distribution functions, for suprema of randomized empirical processes and for chaos of order two.

This Chapter is adapted from the work [19].

2.1 Introduction

Let E be a vector space. A function F from E^n into $\mathbb{R}$ is said to be separately convex if it is convex in each coordinate. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $X_1, \ldots, X_n$ be a finite sequence of independent and centered random variables with values in E. Throughout the chapter, F is a measurable separately convex function from E^n to $\mathbb{R}$. In this work, we are concerned with deviation inequalities for the random variable

$$Z := F(X_1, \ldots, X_n). \tag{2.1.1}$$

Before going further, let us introduce some notations which are used in this chapter. Set $\mathcal{F}_0 := \{\varnothing, \Omega\}$ and for all $k = 1, \dots, n$, $\mathcal{F}_k := \sigma(X_1, \dots, X_k)$ and $\mathcal{F}_n^k := \sigma(X_1, \dots, X_{k-1}, X_{k+1}, \dots, X_n)$. Let $\mathbb{E}_k$ (respectively $\mathbb{E}_n^k$) denote the conditional expectation operator associated to $\mathcal{F}_k$ (resp. $\mathcal{F}_n^k$). Set also

$$Z_k := \mathbb{E}_k[Z], \tag{2.1.2}$$

$$Z^{(k)} := F(X_1, \dots, X_{k-1}, 0, X_{k+1}, \dots, X_n). \tag{2.1.3}$$

Our approach to obtain deviation inequalities is based on the martingale method. The idea is to decompose the random variable $Z - \mathbb{E}[Z]$ as a sum of martingale increments. Precisely, the sequence (Z_k) is an $(\mathcal{F}_k)$-adapted martingale (the Doob martingale associated with $Z - \mathbb{E}[Z]$) and

$$Z - \mathbb{E}[Z] = \sum_{k=1}^{n} \Delta_k, \qquad \text{where } \Delta_k := Z_k - Z_{k-1}.$$

The main problem is to control the increments Δ_k. Classical concentration inequalities for martingales assume that their increments are bounded (see, for example, Chapter 3 of Bercu, Delyon and Rio [8]). In this chapter, our hypotheses on F and on the random variables $X_1, \dots, X_n$ do not imply a deterministic boundedness condition on the martingale increments, but only a symmetric two-sided stochastic one : $-\xi_k \preceq Z_k - Z^{(k)} \preceq \xi_k$, for some stochastic order $\preceq$, where $\xi_1, \dots, \xi_n$ are real-valued nonnegative random variables. Δ_k and $Z_k - Z^{(k)}$ are linked by the following observation:

$$\Delta_k = Z_k - \mathbb{E}_k[Z^{(k)}] - \mathbb{E}_{k-1}[Z_k - \mathbb{E}_k[Z^{(k)}]]. \tag{2.1.4}$$

Note that this observation was already made by Pinelis and Sakhanenko [26] (see their Inequality (9)) when the function F is the norm of the sum. Let us now explain which stochastic order we work with. Let $\alpha > 0$. We define the class $\mathcal{H}_+^\alpha$ of functions φ from $\mathbb{R}$ into $\mathbb{R}$ as follows:

$$\mathcal{H}_+^\alpha := \Big\{\varphi : \varphi(u) = \int_{-\infty}^{\infty} (u-t)_+^\alpha \, \mu(dt) \text{ for some Borel measure } \mu \geq 0 \text{ on } \mathbb{R} \\ \text{and all } u \in \mathbb{R}\Big\}.$$

Here, as usual, $x_+ := x \vee 0 := \max(0, x)$ and $x_+^\alpha := (x_+)^\alpha$ for all reals x. Using the family $\mathcal{H}_+^\alpha$, we define a family of stochastic order by the formula

$$X \underset{\mathcal{H}_+^\alpha}{\preccurlyeq} \xi \quad \text{if} \quad \mathbb{E}[\varphi(X)] \leq \mathbb{E}[\varphi(\xi)] \text{ for all } \varphi \in \mathcal{H}_+^\alpha, \tag{2.1.5}$$

where X and ξ are real-valued random variables. We refer the reader to Pinelis [24] for more on this stochastic order. Our main results in this chapter will be expressed in terms of comparison inequalities with respect to $\underset{\mathcal{H}_+^\alpha}{\preccurlyeq}$ between $Z - \mathbb{E}[Z]$ and a function of $\xi_1, \ldots, \xi_n$.

Concerning general functions of independent random variables, Boucheron, Bousquet, Lugosi and Massart [9] provided general moment inequalities, using an extension of the entropy method proposed by Ledoux [18]. They derived moment inequalities for various functions such as homogeneous tetrahedral polynomials in Rademacher variables or unbounded empirical processes. Recently, Adamczak and Wolff [2] (see their Theorem 1.4) gave a concentration inequality for polynomials of independent sub-Gaussian random variables.
Moreover, if F is separately Lipschitz (E is then assumed equipped with a norm), $Z_k - Z^{(k)}$ satisfies naturally our stochastic boundedness conditions. When F is only separately Lipschitz, a corollary of a result of Pinelis [20] gives that

$$F(X_1, \ldots, X_n) - \mathbb{E}[F(X_1, \ldots, X_n)] \underset{\mathcal{H}_+^1}{\preccurlyeq} 2\sum_{k=1}^{n} \varepsilon_k \|X_k\|,$$

where $\varepsilon_1, \ldots, \varepsilon_n$ is a sequence of independent Rademacher random variables. Kontorovich [16] gave extensions of McDiarmid's inequality for metric spaces with unbounded diameter. He required a sub-Gaussian control of the symmetrized of $\|X_k - X_k'\|$ where X_k' is an independent copy of X_k.
A particular case of separately convex functions is suprema of empirical processes : $F(x_1, \ldots, x_n) = \sup_{t\in\mathcal{T}} \sum_{i=1}^n x_{i,t}$, where $\mathcal{T}$ is a countable index set. Only few results concern concentration inequalities for suprema of unbounded empirical processes : assuming weak tails with respect to suitable Orlicz norms, Adamczak [1], and van de Geer and Lederer [29] obtained exponential bounds. Later van de Geer and Lederer [17] required only weak moment conditions on an envelope of the class of functions and obtained generalized moment inequalities. In this chapter, we will also treat the case of $F(x_1, \ldots, x_n) = \sup_{t\in\mathcal{T}} \sum_{1\le i<j\le n} x_{i,t}x_{j,t}$, which is a particular case of suprema of polynomials in independent random variables.

We shall use the following notations throughout the chapter. The quantile function of a real-valued random variable X which is the general inverse of the nonincreasing and right continuous tail function of X, $\mathbb{P}(X > t)$, is denoted by Q_X. It is defined by

$$Q_X(u) := \inf\{x \in \mathbb{R} : \mathbb{P}(X > x) \le u\}. \tag{2.1.6}$$

Moreover, for $p \ge 1$, let $\mathbb{L}^p$ be the space of real-valued random variables with a finite absolute moment of order p and we denote by $\|X\|_p$ the $\mathbb{L}^p$-norm of X. As usual for any $a := (a_1, \dots, a_n) \in \mathbb{R}^n$ and any real $r \geq 1$, we write

$$\|a\|_r = \Big(\sum_{k=1}^{n} |a_k|^r \Big)^{1/r}, \text{ and } \|a\|_\infty = \max_{1 \leq k \leq n} |a_k|.$$

Finally, for any real function f, we denote by $f(a+)$ (respectively $f(a-)$) the right (resp. left) limit of f at point a.

The chapter is organized as follows. In Section 2.2, we state the main results of this chapter. In Section 2.3, we explain how we can extract a tail comparison inequality from a comparison inequality with respect to the stochastic order associated with the class $\mathcal{H}_+^\alpha$. In Section 2.4, new comparison inequalities for unbounded real-valued random variables are given. The results in this section will allow us to control the increments of the Doob martingale associated to $Z - \mathbb{E}[Z]$. We provide detailed proofs of Sections 2.2 and 2.4 in Section 2.9. We give some applications of the main results in other sections : in Section 2.5 we examine the special case where F is also separately Lipschitz. Section 2.6 considers the weighted empirical distribution functions, Section 2.7 deals with suprema of randomized empirical processes. Finally, Theorems 2.2.1 and 2.2.3 are applied to chaos of order two in Section 2.8.

2.2 Main results

Theorem 2.2.1. *Let Z and $Z^{(k)}$ be defined respectively by* (2.1.1) *and* (2.1.3)*. Assume that for all $k = 1, \dots, n$, there exist nonnegative, square integrable random variables T_k and W_k, such that $\mathbb{E}_k[T_k]$ and $\mathbb{E}_k[W_k]$ are independent of $\mathcal{F}_{k-1}$ and*

$$-T_k \leq Z - Z^{(k)} \leq W_k, \quad \textit{almost surely.} \tag{2.2.1}$$

Let $\xi_1, \dots, \xi_n$ be any finite sequence of nonnegative random variables such that, for any $t \in \mathbb{R}$,

$$\max\Big(\mathbb{E}[(T_k - t)_+], \, \mathbb{E}[(W_k - t)_+] \Big) \leq \mathbb{E}[(\xi_k - t)_+]. \tag{2.2.2}$$

Then

$$Z - \mathbb{E}[Z] \preccurlyeq_{\mathcal{H}_+^2} \sum_{k=1}^{n} \varepsilon_k Q_{\xi_k}(U_k/2), \tag{2.2.3}$$

where $\varepsilon_1, \dots, \varepsilon_n$ are independent Rademacher random variables, $U_1, \dots, U_n$ are independent random variables distributed uniformly on $[0, 1]$ and these two families are independent.

Remark 2.2.2. *Using new results of Pinelis [25] (see his Corollary 5.8), it is straightforward to extend* (2.2.3) *to the larger class of differentiable convex nondecreasing function with a convex derivative.*

In the following result, we relax the assumption (2.2.1) and we instead assume that the bounds have a $\mathcal{F}_n^k$-measurable component.

Theorem 2.2.3. *Let $r > 2$. Let Z and $Z^{(k)}$ be defined respectively by* (2.1.1) *and* (2.1.3)*. Assume that for all $k = 1, \dots, n$, there exist nonnegative, $\mathbb{L}^r$-integrable random variables T_k and W_k such that $\mathbb{E}_k[T_k]$ and $\mathbb{E}_k[W_k]$ are ndependent of $\mathcal{F}_{k-1}$, and nonnegative, $\mathbb{L}^r$-integrable and $\mathcal{F}_n^k$-measurable random variables ψ_k, conditionally independent, with respect to $\mathcal{F}_k$, of T_k and W_k, such that*

$$-T_k\psi_k \le Z - Z^{(k)} \le W_k\psi_k, \quad \textit{almost surely.} \tag{2.2.4}$$

Let $\xi_1, \dots, \xi_n$ be any finite sequence of nonnegative random variables such that, for any $t \in \mathbb{R}$,

$$\max\Big(\mathbb{E}[(T_k - t)_+],\ \mathbb{E}[(W_k - t)_+]\Big) \le \mathbb{E}[(\xi_k - t)_+]. \tag{2.2.5}$$

Then

$$\|(Z - \mathbb{E}[Z])_+\|_r^2 \le (p-1)\sum_{k=1}^n \|\mathbb{E}_{k-1}[\psi_k]\|_r^2\, \|Q_{\xi_k}(U_k/2)\|_r^2, \tag{2.2.6}$$

where $U_1, \dots, U_n$ are independent random variables uniformly distributed on $[0, 1]$.

2.3 Concentration inequalities from comparison inequalities in $\mathcal{H}_+^\alpha$

In this section, we repeat the relevant materials from [22] and [23] without proofs, of how one obtains a deviation inequality from a comparison inequality with respect to the stochastic order associated with the class $\mathcal{H}_+^\alpha$, $\alpha > 0$, such as in Theorem 2.2.1.

First, let us mention some facts about the class $\mathcal{H}_+^\alpha$. It is easy to see that $0 < \beta < \alpha$ implies $\mathcal{H}_+^\alpha \subset \mathcal{H}_+^\beta$. Moreover, for any real t and any positive λ, the functions $x \mapsto (x-t)_+^\alpha$ and $x \mapsto e^{\lambda(x-t)}$ belong to $\mathcal{H}_+^\alpha$. Finally, the following assertions are equivalent:

(*i*) $X \underset{\mathcal{H}_+^\alpha}{\preccurlyeq} \xi$

(*ii*) $\mathbb{E}[(X-t)_+^\alpha] \le \mathbb{E}[(\xi - t)_+^\alpha]$ for all $t \in \mathbb{R}$.

The following is a special case of Theorem 4 of Pinelis [23].

Theorem 2.3.1. *Suppose that $\alpha > 0$, X and ξ are real-valued random variables, and the tail function $x \mapsto \mathbb{P}(\xi \geq x)$ is log-concave on $\mathbb{R}$. Then the comparison inequality $X \underset{\mathcal{H}_+^\alpha}{\preccurlyeq} \xi$ implies that, for all $x \in \mathbb{R}$,*

$$\mathbb{P}(X \geq x) \leq P_\alpha(\xi\,;x) := \inf_{t<x} \frac{\mathbb{E}[(\xi - t)_+^\alpha]}{(x-t)^\alpha} \tag{2.3.1}$$

$$\leq c_{\alpha,0}\,\mathbb{P}(\xi \geq x), \tag{2.3.2}$$

where the constant factor $c_{\alpha,0} := \Gamma(\alpha+1)(e/\alpha)^\alpha$ is the best possible.

Remark 2.3.2. *A thorough study of $P_\alpha(\xi\,;x)$ can be found in Pinelis [24]. See also Bentkus, Kalosha and van Zuijlen [7] for a description of the calculation for specific α and specific families of distribution.*

Remark 2.3.3. *Since the class $\mathcal{H}_+^\alpha$ contains all increasing exponential functions, $P_\alpha(\xi\,;x)$ is also majorized by the exponential bound $\inf_{\lambda>0} e^{-\lambda x}\mathbb{E}[e^{\lambda\xi}]$. For all small enough x, the exponential bound is better than* (2.3.2)*. However, for large values of x, the latter will be significantly better than the exponential one.*

2.4 New comparison inequalities

The purpose of this section is to obtain extensions of an inequality of Hoeffding to unbounded random variables. In particular, Lemma 2.4.6 below will be our main tool to control the increments of the Doob martingale associated to $Z - \mathbb{E}[Z]$. First, let us recall the definition of the usual stochastic order. Let X and Y be two real-valued random variables. X is said to be smaller than Y in the usual stochastic order, denoted by $X \leq_{\mathrm{st}} Y$, if $\mathbb{P}(X \geq x) \leq \mathbb{P}(Y \geq x)$ for all $x \in \mathbb{R}$.

Throughout this section, η and ψ are real-valued random variables such that

$$\eta \in L^1, \quad \psi \in L^2, \text{ and } \quad \eta \leq_{\mathrm{st}} \psi. \tag{2.4.1}$$

We introduce a family of probability distributions related to the distributions of η and ψ. We first recall some classical notations. The distribution function of a real-valued random variable X is denoted by F_X. The generalized inverse of F_X is defined by

$$F_X^{-1}(u) := \inf\{x \in \mathbb{R} : \mathbb{P}(X \leq x) \geq u\}.$$

Definition 2.4.1. *Let assumption* (2.4.1) *hold. For every $q \in]0,1[$, we set $a_q := F_\eta^{-1}(1-q)$, $b_q := F_\psi^{-1}(1-q)$ and let F_q be the distribution function defined by*

$$F_q(x) := \begin{cases} F_\eta(x) & \text{if } x < a_q, \\ 1-q & \text{if } a_q \le x < b_q, \\ F_\psi(x) & \text{if } x \ge b_q. \end{cases}$$

We also set $F_0 := F_\eta$ and $F_1 := F_\psi$. In the following, we always denote by ζ_q a random variable having F_q as distribution function.

Remark 2.4.2. *A similar construction can be found in Bentkus [4], [5] and [6].*

The following bound was obtained by Bentkus [6, Theorem 1] with a little stronger assumption on the stochastic boundedness condition. Indeed Bentkus supposed that $\eta \le_{\mathrm{st}} X \le_{\mathrm{st}} \psi$, which implies our hypothesis (2.4.2).

Lemma 2.4.3. *Let assumption* (2.4.1) *hold. Let ζ_q be as in Definition 2.4.1 and let X be a real-valued integrable random variable such that for any $t \in \mathbb{R}$,*

$$\mathbb{E}[(X-t)_+] \le \mathbb{E}[(\psi-t)_+], \quad \mathbb{E}[(t-X)_+] \le \mathbb{E}[(t-\eta)_+]. \tag{2.4.2}$$

Let q_0 be the highest real in $[0,1]$ such that

$$\int_{1-q_0}^1 (F_\psi^{-1}(u) - F_\eta^{-1}(u))\,du = \mathbb{E}[X] - \mathbb{E}[\eta]. \tag{2.4.3}$$

Then, X and ζ_{q_0} have the same expectation and for any $t \in \mathbb{R}$,

$$\mathbb{E}[(X-t)_+] \le \mathbb{E}[(\zeta_{q_0}-t)_+]. \tag{2.4.4}$$

Consequently, for any convex function φ,

$$\mathbb{E}[\varphi(X-\mathbb{E}[X])] \le \mathbb{E}[\varphi(\zeta_{q_0}-\mathbb{E}[\zeta_{q_0}])]. \tag{2.4.5}$$

Remark 2.4.4. *As noticed by Bentkus ([4], [5], [6]), we can see this lemma as an extension of an inequality of Hoeffding (see Inequalities (4.1) and (4.2) in [14]). Indeed, if η and ψ are two constants, respectively equal to a and b, it easy to see that* (2.4.1) *and* (2.4.2) *imply that $a \le X \le b$ a.s. Then we obtain for all convex functions φ that $\mathbb{E}[\varphi(X)] \le \mathbb{E}[\varphi(\theta)]$ where θ is two-valued random variable taking the values a and b, and such that $\mathbb{E}[X] = \mathbb{E}[\theta]$.*

Remark 2.4.5. *The special case $0 \le X \le_{st} \psi$ was considered by Bentkus ([4], [5]). In [6], Bentkus obtained similar results when $X \le_{st} \psi$ and the variance of X is known.*

The right-hand side of (2.4.5) still depends on the expectation of X by the term $\mathbb{E}[\zeta_{q_0}]$. The next lemma provides a bound in the symmetric case $\eta = -\psi$, which does not depend of $\mathbb{E}[X]$. The drawback is that we have to pick φ in the smaller class of functions $\mathcal{H}_+^2$.

Lemma 2.4.6. *Let ψ and η be two random variables, respectively nonnegative and nonpositive, satisfying* (2.4.1)*. Let ζ_q, a_q and b_q be given by Definition 2.4.1.*

(i) Let $\tilde{q} := \inf\{q \geq 1/2 : b_q + a_q \leq 2\,\mathbb{E}[\zeta_q]\}$. Then for all $t \in \mathbb{R}$,

$$q \mapsto \mathbb{E}[(\zeta_q - \mathbb{E}[\zeta_q] - t)_+^2]$$

is nonincreasing on $[\tilde{q}, 1]$.

(ii) Assume that $\eta = -\psi$ and let X be a real-valued integrable random variable satisfying (2.4.2)*. If $\mathbb{E}[X] \geq 0$, then*

$$X - \mathbb{E}[X] \underset{\mathcal{H}_+^2}{\preccurlyeq} \zeta_{1/2}. \tag{2.4.6}$$

Remark 2.4.7. *We have a better understanding of the random variable $\zeta_{1/2}$ if we observe that it has the same distribution than $\varepsilon\, Q_\psi(U/2)$, where ε is a Rademacher random variable, U is a random variable distributed uniformly on $[0, 1]$ and these random variables are independent.*

The following result is a corollary of the result obtained by Pinelis in [21, Proposition 2.1]. It will be needed in the proof of Theorem 2.2.3 and in Section 2.8.

Proposition 2.4.8. *Let $r > 2$ and let X and Y be random variables in $\mathbb{L}^r$ such that $\mathbb{E}[Y \mid X] = 0$ almost surely. Then*

$$\|(X + Y)_+\|_r^2 \leq \|X_+\|_r^2 + (r - 1)\|Y\|_r^2.$$

Exactly as in Rio [27, Theorem 2.1], we deduce from Proposition 2.4.8 the following inequality by induction on n.

Corollary 2.4.9. *Let $r > 2$ and $(M_n)_{n\geq 0}$ be a sequence of random variables in $\mathbb{L}^r$. Set $\Delta M_k := M_k - M_{k-1}$. Assume that $\mathbb{E}[\Delta M_k \mid M_{k-1}] = 0$ almost surely for any positive k. Then*

$$\|M_{n+}\|_r^2 \leq \|M_{0+}\|_r^2 + (r - 1) \sum_{k=1}^{n} \|\Delta M_k\|_r^2. \tag{2.4.7}$$

2.5 Lipschitz functions of independent random vectors

Throughout this section, we assume that $(E, \|.\|)$ is a separable Banach space. In addition to being separately convex, we suppose that F is separately 1-Lipschitz. Precisely, F satisfies the following Lipschitz type condition:

$$|F(x_1, \dots, x_n) - F(y_1, \dots, y_n)| \leq \sum_{k=1}^{n} \|x_k - y_k\|.$$

Now, $Z = F(X_1, \dots, X_n)$ naturally fulfills the hypotheses (2.2.1) – (2.2.2) of Theorem 2.2.1 with $\xi_k = \|X_k\|$.

2.5.1 Moment inequality

Proposition 2.5.1. *Let $r \geq 2$ and define the function Q by*

$$Q^2(u) := \sum_{k=1}^{n} Q^2_{\|X_k\|}(u).$$

Then

$$\mathbb{E}[(Z - \mathbb{E}[Z])_+^r] \leq \|g\|_r^r \int_0^{1/2} Q^r(u)du, \tag{2.5.1}$$

where g is standard Gaussian random variable.

Example 2.5.2. *Let X be a centered random vector with values in E and $a_1, \dots, a_n$ be deterministic reals. Let $X_1, \dots, X_n$ be n independent copies of X. Define the function F by*

$$Z := F(\widetilde{X}_1, \dots, \widetilde{X}_n) := \Big\| \sum_{k=1}^{n} a_k X_k \Big\|,$$

where $\widetilde{X}_k := a_k X_k$. The definition of the quantile function (2.1.6) *implies that for any $k = 1, \dots, n$, $Q_{\|\widetilde{X}_k\|}(u) = |a_k| Q_{\|X\|}(u)$. Then Proposition 2.5.1 yields for any $r \geq 2$ that*

$$\begin{aligned}\mathbb{E}[(Z - \mathbb{E}[Z])_+^r] &\leq \|g\|_r^r \Big(\sum_{k=1}^{n} a_i^2 \Big)^{r/2} \int_0^{1/2} Q^r_{\|X\|}(u)du \\ &\leq \|g\|_r^r \Big(\sum_{k=1}^{n} a_i^2 \Big)^{r/2} \int_0^{1} Q^r_{\|X\|}(u)du \end{aligned} \tag{2.5.2}$$

We recall now that for any real-valued random variable ξ, $Q_\xi(U)$ has the same distribution as ξ for any random variable U with the uniform distribution over

$[0,1]$, *and that* $\|g\|_r^r = \pi^{-1/2}\, 2^{r/2}\, \Gamma((r+1)/2)$, *where* $\Gamma(.)$ *is the usual Gamma function. Then we get from* (2.5.2),

$$\mathbb{E}[(Z-\mathbb{E}[Z])_+^r] \le \pi^{-1/2}\Big(2\sum_{k=1}^n a_i^2\Big)^{r/2}\mathbb{E}[\|X\|^r]\,\Gamma\Big(\frac{r+1}{2}\Big). \tag{2.5.3}$$

We now apply this result to suprema of empirical processes. Let Y *be a random variable valued in some measurable space* $(\mathcal{X},\mathcal{A})$ *and let* $Y_1,\dots,Y_n$ *be* n *independent copies of* Y. *Let* $\mathscr{F}$ *be a countable class of measurable functions from* $\mathcal{X}$ *into* $\mathbb{R}$ *such that* $\mathbb{E}[f(Y)]=0$ *for all* $f\in\mathscr{F}$. *We assume that* $\mathscr{F}$ *has an* r*-integrable envelope function* Φ, *that is*

$$|f|\le\Phi \quad \textit{for all } f\in\mathscr{F}, \textit{ and} \quad \Phi\in\mathbb{L}^r. \tag{2.5.4}$$

Let $M^r := \mathbb{E}[\Phi^r(Y)]$. *We denote by* $l_\infty(\mathscr{F})$ *the space of all bounded real functions on* $\mathscr{F}$ *equipped with the norm* $\|x\|_{\mathscr{F}} := \sup_{f\in\mathscr{F}}|x(f)|$, *making* $(l_\infty(\mathscr{F}),\|.\|_{\mathscr{F}})$ *a Banach space. We define*

$$Z := F(\widetilde{X}_1,\dots,\widetilde{X}_n) := \sup_{f\in\mathscr{F}}\Big|\sum_{k=1}^n a_k f(Y_k)\Big|,$$

where $\widetilde{X}_k := (a_k f(Y_k))_{f\in\mathscr{F}}$. *We first assume that* $\mathscr{F}$ *is finite which implies that* $(l_\infty(\mathscr{F}),\|.\|_{\mathscr{F}})$ *is separable. The countable case will follow from the finite case by the virtue of the monotone convergence theorem. Then* (2.5.3) *yields*

$$\|(Z-\mathbb{E}[Z])_+\|_r \le \pi^{-1/2r}\sqrt{2\sum_{k=1}^n a_i^2}\; M\Big(\Gamma\Big(\frac{r+1}{2}\Big)\Big)^{1/r}, \tag{2.5.5}$$

where $\Gamma(.)$ *is the usual Gamma function. This result improves Theorem 4.1 of Lederer and van de Geer [17].*

Proof of Proposition 2.5.1. Theorem 2.2.1 applied with $\xi_k = \|X_k\|$ and (2.1.5) specified to $\varphi(x)=x_+^r$ yield that

$$\mathbb{E}[(Z-\mathbb{E}[Z])_+^r] \le \mathbb{E}\Big[\Big(\sum_{k=1}^n \varepsilon_k Q_{\|X_k\|}(U_k/2)\Big)_+^r\Big]. \tag{2.5.6}$$

Since the random variables $\varepsilon_k Q_{\|X_k\|}(U_k/2)$ are symmetric,

$$\mathbb{E}\Big[\Big(\sum_{k=1}^n \varepsilon_k Q_{\|X_k\|}(U_k/2)\Big)_+^r\Big] = \frac{1}{2}\,\mathbb{E}\Big[\Big|\sum_{k=1}^n \varepsilon_k Q_{\|X_k\|}(U_k/2)\Big|^r\Big]. \tag{2.5.7}$$

Conditioning by $\mathcal{F}_n$ and using the classical Khintchine inequality with the best possible constant founded by Whittle (for $r\ge 3$) and Haagerup (for

$r > 0$) (see the Introduction of Figiel et al. [13] and references therein for a statement of these results), one has

$$\frac{1}{2}\,\mathbb{E}\Big[\Big|\sum_{k=1}^{n}\varepsilon_k Q_{\|X_k\|}(U_k/2)\Big|^r\Big] \le \frac{1}{2}\,\|g\|_r^r\,\mathbb{E}\Big[\Big(\sum_{k=1}^{n} Q^2_{\|X_k\|}(U_k/2)\Big)^{r/2}\Big], \tag{2.5.8}$$

where g is a standard Gaussian random variable. Furthermore let us prove the following general lemma:

Lemma 2.5.3. *Let $r \ge 1$. Let $\xi_1, \ldots, \xi_n$ be nonnegative random variables. Then*

$$\|\xi_1 + \ldots + \xi_n\|_r \le \|Q_{\xi_1}(U) + \ldots + Q_{\xi_n}(U)\|_r,$$

where U has the uniform distribution over $[0,1]$.

Proof. First, by Riesz Representation Theorem we get

$$\|\xi_1 + \ldots + \xi_n\|_r = \sup_{\|W\|_{r'} \le 1} |\mathbb{E}[W(\xi_1 + \ldots + \xi_n)]|, \tag{2.5.9}$$

where r' is such that $1/r + 1/r' = 1$. Moreover, observe that

$$\mathbb{E}[|W|(\xi_1 + \ldots + \xi_n)] \le \int_0^1 Q_{|W|}(u) Q_{\xi_1+\ldots+\xi_n}(u)du \tag{2.5.10}$$

$$\le \int_0^1 Q_{|W|}(u)(Q_{\xi_1}(u) + \ldots + Q_{\xi_n}(u))du, \tag{2.5.11}$$

where we use Lemma 2.1 (a) of Rio [28] in (2.5.10) and Lemma 2.1 (c) of [28] in (2.5.11). Finally, combining (2.5.9), (2.5.11) and an application of Riesz Representation Theorem concludes the proof of Lemma 2.5.3. □

Then this lemma applied to the right-hand side of 2.5.8 leads to the inequality:

$$\mathbb{E}\Big[\Big(\sum_{k=1}^{n} Q^2_{\|X_k\|}(U_k/2)\Big)^{r/2}\Big] \le \mathbb{E}\Big[\Big(\sum_{k=1}^{n} Q^2_{\|X_k\|}(U/2)\Big)^{r/2}\Big], \tag{2.5.12}$$

where U is a random variable distributed uniformly on $[0,1]$. Finally, combining (2.5.6) – (2.5.12), one has (2.5.1) which ends the proof. □

2.5.2 A deviation inequality for the bounded case

Consider the bounded case $\|X_k\| \le a_k$ a.s., for some reals $a_k > 0$. Set also $a = (a_1, \ldots, a_n)$. Theorem 2.2.1 with $\xi_k = a_k$ implies that

$$Z - \mathbb{E}[Z] \underset{\mathcal{H}_+^2}{\preccurlyeq} \sum_{k=1}^{n} a_k\varepsilon_k, \tag{2.5.13}$$

where ε_k are i.i.d. Rademacher random variables.

Remark 2.5.4. *Inequality* (2.5.13) *can be directly obtained via Lemma 4.4. of Bentkus [3].*

Proposition 2.5.5. *Let* ℓ *be the function defined by* $\ell(t) := \log(\cosh(t))$ *and let* ℓ^* *denote the Legendre-Fenchel transform of* ℓ*, which is defined for any* $x > 0$ *by* $\ell^*(x) := \sup_{t>0}\{xt - \ell(t)\}$*. Then for any* $x \in [0,1]$*,*

$$\ell^*(x) = \frac{1}{2}\Big((1+x)\log(1+x) + (1-x)\log(1-x)\Big),$$

and

$$\mathbb{P}(Z - \mathbb{E}[Z] \geq \|a\|_1\, x) \leq \exp\Big(-\frac{\|a\|_1^2}{\|a\|_2^2}\,\ell^*(x)\Big). \tag{2.5.14}$$

Proof. Since the exponential function $x \mapsto e^{tx}$ belongs to $\mathcal{H}_+^2$ for any $t > 0$, (2.5.13) ensures that

$$\log\mathbb{E}[\exp(t(Z - \mathbb{E}[Z]))] \leq \log\mathbb{E}\Big[\exp\Big(t\sum_{k=1}^n \varepsilon_k a_k\Big)\Big] = \sum_{k=1}^n \ell(a_k t). \tag{2.5.15}$$

Note that $\ell'(.) = \tanh(.)$ is a concave function on $[0,\infty[$. Now, from (2.5.15), proceeding exactly as in Bercu, Delyon and Rio [8] (see their Inequality (2.98)) we get

$$\log\mathbb{E}[\exp(t(Z - \mathbb{E}[Z]))] \leq \frac{\|a\|_1^2}{\|a\|_2^2}\,\ell\Big(\frac{\|a\|_2^2\, t}{\|a\|_1}\Big). \tag{2.5.16}$$

Finally, (2.5.14) follows from Markov's inequality together with (2.5.16). □

2.6 Weighted empirical distribution functions

Let U be a random variable uniformly distributed on $[0,1]$, $\widetilde{U}_1, \ldots, \widetilde{U}_n$ be n independent copies of U, and denote the uniform empirical process by

$$e_n(t) = \frac{1}{\sqrt{n}}\sum_{k=1}^n (\mathbb{1}_{\widetilde{U}_k \leq t} - t) \quad \text{for any } t \in [0,1].$$

Let $q :]0,1[\to \mathbb{R}$ be a weight function, continuous, such that

$$q(t) = q(1-t), \quad q(t) > 0, \quad \int_0^{1/2}\frac{dt}{q^2(t)} < \infty,$$
$$\text{and} \quad t \mapsto \frac{q(t)}{t} \text{ is nonincreasing.}$$

Remark 2.6.1. *The previous assumptions implies that* $t \mapsto q(t)/(1-t)$ *is nondecreasing.*

Example 2.6.2. *The most common such weight functions* q *are*

$$q(t) = \left(\sqrt{t(1-t)}\right)^{\alpha}, \text{ for any } \alpha \in]0,1[,$$
$$q(t) = \max\left(\sqrt{t(1-t)}, \sqrt{\delta(1-\delta)}\right), \text{ for some } 0 < \delta < 1.$$

In this section, the quantity of interest is

$$Z := \sup_{0<t<1} \frac{e_n(t)}{q(t)}.$$

We refer the reader to Csörgő and Horváth [11] for asymptotic results on this object. Setting now the class of functions $\mathscr{F} := \left\{ \frac{\mathbb{1}_{[0,t]} - t}{q(t)} \,:\, t \in]0,1[\right\}$ and $\widetilde{X}_k := (f(\widetilde{U}_k))_{f \in \mathscr{F}}$, we can write Z as

$$Z = F(\widetilde{X}_1, \ldots, \widetilde{X}_n) := \frac{1}{\sqrt{n}} \sup_{f \in \mathscr{F}} \sum_{k=1}^{n} f(\widetilde{U}_k).$$

Proposition 2.6.3. *We have*

$$Z - \mathbb{E}[Z] \underset{\mathcal{H}_+^2}{\preccurlyeq} \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \varepsilon_k \frac{1 - U_k/2}{q(U_k/2)}, \tag{2.6.1}$$

where $\varepsilon_1, \ldots, \varepsilon_n$ *are independent Rademacher random variables,* $U_1, \ldots, U_n$ *are independent random variables distributed uniformly on* $[0,1]$ *and these two families are independent.*

Remark 2.6.4. *Define* $\ell^\infty(\mathscr{F}) := \{x : \mathscr{F} \to \mathbb{R} : \sup_{f \in \mathscr{F}} |x(f)| < \infty\}$ *equipped with the norm* $\|x\|_{\mathscr{F}} := \sup_{f \in \mathscr{F}} |x(f)|$*. Then, the summands in the right-hand side of* (2.6.1) *are equal to* $\|\widetilde{X}_k\|_{\mathscr{F}}$*, leading to*

$$Z - \mathbb{E}[Z] \underset{\mathcal{H}_+^2}{\preccurlyeq} \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \varepsilon_k \|\widetilde{X}_k\|_{\mathscr{F}}.$$

Remark 2.6.5. *The uniform case also treats the general one. Precisely, let* $X_1, \ldots, X_n$ *be* n *independent copies of a real-valued random variable* X *with a continuous distribution function* F_X*. Then*

$$Z := \frac{1}{\sqrt{n}} \sup_{\substack{t \in \mathbb{R} \\ 0 < F_X(t) < 1}} \frac{\sum_{k=1}^{n} (\mathbb{1}_{X_i \le t} - F_X(t))}{q(F_X(t))} = \sup_{\substack{t \in \mathbb{R} \\ 0 < F_X(t) < 1}} \frac{e_n(F_X(t))}{q(F_X(t))}.$$

Proceeding in the same way as in the proof of Proposition 2.5.1, we obtain the following moment inequality:

Corollary 2.6.6. *Let $\alpha \in]0,1[$, $q(t) = \left(\sqrt{t(1-t)}\right)^\alpha$, and $r \geq 2$ such that $r\alpha < 2$. Then*

$$\mathbb{E}[(Z - \mathbb{E}[Z])_+^r] \leq \|g\|_r^r \int_0^{1/2} (1-u)^{(2-\alpha)(r/2)} u^{-\alpha r/2} \, du, \tag{2.6.2}$$

where g is a standard Gaussian random variable.

Example 2.6.7. *With $r = 2$ and $\alpha = 1/2$,*

$$\mathbb{E}[(Z - \mathbb{E}[Z])_+^2] \leq \frac{1}{2} + \frac{3}{16}\pi \approx 1.089.$$

Proof of Proposition 2.6.3. For any function f in $\mathscr{F}$ and for all $x \in]0,1[$,

$$-\frac{x}{q(x)} \leq f(x) \leq \frac{1-x}{q(x)}.$$

Since $q(t) = q(1-t)$, $W := (1-U)/q(1-U)$ and $T := U/q(U)$ have the same distribution. Moreover $Q_T(U_k/2) = (1-U_k/2)/q(U_k/2)$. Then Theorem 2.2.1 implies (2.6.1), and the proof is completed. □

2.7 Suprema of randomized empirical processes

Let $X_1, \ldots, X_n$ be a sequence of independent random variables with values in some Polish space $\mathcal{X}$ and $Y_1, \ldots, Y_n$ be a sequence of independent real-valued symmetric random variables such that the two sequences are independent. Let $\mathscr{F}$ be a countable class of measurable real-valued functions and define the function F by

$$Z := F(\widetilde{X}_1, \ldots, \widetilde{X}_n) := \sup_{f \in \mathscr{F}} \sum_{k=1}^n Y_k f(X_k), \tag{2.7.1}$$

where $\widetilde{X}_k := (Y_k f(X_k))_{f \in \mathscr{F}}$. Assume that there exist nonnegative functions G and H such that for any function f in $\mathscr{F}$, $-G \leq f \leq H$. It thus follows that $-T_k \leq Z - Z^k \leq W_k$, where

$$W_k := Y_{k+}H(X_k) + Y_{k-}G(X_k), \quad T_k := Y_{k-}H(X_k) + Y_{k+}G(X_k).$$

Throughout, we assume that $\mathbb{E}[G^2(X_k)] < \infty$ and $\mathbb{E}[H^2(X_k)] < \infty$ for any $k = 1, \ldots, n$. Since Y is symmetric, W_k and T_k have the same distribution. Then Theorem 2.2.1 yields

$$Z - \mathbb{E}[Z] \preccurlyeq_{\mathcal{H}_+^2} \sum_{k=1}^n \varepsilon_k Q_{W_k}(U_k/2). \tag{2.7.2}$$

Throughout this section we will use the following notation:

$$s^2 := \sum_{k=1}^n s_k^2 := \sum_{k=1}^n \mathbb{E}[Y_k^2], \tag{2.7.3}$$

$$\sigma^2 := \sum_{k=1}^n \sigma_k^2 := \sum_{k=1}^n \mathbb{E}[Q_{W_k}^2(U_k/2)]. \tag{2.7.4}$$

In the rest of this section, we present how (2.7.2) may be used to derive concentration inequalities through several examples. However, in some cases, this bound can prove difficult to manipulate. Now, we show that, due to the symmetry of the Y_k, we can derive a more tractable comparison moment inequality, which is, however, less efficient. Precisely, let us define $\xi_k := Y_{k+}(G(X_k) + H(X_k))$. Let us first observe the following lemma:

Lemma 2.7.1. *For any $t \in \mathbb{R}$,*

$$\mathbb{E}[(W_k - t)_+] \leq \mathbb{E}[(\xi_k - t)_+].$$

Proof. Since W_k and ξ_k are nonnegative, the inequality is trivially true for $t < 0$. Let now $t \geq 0$. By the definition of W_k and since Y_k is symmetric and independent of X_k,

$$\mathbb{E}[(W_k - t)_+] = \mathbb{E}[(Y_{k+}H(X_k) - t)_+] + \mathbb{E}[(Y_{k+}G(X_k) - t)_+]$$

Now, the superadditivity of the function $x \mapsto (x-t)_+$ on $[0, \infty[$ yields

$$\mathbb{E}[(W_k - t)_+] \leq \mathbb{E}[(Y_{k+}(H(X_k) + G(X_k)) - t)_+] = \mathbb{E}[(\xi_k - t)_+],$$

which ends the proof. □

Moreover, since Y_k is symmetric and independent of X_k, $\varepsilon_k Q_{\xi_k}(U_k/2)$ and $Y_k(G(X_k) + H(X_k))$ have the same distribution. Consequently, we derive from Lemma 2.7.1 and Theorem 2.2.1 that

$$Z - \mathbb{E}[Z] \underset{\mathcal{H}_+^2}{\preccurlyeq} \sum_{k=1}^n Y_k \left(G(X_k) + H(X_k)\right). \tag{2.7.5}$$

Example 2.7.2. *Let $F_0 : \mathbb{R} \to [0,1]$ be a nondecreasing function. Let $X_1, \ldots, X_n$ be independent real-valued random variables and let $q :]0,1[\to \mathbb{R}$ be a weight function, continuous, such that*

$$q(t) > 0, \quad \int_0^{1/2} \frac{dt}{q^2(t)} < \infty,$$

$$t \mapsto \frac{q(t)}{t} \text{ is nonincreasing,} \quad \text{and} \quad t \mapsto \frac{q(t)}{1-t} \text{ is nondecreasing.}$$

Define now

$$Z := \sup_{\substack{t\in\mathbb{R}\\ 0<F_0(t)<1}} \sum_{k=1}^{n} Y_k \frac{\mathbb{1}_{X_k\le t} - F_0(t)}{q(F_0(t))}.$$

In this case

$$H(x) = \frac{1-F_0(x)}{q(F_0(x))}, \ \textit{and}\ G(x) = \frac{F_0(x-)}{q(F_0(x-))},$$

whence

$$Z - \mathbb{E}[Z] \underset{\mathcal{H}_+^2}{\preccurlyeq} \sum_{k=1}^{n} Y_k \left(\frac{1-F_0(X_k)}{q(F_0(X_k))} + \frac{F_0(X_k-)}{q(F_0(X_k-))}\right) \tag{2.7.6}$$

$$\underset{\mathcal{H}_+^2}{\preccurlyeq} \sum_{k=1}^{n} \frac{Y_k}{q(F_0(X_k))}, \tag{2.7.7}$$

where the last inequality come from Theorem 2.2.1 applied to the right-hand side of (2.7.6) *since* $Y_k \le Y_{k+}$ *and* $0 \le G(X_k) + H(X_k) \le 1/q(F_0(X_k))$ *by the assumption on* q. *Let us now give a relevant example. We assume that* $X_1, \ldots, X_n$ *are* n *independent copies of a random variable* U *distributed uniformly on* $[0,1]$. *Let* $F_0 = F_U$ *be the distribution function of* U *and let* $q(t) = \sqrt{\max(t,\delta)}$ *for some* $0<\delta<1$. *Then* (2.7.7) *gives*

$$\mathbb{E}[(Z-\mathbb{E}[Z])_+^2] \le \frac{1}{2}\sum_{k=1}^{n} \mathbb{E}\Big[\frac{Y_k^2}{\max(X_k,\delta)}\Big] = \frac{1}{2}\, s^2 \log\Big(\frac{e}{\delta}\Big).$$

2.7.1 Case $G = 0$

Here $W_k = Y_{k+}H(X_k)$. Since Y_k is symmetric and independent of X_k, $\varepsilon_k Q_{W_k}(U_k/2)$ and $Y_k H(X_k)$ have the same distribution. Then (2.7.4) becomes

$$\sigma^2 = \sum_{k=1}^{n} \mathbb{E}[Y_k^2]\,\mathbb{E}[H^2(X_k)], \tag{2.7.8}$$

and (2.7.2) becomes

$$Z - \mathbb{E}[Z] \underset{\mathcal{H}_+^2}{\preccurlyeq} \sum_{k=1}^{n} Y_k H(X_k). \tag{2.7.9}$$

2.7.1.1 Chebyshev type inequality

Proposition 2.7.3. *Let* Z *be defined by* (2.7.1) *and let* σ^2 *defined by* (2.7.8). *For any* $x>0$,

$$\mathbb{P}(Z - \mathbb{E}[Z] \ge \sigma x) \le \min\Big(\frac{1}{1+x^2}\,,\ \frac{1}{2x^2}\Big). \tag{2.7.10}$$

Proof. Let $\xi := \sum_{k=1}^n Y_k H(X_k)$. Let $x > 0$. From (2.7.9) and the result of Pinelis (2.3.1), we derive

$$\mathbb{P}(Z - \mathbb{E}[Z] \geq x) \leq P_2(\xi\,; x) = \inf_{t<x} \frac{\mathbb{E}[(\xi - t)_+^2]}{(x-t)^2}. \tag{2.7.11}$$

Taking now $t = 0$ in the right-hand side of (2.7.11) gives the bound

$$P_2(\xi\,; x) \leq \sigma^2/(2x^2). \tag{2.7.12}$$

Moreover, since $z_+^2 \leq z^2$ for all $z \in \mathbb{R}$, we get the so-called Cantelli's inequality

$$P_2(\xi\,; x) \leq \inf_{t<x} \frac{\mathbb{E}[(\xi - t)^2]}{(x-t)^2} = \inf_{t<x} \frac{\sigma^2 + t^2}{(x-t)^2} = \frac{\sigma^2}{\sigma^2 + x^2}. \tag{2.7.13}$$

Finally, combining (2.7.11), (2.7.12) and (2.7.13) ends the proof. □

2.7.1.2 $H = 1$ and Gaussian case

Here, we assume that $Y_1, \ldots, Y_n$ is a sequence of independent centered Gaussian random variables. Since $H = 1$, σ defined by (2.7.8) is the standard deviation of $\sum_{k=1}^n Y_k$. We obtain the following inequality:

Proposition 2.7.4. *Let $Y_1, \ldots, Y_n$ be a sequence of independent centered Gaussian random variables. Let g be a standard Gaussian random variable. Let σ denote the standard deviation of $\sum_{k=1}^n Y_k$. Let Z be defined by* (2.7.1). *Then for any $x > 0$,*

$$\mathbb{P}(Z - \mathbb{E}[Z] \geq \sigma x) \leq \min\left\{\frac{1}{1+x^2}, \frac{1}{2x^2}, \frac{e^2}{2}\,\mathbb{P}(g \geq x)\right\} := h(x). \tag{2.7.14}$$

Remark 2.7.5. *Note that $h(x) = 1/(1+x^2)$ for any $0 < x \leq 1$, $h(x) = 1/2x^2$ for any $1 < x \leq x_0$ and $h(x) = (e^2/2)\,\mathbb{P}(g \geq x)$ for any $x > x_0$, where x_0 is the unique root of the equation $(e^2/2)\,\mathbb{P}(g \geq x) = 1/2x^2$. A numerical calculation gives $x_0 \approx 1.6443$. Furthermore the function h is always better than the usual exponential bound (i.e. $h(x) \leq \exp(-x^2/2)$).*

Proof of Proposition 2.7.4. Let $\xi := \sum_{k=1}^n Y_k$. Then σg and ξ have the same distribution. Starting as in the proof of Proposition 2.7.3, we have for any $x > 0$,

$$\mathbb{P}(Z - \mathbb{E}[Z] \geq x) \leq P_2(\xi; x). \tag{2.7.15}$$

Then Inequality (2.3.2) provides

$$P_2(\xi; x) \leq \frac{e^2}{2}\mathbb{P}(\xi \geq x) = \frac{e^2}{2}\mathbb{P}(\sigma g \geq x), \tag{2.7.16}$$

which gives the third bound in (2.7.14). The two other bounds are given by Proposition 2.7.3 since this case is a particular case. □

2.7.1.3 $0 \leq H \leq 1$ and Gaussian case

Here, we assume that $Y_1, \ldots, Y_n$ is a sequence of independent centered Gaussian random variables and $X_1, \ldots, X_n$ are identically distributed according to some distribution P.

Proposition 2.7.6. *Let $g_1, \ldots g_n$ be a sequence of independent standard Gaussian random variables and let $a_1, \ldots, a_n$ be a sequence of positive deterministic reals. Set for any $k = 1, \ldots, n$, $Y_k = a_k g_k$. Let $X_1, \ldots, X_n$ be a sequence of independent random variables valued in some Polish space $\mathcal{X}$ with common distribution P and independent of the sequence $g_1, \ldots, g_n$. Let Z be defined by (2.7.1). Let $a := (a_1, \ldots, a_n)$, $v := \mathbb{E}[H^2(X_1)]$ and let γ be the function defined on $]0, \infty[$ by*

$$\gamma(x) := x\sqrt{2}/\sqrt{\log(1 + v^{-1}(e^x - 1))}.$$

Then for any $x > 0$,

$$\mathbb{P}\left(Z - \mathbb{E}[Z] > \frac{\|a\|_2^2}{\|a\|_\infty}\gamma\left(\frac{\|a\|_\infty^2 x}{\|a\|_2^2}\right)\right) \leq \exp(-x). \tag{2.7.17}$$

Remark 2.7.7. *As x goes to zero, the function γ has the asymptotic expansion*

$$\gamma(x) = \sqrt{2vx}\,(1 + O(x)),$$

and as x goes to infinity, $\gamma(x) \sim \sqrt{2x}$.

Proof of Proposition 2.7.6. Starting as in the proof of Proposition 2.5.5 and conditioning by X_k, one has

$$\begin{aligned}\log \mathbb{E}[\exp(t(Z - \mathbb{E}[Z]))] &\leq \sum_{k=1}^n \log \mathbb{E}\Big[\mathbb{E}[\exp(ta_k g_k H(X_k)) \mid X_k]\Big] \\ &\leq \sum_{k=1}^n \log \mathbb{E}\Big[\exp\Big(\frac{a_k^2 t^2}{2} H^2(X_1)\Big)\Big].\end{aligned}$$

Define next the function ℓ_v by

$$\ell_v(t) := \log(1 + v(e^{t^2/2} - 1)) \quad \text{for any } t > 0. \tag{2.7.18}$$

Now by the convexity of the function $\lambda \mapsto e^{\alpha\lambda}$ for any $\alpha \geq 0$, the chordal slope lemma yields

$$\log\Big(\mathbb{E}\Big[\exp\Big(\frac{a_k^2 t^2}{2} H^2(X_1)\Big)\Big]\Big) \leq \ell_v(a_k t). \tag{2.7.19}$$

In order to bound up the right-hand side term, we will use the property below concerning ℓ_v.

Lemma 2.7.8. *Let h_v be the function defined by $h_v(t) := \ell_v'(t)/t$ for any $t > 0$. Then h_v is nondecreasing.*

Proof of Lemma 2.7.8. A straightforward calculation leads to

$$\ell_v''(t) = \ell_v'(t)\left(\frac{1}{t} + \frac{t(1-v)}{1+v(e^{t^2/2}-1)}\right).$$

Then

$$h_v'(t) = \frac{t\ell_v''(t) - \ell_v'(t)}{t^2} = \ell_v'(t)\left(\frac{1-v}{1+v(e^{t^2/2}-1)}\right).$$

Since $v \leq 1$, we get $h_v'(t) \geq 0$ and the lemma follows. $\square$

Next, from (2.7.19) and Lemma 2.7.8, proceeding exactly as in Bercu, Delyon and Rio [8] (see their Inequality (2.97)), one has

$$\log \mathbb{E}\left[\exp(t(Z - \mathbb{E}[Z]))\right] \leq \frac{\|a\|_2^2}{\|a\|_\infty^2}\, \ell_v(\|a\|_\infty t).$$

From the inversion formula for ℓ_v^* given in [8] (see Exercise 1, p. 57):

$$\ell_v^{*-1}(x) = \inf\{t^{-1}(\ell_v(t) + x) : t > 0\}, \tag{2.7.20}$$

we can see that for any $x > 0$,

$$\inf_{t>0} \frac{1}{t}\left(\frac{\|a\|_2^2}{\|a\|_\infty^2}\ell_v(\|a\|_\infty t) + x\right) = \frac{\|a\|_2^2}{\|a\|_\infty}\, \ell_v^{*-1}\left(\frac{\|a\|_\infty^2 x}{\|a\|_2^2}\right). \tag{2.7.21}$$

Then (see Lemma 2.7 of [8]),

$$\mathbb{P}\left(Z - \mathbb{E}[Z] > \frac{\|a\|_2^2}{\|a\|_\infty}\ell_v^{*-1}\left(\frac{\|a\|_\infty^2 x}{\|a\|_2^2}\right)\right) \leq \exp(-x). \tag{2.7.22}$$

However, it seems difficult to calculate the inverse function of ℓ_v^*. Then to obtain a "ready-to-use" inequality, we will bound up $\ell_v^{*-1}(x)$.
Let $t_x := \sqrt{2\log(1 + v^{-1}(e^x - 1))}$. Remark that t_x is optimal in the Gaussian case ($v = 1$). Hence, putting t_x in (2.7.20), we get $\ell_v^{*-1}(x) \leq \gamma(x)$ and the proposition follows. $\square$

2.7.1.4 Unbounded function H

Here we assume that Y_k and $H(X_k)$ are $\mathbb{L}^r$-integrable random variables with $2 \le r \le 4$.

Proposition 2.7.9. *Let $2 \le r \le 4$. Then*

$$\|(Z - \mathbb{E}[Z])_+\|_r^r \le \frac{1}{2} \sum_{k=1}^{n} \|Y_k\|_r^r \, \|H(X_k)\|_r^r + \frac{1}{2}\sigma^r \|g\|_r^r \, \mathbb{1}_{r>2},$$

where σ^2 is defined by (2.7.8) *and g is a standard Gaussian random variable.*

Proof. Applying (2.7.9) with $\varphi(x) = x_+^r$, we get

$$\|Z - \mathbb{E}[Z]\|_r^r \le \mathbb{E}\Big[\Big(\sum_{k=1}^{n} Y_k H(X_k)\Big)_+^r\Big]. \tag{2.7.23}$$

Moreover, we have already noticed in the proof of Proposition 2.5.1 that the symmetry of $Y_k H(X_k)$ allows us to write

$$\mathbb{E}\Big[\Big(\sum_{k=1}^{n} Y_k H(X_k)\Big)_+^r\Big] = \frac{1}{2}\mathbb{E}\Big[\Big|\sum_{k=1}^{n} Y_k H(X_k)\Big|^r\Big]. \tag{2.7.24}$$

Now, Corollary 6.2 of Figiel et al. [13] (see also their Theorems 6.1 and 7.1) and the independence between X_k and Y_k lead to

$$\mathbb{E}\Big[\Big|\sum_{k=1}^{n} Y_k H(X_k)\Big|^r\Big] \le \sum_{k=1}^{n} \mathbb{E}[|Y_k H(X_k)|^r] + \|g\|_r^r \Big(\sum_{k=1}^{n} \mathbb{E}[Y_k^2 H^2(X_k)]\Big)^{r/2} \mathbb{1}_{r>2}, \tag{2.7.25}$$

where g is a standard Gaussian random variable. Thus, combining (2.7.23)–(2.7.25) concludes the proof. □

2.7.2 Case $G \neq 0$

In the sequel, we assume that the underlying probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is rich enough to contain a random variable δ with uniform distribution over $[0, 1]$, independent of all other considered random variables. First we present a duality formula for the r-th moment of $\varepsilon_k Q_{W_k}(U_k/2)$ for any $r \ge 2$. It will allow us to derive a simpler bound of these moments which we will use thereafter to obtain concentration inequalities.

2.7.2.1 Duality formula

Lemma 2.7.10. *Let $r \geq 2$. One has*

$$\begin{aligned}&\mathbb{E}\Big[|\varepsilon_k Q_{W_k}(U_k/2)|^r\Big] \qquad (2.7.26)\\ &\quad = \sup\Big\{\mathbb{E}\Big[|Y_k|^r(H^r(X_k)\mathbb{1}_A + G^r(X_k)\mathbb{1}_B)\Big] : A, B \in \mathcal{F},\, \mathbb{P}(A) + \mathbb{P}(B) = 1\Big\}.\end{aligned}$$

Remark 2.7.11. *The duality formula gives us directly a more tractable bound*

$$\mathbb{E}\Big[|\varepsilon_k Q_{W_k}(U_k/2)|^r\Big] \leq \mathbb{E}\Big[|Y_k|^r\Big]\,\mathbb{E}\Big[H^r(X_k) + G^r(X_k)\Big]. \qquad (2.7.27)$$

Proof of Lemma 2.7.10. Let us recall the general following fact. Let $\alpha \in]0,1[$ and let θ_α be a Bernoulli random variable with parameter α. Let X be an integrable random variable. Then

$$\int_0^\alpha Q^r_{|X|}(u)du = \int_0^1 Q^r_{|X|}(u)\, Q_{\theta_\alpha}(u)du = \sup_\theta \mathbb{E}[|X|^r\theta],$$

where the supremum is taken over the set of all Bernoulli random variables with parameter α. Consequently,

$$\mathbb{E}\Big[|\varepsilon_k Q_{W_k}(U_k/2)|^r\Big] = 2\sup\Big\{\mathbb{E}[W_k^r\,\mathbb{1}_C] : C \in \mathcal{F},\, \mathbb{P}(C) = 1/2\Big\}. \qquad (2.7.28)$$

Then, recalling the definition of W_k and since $Y_{k+} = |Y_k|\mathbb{1}_{Y\geq 0}$ and $Y_{k-} = |Y_k|\mathbb{1}_{Y<0}$, one has

$$\begin{aligned}&\mathbb{E}\Big[|\varepsilon_k Q_{W_k}(U_k/2)|^r\Big] = \qquad (2.7.29)\\ &2\sup\Big\{\mathbb{E}\Big[|Y_k|^r(H^r(X_k)\mathbb{1}_{C\cap\{Y\geq 0\}} + G^r(X_k)\mathbb{1}_{C\cap\{Y<0\}})\Big] : C \in \mathcal{F},\, \mathbb{P}(C) = 1/2\Big\}.\end{aligned}$$

Since $C\cap\{Y \geq 0\}$ and $C \cap \{Y < 0\}$ are disjoint, the supremum in the right hand-side of (2.7.29) is equal to

$$\begin{aligned}&\sup\Big\{\mathbb{E}\Big[|Y_k|^r(H^r(X_k)\mathbb{1}_A + G^r(X_k)\mathbb{1}_B)\Big] :\\ &\qquad A, B \in \mathcal{F},\, A \subset \{Y \geq 0\}, B \subset \{Y < 0\}, \mathbb{P}(A) + \mathbb{P}(B) = 1/2\Big\}. \qquad (2.7.30)\end{aligned}$$

Now, recalling that W_k and T_k have the same distribution and repeating the above proof for $\mathbb{E}[|\varepsilon_k Q_{T_k}(U_k/2)|^r]$, we deduce that the supremum in (2.7.30) is equal to

$$\begin{aligned}&\sup\Big\{\mathbb{E}\Big[|Y_k|^r(H^r(X_k)\mathbb{1}_{\tilde{A}} + G^r(X_k)\mathbb{1}_{\tilde{B}})\Big] :\\ &\qquad \tilde{A}, \tilde{B} \in \mathcal{F},\, \tilde{A} \subset \{Y < 0\}, \tilde{B} \subset \{Y \geq 0\}, \mathbb{P}(\tilde{A}) + \mathbb{P}(\tilde{B}) = 1/2\Big\}. \qquad (2.7.31)\end{aligned}$$

Finally, combining (2.7.28) – (2.7.30) ends the proof of Lemma 2.7.9. □

2.7.2.2 Chebyshev type inequality

Proposition 2.7.12. *Define*

$$V := \sum_{k=1}^{n} \mathbb{E}[\, Y_k^2 (G \vee H)^2(X_k)\,], \;\; \textit{and} \;\; V_1 := \sum_{k=1}^{n} \mathbb{E}[\, Y_k^2\, (G^2(X_k) + H^2(X_k))\,].$$

Let Z *be defined by* (2.7.1)*. Then for any* $x > 0$*,*

$$\begin{aligned} \mathbb{P}(Z - \mathbb{E}[Z] \geq x) &\leq \min\left(\frac{V}{V + x^2}\,, \,\frac{\sigma^2}{2x^2}\right) && (2.7.32)\\ &\leq \min\left(\frac{V}{V + x^2}\,, \,\frac{V_1}{2x^2}\right), && (2.7.33)\end{aligned}$$

where σ^2 *is defined by* (2.7.4)*.*

Remark 2.7.13. *Note that* $V_1/2x^2 \leq V/(V + x^2)$ *for all* x *such that*

$$x^2 \geq \sum_{k=1}^{n} \mathbb{E}[\, Y_k^2\, (G \vee H)(X_k)\,]\, \frac{\sum_{k=1}^{n} \mathbb{E}[\, Y_k^2\, (G^2(X_k) + H^2(X_k))\,]}{\sum_{k=1}^{n} \mathbb{E}[\, Y_k^2\, |G^2(X_k) - H^2(X_k)|\,]}.$$

Example 2.7.14. *Let* $\mathscr{S}$ *be a countable class of sets such that for any* S *in* $\mathscr{S}$*,* $\mathbb{P}(S) \leq p$*. We consider the class of functions* $\mathscr{F} = \{\mathbb{1}_S - \mathbb{P}(S) : S \in \mathscr{S}\}$*. Here* $H = 1$ *and* $G = p$ *which imply* $V = s^2$ *and* $V_1 = (1 + p^2)s^2$ *where* s^2 *is defined by* (2.7.3)*. Then Proposition 2.7.12 yields for any* $x > 0$*,*

$$\mathbb{P}(Z - \mathbb{E}[Z] \geq sx) \leq \begin{cases} \dfrac{1}{1 + x^2} & \textit{if } x < x_0 \\ \dfrac{1 + p^2}{2x^2} & \textit{if } x \geq x_0 \end{cases},$$

where $x_0^2 = (1 + p^2)/(1 - p^2)$*.*

Proof of Proposition 2.7.12. Since $\sigma^2 \leq V_1$ by Remark 2.7.11, (2.7.33) follows from (2.7.32). Let us prove now (2.7.32). The Cantelli inequality (see, for instance, Exercise 2.3 in Boucheron et al. [10]), states that for any $x > 0$,

$$\mathbb{P}(Z - \mathbb{E}[Z] \geq x) \leq \frac{\mathrm{Var}(Z)}{\mathrm{Var}(Z) + x^2}.$$

Moreover, Theorem 4 in Pinelis and Sakhanenko [26] (or Theorem 11.1 [10]) implies that $\mathrm{Var}(Z) \leq V$. Since $t \mapsto t/(t + x^2)$ is increasing on $[0, \infty[$, we get the bound

$$\mathbb{P}(Z - \mathbb{E}[Z] \geq x) \leq \frac{V}{V + x^2}.$$

Furthermore, set $\xi := \sum_{k=1}^n \varepsilon_k Q_{W_k}(U_k/2)$. Starting from (2.7.2) and proceeding as in the proof of Proposition 2.7.3, we get

$$\mathbb{P}(Z - \mathbb{E}[Z] \geq x) \leq P_2(\xi; x) \leq \frac{\sigma^2}{2x^2},$$

which ends the proof. □

2.7.2.3 Moment inequality

Here we assume that the random variables Y_k, $G(X_k)$ and $H(X_k)$ are $\mathbb{L}^r$-integrable with $2 \leq r \leq 4$. The following moment inequality is similar to Proposition 2.7.9.

Proposition 2.7.15. *Let $2 \leq r \leq 4$. Let Z be defined by* (2.7.1)*. Then*

$$\|(Z - \mathbb{E}[Z])_+\|_r^r \leq \frac{1}{2}\sum_{k=1}^n \mathbb{E}\Big[|\varepsilon_k Q_{W_k}(U_k/2)|^r\Big] + \frac{1}{2}\,\sigma^r \|g\|_r^r \,\mathbb{1}_{r>2}, \qquad (2.7.34)$$

where σ is defined by (2.7.4) *and g is a standard Gaussian random variable. Consequently, using Remark 2.7.11, we get*

$$\|(Z-\mathbb{E}[Z])_+\|_r^r \leq \frac{1}{2}\sum_{k=1}^n \mathbb{E}\Big[|Y_k|^r (H^r(X_k)+G^r(X_k))\Big] + \frac{1}{2}\,\sigma^r \|g\|_r^r \,\mathbb{1}_{r>2}. \quad (2.7.35)$$

Example 2.7.14 (continued). We derive immediately by Markov's inequality, for any $2 \leq r \leq 4$,

$$\mathbb{P}(Z - \mathbb{E}[Z] \geq sx) \leq \frac{1}{2}\Bigg(\|g\|_r^r \bigg(\frac{\sqrt{1+p^2}}{x}\bigg)^r \mathbb{1}_{r>2} + \frac{1+p^r}{s^r x^r}\sum_{k=1}^n \mathbb{E}[Y_k^r]\Bigg).$$

2.7.2.4 Exponential inequality

Proposition 2.7.16. *Let Z be defined by* (2.7.1) *and let L denote the logarithm of the Laplace transform of $Z - \mathbb{E}[Z]$. Then for any $t > 0$,*

$$L(t) \leq \sum_{k=1}^n \log\Big(\mathbb{E}\left[\cosh\left(t\, Y_k H(X_k)\right) + \cosh\left(t\, Y_k G(X_k)\right) - 1\right]\Big).$$

Example 2.7.14 (continued). Here, we make the additional assumptions that the Y_k are standard Gaussian random variables. In this situation, the inequality above implies that

$$L(t) \leq n \log\Big(e^{t^2/2} + e^{p^2 t^2/2} - 1\Big) \leq n\, \ell_{1+p^2}(t),$$

where $\ell_v(t) := \log(1 + v(e^{t^2/2} - 1))$. Hence, for any $x > 0$,

$$\mathbb{P}\left(Z - \mathbb{E}[Z] \geq \frac{n\,\ell_{1+p^2}(t) + x}{t}\right) \leq \exp(-x). \tag{2.7.36}$$

Let $t_x := \sqrt{2\log(1 + (1+p^2)^{-1}(e^{x/n} - 1))}$. Then putting t_x in (2.7.36), we obtain for any $x > 0$,

$$\mathbb{P}\left(Z - \mathbb{E}[Z] \geq \frac{x\sqrt{2}}{\sqrt{\log(1 + (1+p^2)^{-1}(e^{x/n} - 1))}}\right) \leq \exp(-x). \tag{2.7.37}$$

Proof of Proposition 2.7.16. Applying (2.7.2) with $\varphi(x) = e^{tx}$, $t > 0$, we get

$$L(t) \leq \sum_{k=1}^{n} \log\left(\mathbb{E}\left[\exp(t\varepsilon_k Q_{W_k}(U_k/2))\right]\right). \tag{2.7.38}$$

Now, using Remark 2.7.11,

$$\begin{aligned}\mathbb{E}\left[\exp(t\varepsilon_k Q_{W_k}(U_k/2))\right] &= \sum_{j=0}^{\infty} \frac{t^{2j}}{(2j)!}\mathbb{E}\left[(Q_{W_k}(U_k/2))^{2j}\right] \\ &\leq \mathbb{E}\left[\cosh\left(tY_k H(X_k)\right) + \cosh\left(tY_k G(X_k)\right) - 1\right].\end{aligned}$$

Putting then this inequality in (2.7.38) ends the proof. $\square$

2.8 Chaos of order two

Let $\mathcal{X}$ be a Polish space and $\mathscr{F}$ be a countable class of measurable functions from $\mathcal{X}$ into $\mathbb{R}$ and let Γ be a subset of $\mathscr{F} \times \mathscr{F}$. Let $A = (a_{i,j})_{1\leq i,j\leq n}$ be a symmetric real matrix with zero diagonal entries (i.e. $a_{i,i} = 0$ for all i) and let $\|.\|$ denote the Hilbert-Schmidt norm which is $\|A\|_{\mathrm{HS}} = \sqrt{\mathrm{Tr}(A^{\mathrm{T}}A)}$. Let X be a random variable with values in $\mathcal{X}$ such that for any function f of $\mathscr{F}$, $f(X)$ is a centered random variable. Let $X_1, \ldots, X_n$ be n independent copies of X. Define now the function F by

$$Z := F(\widetilde{X}_1, \ldots, \widetilde{X}_n) := \sup_{(f,g)\in\Gamma} \sum_{1\leq i<j\leq n} a_{ij} f(X_i)g(X_j), \tag{2.8.1}$$

where $\widetilde{X}_k := (f(X_k)g(X_k))_{(f,g)\in\Gamma}$. We say that $\mathscr{F}$ is a Vapnik–Červonenkis (VC for short) subgraph class if the collection of all subgraphs of the functions in $\mathscr{F}$ (i.e. the collection of sets $\{(x,s) \in \mathcal{X} \times \mathbb{R} : s < f(x)\}$ for $f \in \mathscr{F}$) forms a VC-class of sets in $\mathcal{X} \times \mathbb{R}$ (see, for instance, van der Vaart and Wellner [30]).

Proposition 2.8.1. *Let Z be defined by* (2.8.1) *and $p > 2$. Assume that $\mathscr{F}$ is a VC-subgraph class of functions with square integrable envelope function Φ. Then there exists a constant $K(\mathscr{F})$ depending only on $\mathscr{F}$ such that*

$$\begin{aligned} &\|(Z - \mathbb{E}[Z])_+\|_p \\ &\leq (p-1)\frac{\|A\|_{HS}}{\sqrt{2}}\,\|Q_{\Phi(X)}(U/2)\|_p^2\left(\sqrt{1+\frac{K(\mathscr{F})}{(p-1)}\frac{\|\Phi(X)\|_2^2}{\|Q_{\Phi(X)}(U/2)\|_p^2}}\right), \end{aligned} \tag{2.8.2}$$

where U is a random variable distributed uniformly on $[0,1]$.

Remark 2.8.2. *See that*

$$\|\Phi(X)\|_p \leq \|Q_{\Phi(X)}(U/2)\|_p \leq 2^{1/p}\|\Phi(X)\|_p. \tag{2.8.3}$$

Suppose now that $\Phi(X)$ is in $\mathbb{L}^p$ for all $p > 2$ and $\|\Phi(X)\|_p$ tends to infinity as p tends to infinity. Then as p tends to infinity, we obtain the following behavior of the right-hand side of (2.8.2)

$$\begin{aligned} &(p-1)\frac{\|A\|_{HS}}{\sqrt{2}}\,\|Q_{\Phi(X)}(U/2)\|_p^2\left(\sqrt{1+\frac{K(\mathscr{F})}{(p-1)}\frac{\|\Phi(X)\|_2^2}{\|Q_{\Phi(X)}(U/2)\|_p^2}}\right) \\ &\qquad = (p-1)\frac{\|A\|_{HS}}{\sqrt{2}}\,\|\Phi(X)\|_p^2\left(1+O\Big(\frac{1}{p}\Big)\right). \end{aligned} \tag{2.8.4}$$

Example 2.8.3. *Here we assume that $\Phi \leq 1$ and we show how* (2.8.2) *can be used to obtain an exponential bound for the tail probability.*

Let $p > 2$, $x \geq 0$, and define the function f_x on $]2,\infty[$ by

$$f_x(q) := \left((q-1)Cx^{-1}\,\|Q_{\Phi(X)}(U/2)\|_q^2\left(\sqrt{1+\frac{K(\mathscr{F})}{(q-1)}\frac{\|\Phi(X)\|_2^2}{\|Q_{\Phi(X)}(U/2)\|_q^2}}\right)\right)^q,$$

where $C := \|A\|_{HS}/\sqrt{2}$. Using (2.8.3) *and $\Phi \leq 1$, one has*

$$\begin{aligned} f_x(q) &\leq \left((q-1)Cx^{-1}\,2^{2/q}\,\|\Phi\|_q^2\left(\sqrt{1+\frac{K(\mathscr{F})}{(p-1)}\frac{\|\Phi\|_2^2}{\|\Phi\|_q^2}}\right)\right)^q \\ &\leq 4\left(Cx^{-1}\,q\sqrt{1+K(\mathscr{F})}\,\right)^q := 4\exp(h_x(q)). \end{aligned}$$

Clearly, $\inf_{q>0} h_x(q) = h_x\Big(x/(e\,C\sqrt{1+K(\mathscr{F})}\,)\Big)$. Now, by Markov's inequality, for any $x \geq 0$ such that $2\,e\,Cx^{-1}\sqrt{1+K(\mathscr{F})} \leq 1$,

$$\begin{aligned} \mathbb{P}(Z - \mathbb{E}[Z] \geq x) &\leq \inf_{p>2}\frac{\mathbb{E}[(Z-\mathbb{E}[Z])_+^p]}{x^p} \\ &\leq 4\exp\left(-\frac{x}{e\,C\sqrt{1+K(\mathscr{F})}}\right). \end{aligned}$$

Proof of Proposition 2.8.1. For any m and l belonging to $\{1,\dots,n\}$, we set $S_k(l,m) := \sup_{f\in\mathscr{F}} \left|\sum_{i=l}^m a_{ik} f(X_i)\right|$. Noting that

$$|Z - Z^{(k)}| \le \Phi(X_k)\left(S_k(1,k-1) + S_k(k+1,n)\right),$$

it follows from Theorem 2.2.3 that

$$\begin{aligned}&\|(Z-\mathbb{E}[Z])_+\|_p^2\\ &\quad\le (p-1)\|Q_{\Phi(X)}(U/2)\|_p^2 \sum_{k=1}^n \Big(\|S_k(1,k-1)+\mathbb{E}[S_k(k+1,n)]\|_p^2\Big).\end{aligned} \tag{2.8.5}$$

Define the function $\widetilde{F}$ such that $\widetilde{Z} := \widetilde{F}(\widetilde{X}_1,\dots,\widetilde{X}_{k-1}) := S(1,k-1)$. Define also for each $l = 1,\dots,k-1$,

$$\widetilde{Z}^{(l)} := \widetilde{F}(\widetilde{X}_1,\dots,\widetilde{X}_{l-1},0,\widetilde{X}_{l+1},\dots,\widetilde{X}_{k-1}).$$

Hence, it follows that

$$|\widetilde{Z} - \widetilde{Z}^{(l)}| \le |a_{lk}|\Phi(X_l).$$

Since $S_k(1,k-1)+\mathbb{E}[S_k(k+1,n)]$ is a nonnegative random variable, we can replace its p-norm in (2.8.5) by $\|(S_k(1,k-1)+\mathbb{E}[S_k(k+1,n)])_+\|_p$. Then an application of Theorem 2.2.1 leads to

$$\begin{aligned}&\|(S_k(1,k-1)+\mathbb{E}[S_k(k+1,n)])_+\|_p^2\\ &\quad\le \left\|\Big(\sum_{i=1}^{k-1} a_{ik}\,\varepsilon_i Q_{\Phi(X)}(U/2) + \mathbb{E}[S_k(1,k-1)+S_k(k+1,n)]\Big)_+\right\|_p^2.\end{aligned} \tag{2.8.6}$$

Now by Corollary 2.4.9, Inequality (2.8.6) becomes

$$\begin{aligned}&\|S_k(1,k-1)+\mathbb{E}[S_k(k+1,n)]\|_p^2\\ &\qquad\le (p-1)\Big(\sum_{i=1}^{k-1} a_{ik}^2\Big)\|Q_{\Phi(X)}(U/2)\|_p^2\\ &\qquad\quad + \Big(\mathbb{E}[S_k(1,k-1)+S_k(k+1,n)]\Big)^2.\end{aligned} \tag{2.8.7}$$

Combining (2.8.5) and (2.8.7), we get

$$\begin{aligned}\|(Z-\mathbb{E}[Z])_+\|_p^2 &\le (p-1)^2\Big(\sum_{1\le i<k\le n} a_{ik}^2\Big)\|Q_{\Phi(X)}(U/2)\|_p^4\\ &\quad + (p-1)\|Q_{\Phi(X)}(U/2)\|_p^2 \sum_{k=1}^n \Big(\mathbb{E}[S_k(1,k-1)+S_k(k+1,n)]\Big)^2.\end{aligned} \tag{2.8.8}$$

Let us now bound up $\mathbb{E}[S_k(1,k-1)]$. Define the probability measure

$$\mathbb{P}_{k-1} = \sum_{i=1}^{k-1} a_{i,k}^2 \, \delta_{X_i}.$$

Exactly as in the proof of Theorem 2.5.2 in van der Vaart and Wellner [30], it can be shown that for some universal constant K,

$$\begin{aligned} &\mathbb{E}[S_k(1,k-1)] \\ &\quad \leq K \, \mathbb{E}\left[\int_0^1 \sqrt{\log N(\eta(\mathbb{P}_{k-1}\Phi^2)^{\frac{1}{2}}, \mathscr{F}, L_2(\mathbb{P}_{k-1}))} d\eta \times (\mathbb{P}_{k-1}\Phi^2)^{\frac{1}{2}}\right], \end{aligned} \quad (2.8.9)$$

where for any semimetric space (T,d), the covering number $N(\eta, T, d)$ is the minimal number of balls of radius η needed to cover T. Then, recalling that a VC-subgraph class satisfies the uniform entropy condition (see for instance [30], Theorem 2.6.7), there exists a constant $C(\mathscr{F})$ which depends only on $\mathscr{F}$ such that

$$\mathbb{E}[S_k(1,k-1)] \leq C(\mathscr{F}) \, \mathbb{E}\left[\Big(\sum_{i=1}^{k-1} a_{i,k}^2 \Phi^2(X_i)\Big)^{\frac{1}{2}}\right]. \quad (2.8.10)$$

Proceeding in the same way for $\mathbb{E}[S(k+1,n)]$, we finally obtain that

$$\begin{aligned} &(\mathbb{E}[S_k(1,k-1) + S_k(k+1,n)])^2 \\ &\quad \leq 4\, C^2(\mathscr{F}) \left(\mathbb{E}\left[\Big(\sum_{i=1}^{k-1} a_{i,k}^2 \Phi^2(X_i)\Big)^{\frac{1}{2}} + \Big(\sum_{i=k+1}^{n} a_{i,k}^2 \Phi^2(X_i)\Big)^{\frac{1}{2}}\right]\right)^2. \end{aligned} \quad (2.8.11)$$

Since $(\sqrt{x}+\sqrt{y})^2 \leq 2\,(x+y)$ for any nonnegative x and y, and $\sum_{k=1}^n \sum_{i\neq k} a_{ik} = 2\sum_{1\leq i<k\leq n} a_{ik}$, we then get by Jensen's inequality

$$\Big(\mathbb{E}[S_k(1,k-1) + S_k(k+1,n)]\Big)^2 \leq 8\, C^2(\mathscr{F}) \Bigg(\sum_{\substack{i=1\\ i\neq k}}^{n} a_{ik}^2\Bigg) \mathbb{E}\left[\Phi^2(X)\right]. \quad (2.8.12)$$

Combining this inequality with (2.8.8), one has (2.8.2) which ends the proof. $\square$

Remark 2.8.4. *If we are concerned with*

$$Z := \sup_{f\in\mathscr{F}} \Big\{ \sum_{1\leq i<j\leq n} a_{ij} f(X_i) f(X_j) \Big\},$$

the same proof applies and we obtain exactly the same inequality (2.8.2).

2.9 Proofs of the results of Sections 2.2 and 2.4

2.9.1 Proofs of Section 2.4

Proof of Lemma 2.4.3. The case $q_0 \in \{0, 1\}$ is straightforward. We now turn to the case $q_0 \in]0, 1[$. By the definition of q_0, it is clear that $\mathbb{E}[X] = \mathbb{E}[\zeta_{q_0}]$. We set in the following $a := a_{q_0}$ and $b := b_{q_0}$. To prove (2.4.4), we consider the following cases separately:

$$(i)\ t \le a, \quad (ii)\ a < t < b, \quad (iii)\ t \ge b.$$

Case *(i)*. Let $t \le a$. Noting that $(y - t)_+ = (y - t) + (t - y)_+$, one has

$$\mathbb{E}[(\zeta_{q_0} - t)_+] = \mathbb{E}[X - t] + \mathbb{E}[(t - \eta)_+].$$

Hence, using the second inequality of (2.4.2), $\mathbb{E}[(X-t)_+] \le \mathbb{E}[(\zeta_{q_0}-t)_+]$.

Case *(ii)*. Let $a < t < b$. A direct calculation leads to

$$\mathbb{E}[(\zeta_{q_0} - t)_+] = q_0(b - t) + \mathbb{E}[(\psi - b)_+].$$

Let $c = (b - t)/(b - a)$, f and g be the functions defined by

$$f(x) := \max\{c(x - a)_+, (x - t)_+\},$$
$$g(x) := ((x - t) - c(x - a))\mathbb{1}_{x \ge b} = (1 - c)(x - b)_+,$$

for all $x \in \mathbb{R}$. Clearly $f(x) = c(x - a)_+ + g(x)$, whence

$$\mathbb{E}[(X - t)_+] \le c\,\mathbb{E}[(X - a)_+] + \mathbb{E}[g(X)]. \tag{2.9.1}$$

Now by the Case (i),

$$\mathbb{E}[(X - a)_+] \le q(b - a) + \mathbb{E}[(\psi - b)_+], \tag{2.9.2}$$

and by the first inequality of (2.4.2),

$$\mathbb{E}[g(X)] \le \mathbb{E}[g(\psi)] = (1 - c)\,\mathbb{E}[(\psi - b)_+]. \tag{2.9.3}$$

Case (ii) follows then from combining (2.9.1)–(2.9.3).

Case *(iii)*. Let $t \ge b$. Clearly, $\mathbb{E}[(\zeta_{q_0} - t)_+] = \mathbb{E}[(\psi - t)_+]$ and the first inequality of (2.4.2) gives the desired inequality.

The proof of (2.4.4) is completed. The extension (2.4.5) to convex functions is classical (see, for example, Proposition 3 in Bentkus [5] or the proof of Theorem 3.3 in Klein, Ma and Privault [15]). $\square$

Proof of Lemma 2.4.6. We recall the assumption that the random variables η and ψ are respectively nonpositive and nonnegative, satisfying (2.4.1). If $\eta = \psi = 0$ almost surely, then all the considered variables are equal to zero and Lemma 2.4.6 is trivial. Assume now that at least one of η or ψ is different from zero. For every $t \in \mathbb{R}$, define the function g_t on $]0,1[$ by $g_t(q) := \mathbb{E}\left[(\zeta_q - \mathbb{E}[\zeta_q] - t)_+^2\right]$. In the sequel, g_t' denote the left derivative of the continuous function g_t. Let (C_q) denote the condition

$$(C_q): \quad 2\,\mathbb{E}[\zeta_q] - a_q - b_q \geq 0.$$

Remark that the left-hand side of (C_q) is nondecreasing in q and tends to a positive value as q tends to 1. Hence $\tilde{q} := \inf\{q \geq 1/2 : b_q + a_q \leq 2\,\mathbb{E}[\zeta_q]\}$ exists, $(C_{\tilde{q}})$ is true and for any $q \geq \tilde{q}$, (C_q) is also verified.

In the following, we link the sign of $g_t'(q)$ with the verification of the condition (C_q). Now,

$$\begin{aligned} g_t'(q) = &- (a_q - \mathbb{E}[\zeta_q] - t)_+^2 - 2\,(b_q - a_q) \int_0^{1-q} (F_\eta^{-1}(u) - \mathbb{E}[\zeta_q] - t)_+ \, du \\ &+ (b_q - \mathbb{E}[\zeta_q] - t)_+^2 - 2\,(b_q - a_q) \int_{1-q}^{1} (F_\psi^{-1}(u) - \mathbb{E}[\zeta_q] - t)_+ \, du. \end{aligned} \tag{2.9.4}$$

We consider the following cases separately:

(i) $t + \mathbb{E}[\zeta_q] \geq F_\psi^{-1}(u)$ for all $u \in]0,1[$.

(ii) $t + \mathbb{E}[\zeta_q] \leq F_\eta^{-1}(u)$ for all $u \in]0,1[$.

(iii) $a_q \leq t + \mathbb{E}[\zeta_q] \leq b_q$.

(iv) $b_q < t + \mathbb{E}[\zeta_q] < F_\psi^{-1}(1-)$.

(v) $F_\eta^{-1}(0+) < t + \mathbb{E}[\zeta_q] < a_q$.

Case *(i)*. All the terms in the right-hand side of (2.9.4) are equal to zero.

Case *(ii)*. ζ_q has a finite second moment and we have $g_t'(q) = d/dq\,\mathrm{Var}(\zeta_q)$. Then, it is elementary to see that $g_t'(q) \leq 0$ if and only if (C_q) is true.

Case *(iii)*. One has

$$g_t'(q) \leq (b_q - \mathbb{E}[\zeta_q] - t)\,(b_q - \mathbb{E}[\zeta_q] - t - 2q(b_q - a_q))\,. \tag{2.9.5}$$

The first factor of the right-hand side of (2.9.5) is nonnegative. Hence the right-hand side of (2.9.5) is negative if and only if, for all t in $[a_q - \mathbb{E}[\zeta_q], b_q - \mathbb{E}[\zeta_q]]$,

$$-\,2q(b_q - a_q) \leq -b_q + \mathbb{E}[\zeta_q] + t. \tag{2.9.6}$$

See now that the right-hand side of (2.9.6) is nondecreasing in t. It thus follows that $-2q(b_q - a_q) \leq -(b_q - a_q)$, or equivalently $q \geq 1/2$, implies that $g_t'(q) \leq 0$.

Case (*iv*). One has directly in this case

$$g_t'(q) = -2\,(b_q - a_q)\int_{1-q}^{1}\left(F_\psi^{-1}(u) - \mathbb{E}[\zeta_q] - t\right)_+ du \le 0.$$

Case (*v*). Define

$$\varepsilon_t := \sup\left\{\theta \in]0,1[: F_\eta^{-1}(\theta) \le t + \mathbb{E}[\zeta_q]\right\}.$$

Then

$$g_t'(q) = (b_q - a_q)\,A_{q,t},$$

where

$$\begin{aligned}
A_{q,t} &= b_q + a_q - 2\,\mathbb{E}[\zeta_q] - 2t \\
&\qquad - 2\int_{1-q}^{1}(F_\psi^{-1}(u) - \mathbb{E}[\zeta_q] - t)du - 2\int_{\varepsilon_t}^{1-q}(F_\eta^{-1}(u) - \mathbb{E}[\zeta_q] - t)\,du \\
&= b_q + a_q - 2\,\mathbb{E}[\zeta_q] + 2\int_0^{\varepsilon_t}(F_\eta^{-1}(u) - \mathbb{E}[\zeta_q] - t)\,du \\
&\le b_q + a_q - 2\,\mathbb{E}[\zeta_q].
\end{aligned}$$

We note that if (C_q) is true, then $A_{q,t} \le 0$, whence $g_t'(q) \le 0$.

Finally, if $q \ge 1/2$ and (C_q) is verified, then $g_t'(q) \le 0$ and the proof of (i) is completed. Let us prove now (ii). Starting with Lemma 2.4.3, we get

$$\mathbb{E}\left[(X - \mathbb{E}[X] - t)_+^2\right] \le \mathbb{E}\left[(\zeta_{q_0} - \mathbb{E}[\zeta_{q_0}] - t)_+^2\right],$$

where q_0 is given by (2.4.3). In particular, $\mathbb{E}[\zeta_{q_0}] = \mathbb{E}[X] \ge 0$. Moreover, since $\eta = -\psi$, $\mathbb{E}[\zeta_{1/2}] = 0$. Recalling that $\mathbb{E}[\zeta_q]$ is nondecreasing with respect to q, it implies that $q_0 \ge 1/2$. Now, see that $a_{1/2} + b_{1/2} \le 0 = \mathbb{E}[\zeta_{1/2}]$. Indeed, for any $u \in]0,1[$,

$$\begin{aligned}
F_\psi^{-1}(u) &\le -\sup\{t \in \mathbb{R} : F_{-\psi}(t) \le 1-u\} \\
&= -F_{-\psi}^{-1}((1-u)+) \\
&\le -F_{-\psi}^{-1}(1-u).
\end{aligned}$$

Thus, $a_{1/2} + b_{1/2} = F_{-\psi}^{-1}(\frac{1}{2}) + F_\psi^{-1}(\frac{1}{2}) \le 0$. By the point (i), we obtain (2.4.6), which concludes the proof. □

Proof of Proposition 2.4.8. Since $X \le X_+$,

$$\|(X+Y)_+\|_p^2 \le \|(X_+ + Y)_+\|_p^2 \le \|X_+ + Y\|_p^2. \tag{2.9.7}$$

Moreover, under the same hypotheses of Proposition 2.4.8, one has the inequality $\|X+Y\|_p^2 \le \|X\|_p^2 + (p-1)\|Y\|_p^2$ as a corollary of Proposition 2.1 of Pinelis [20] (see also Lemma 2.4 of [12] and Proposition 2.1 of [27]). Combining this with (2.9.7) completes the proof. □

2.9.2 Proofs of Section 2.2

Proof of Theorem 2.2.1. Starting from (2.2.1) and projecting on $\mathcal{F}_k$, we obtain $-\mathbb{E}_k[T_k] \leq Z_k - \mathbb{E}_k[Z^{(k)}] \leq \mathbb{E}_k[W_k]$ almost surely. Note that, combining (2.2.2) and the conditional Jensen inequality leads to

$$\begin{aligned}\max\Big(\mathbb{E}[(\mathbb{E}_k[T_k]-t)_+],\ \mathbb{E}[(\mathbb{E}_k[W_k]-t)_+]\Big)\\ \leq \max\Big(\mathbb{E}[(T_k-t)_+],\ \mathbb{E}[(W_k-t)_+]\Big) \leq \mathbb{E}[(\xi_k-t)_+].\end{aligned}$$

Moreover, recalling that the random variables X_k are centered and since F is separately convex, an application of Jensen's inequality ensures that $\mathbb{E}_{k-1}[Z_k - Z^{(k)}] \geq 0$. Thus, conditionally to $\mathcal{F}_{k-1}$, we can apply the second part of Lemma 2.4.6 with $X = Z_k - \mathbb{E}_k[Z^{(k)}]$ and $\psi = \xi_k$. Recalling (2.1.4) and Remark 2.4.7, it yields that for any function φ in $\mathcal{H}_+^2$,

$$\mathbb{E}_{k-1}[\varphi(\Delta_k)] \leq \mathbb{E}\left[\varphi\left(\varepsilon_k Q_{\xi_k}(U_k/2)\right)\right]. \tag{2.9.8}$$

We now prove (2.2.3) by induction on n. The case $n=1$ is given by (2.9.8) with $k=1$. Let $n>1$ and assume that (2.2.3) holds for $n-1$. We then have

$$\begin{aligned}\mathbb{E}[\varphi(Z-\mathbb{E}[Z])] &= \mathbb{E}\left[\mathbb{E}_{n-1}\left[\varphi\left(Z_{n-1}+\Delta_n\right)\right]\right]\\ &\leq \mathbb{E}\left[\varphi\left(Z_{n-1}+\varepsilon_n Q_{\xi_n}(U_n/2)\right)\right]\\ &\leq \mathbb{E}\left[\varphi\Big(\sum_{k=1}^n \varepsilon_k Q_{\xi_k}(U_k/2)\Big)\right],\end{aligned}$$

where we use (2.9.8) in the first inequality and the induction assumption in the second inequality. $\square$

Proof of Theorem 2.2.3. As in the proof of Theorem 2.2.1, we obtain for any function φ in $\mathcal{H}_+^2$,

$$\mathbb{E}_{k-1}[\varphi(\Delta_k)] \leq \mathbb{E}\left[\varphi\left(\varepsilon_k\, \mathbb{E}_{k-1}[\psi_k] Q_{\xi_k}(U_k/2)\right)\right]. \tag{2.9.9}$$

We now prove (2.2.6) by induction on n. For $n=1$, it follows from (2.9.9) for $k=1$ and (2.4.8). Let $n>1$ and assume that (2.2.6) holds for $n-1$. Then

$$\begin{aligned}\|(Z-\mathbb{E}[Z])_+\|_p^2 &\leq \left\|\left(Z_{n-1}+\varepsilon_n\, \mathbb{E}_{n-1}[\psi_n] Q_{\xi_n}(U_n/2)\right)_+\right\|_p^2\\ &\leq \|(Z_{n-1})_+\|_p^2 + (p-1)\|\mathbb{E}_{n-1}[\psi_n] Q_{\xi_n}(U_n/2)\|_p^2\\ &\leq (p-1)\sum_{k=1}^n \|\mathbb{E}_{k-1}[\psi_k]\|_p^2\, \|Q_{\xi_k}(U_k/2)\|_p^2,\end{aligned}$$

where we use (2.9.9) in the first inequality, Proposition 2.4.8 in the second inequality and the induction assumption in the third inequality. $\square$

Remark 2.9.1. *See that, contrary to* (2.9.8)*, there is a* $\mathcal{F}_{k-1}$*-measurable term in the expectation in the right-hand side of* (2.9.9)*, which prevents us to proceed as in the proof of Theorem 2.2.1.*

References

[1] R. Adamczak. A tail inequality for suprema of unbounded empirical processes with applications to Markov chains. *Electron. J. Probab.*, 13:no. 34, 1000–1034, 2008.

[2] R. Adamczak and P. Wolff. Concentration inequalities for non-lipschitz functions with bounded derivatives of higher order. *Probability Theory and Related Fields*, 162(3):531–586, 2015.

[3] V. Bentkus. On Hoeffding's inequalities. *Ann. Probab.*, 32(2):1650–1673, 04 2004.

[4] V. Bentkus. An extension of the Hoeffding inequality to unbounded random variables. *Lithuanian Mathematical Journal*, 48(2):137–157, 2008.

[5] V. Bentkus. Addendum to "An extension of an inequality of Hoeffding to unbounded random variables": The non-i.i.d. case. *Lithuanian Mathematical Journal*, 48(3):237–255, 2008.

[6] V. Bentkus. Bounds for the stop loss premium for unbounded risks under the variance constraints. Preprint on https://www.math.uni-bielefeld.de/sfb701/preprints/view/423, 2010.

[7] V. Bentkus, N. Kalosha, and M. Zuijlen. On domination of tail probabilities of (super)martingales: Explicit bounds. *Lithuanian Mathematical Journal*, 46(1):1–43, 2006.

[8] B. Bercu, B. Delyon, and E. Rio. *Concentration Inequalities for Sums and Martingales*. SpringerBriefs in Mathematics. Springer, 2015.

[9] S. Boucheron, O. Bousquet, G. Lugosi, and P. Massart. Moment inequalities for functions of independent random variables. *Ann. Probab.*, 33(2):514–560, 03 2005.

[10] S. Boucheron, G. Lugosi, and P. Massart. *Concentration Inequalities: A Nonasymptotic Theory of Independence*. OUP Oxford, 2013.

[11] M. Csörgö and L. Horváth. *Weighted Approximations in Probability and Statistics*. Wiley Series in Probability and Statistics. Wiley, 1993.

[12] L. Dümbgen, S. van de Geer, M.C. Veraar, and J.A. Wellner. Nemirovski's inequalities revisited. *The American mathematical monthly: the official journal of the Mathematical Association of America*, 117(2):138, 2010.

[13] T. Figiel, P. Hitczenko, W.B. Johnson, G. Schechtman, and J. Zinn. Extremal properties of Rademacher functions with applications to the Khintchine and Rosenthal inequalities. *Trans. Amer. Math. Soc.*, 349(3):997–1027, 1997.

[14] W. Hoeffding. Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.*, 58:13–30, 1963.

[15] T. Klein, Y. Ma, and N. Privault. Convex concentration inequalities and forward-backward stochastic calculus. *Electron. J. Probab.*, 11:no. 20, 486–512, 2006.

[16] A. Kontorovich. Concentration in unbounded metric spaces and algorithmic stability. In *ICML*, pages 28–36, 2014.

[17] J. Lederer and S. van de Geer. New concentration inequalities for suprema of empirical processes. *Bernoulli*, 20(4):2020–2038, 11 2014.

[18] M. Ledoux. On Talagrand's deviation inequalities for product measures. *ESAIM: Probability and Statistics*, 1:63–87, 1997.

[19] A. Marchina. Concentration inequalities for separately convex functions. *To appear in Bernoulli Journal*, 2018+. Preprint on <hal-01344861>.

[20] I. Pinelis. On a majorization inequality for sums of independent random vectors. *Statistics & Probability Letters*, 19(2):97 – 99, 1994.

[21] I. Pinelis. Optimum bounds for the distributions of martingales in banach spaces. *The Annals of Probability*, pages 1679–1706, 1994.

[22] I. Pinelis. Optimal tail comparison based on comparison of moments. In *High dimensional probability*, pages 297–314. Springer, 1998.

[23] I. Pinelis. Fractional sums and integrals of r-concave tails and applications to comparison probability inequalities. *Advances in stochastic inequalities (Atlanta, GA, 1997)*, 234:149–168, 1999.

[24] I. Pinelis. An optimal three-way stable and monotonic spectrum of bounds on quantiles: a spectrum of coherent measures of financial risk and economic inequality. *Risks*, 2(3):349–392, 2014.

[25] I. Pinelis. Convex cones of generalized multiply monotone functions and the dual cones. *Banach J. Math. Anal.*, 10(4):864–897, 10 2016.

[26] I.F. Pinelis and A.I. Sakhanenko. Remarks on inequalities for large deviation probabilities. *Theory Probab. Appl.*, 30:143–148, 1986.

[27] E. Rio. Moment inequalities for sums of dependent random variables under projective conditions. *Journal of Theoretical Probability*, 22(1):146–163, 2009.

[28] E. Rio. *Asymptotic theory of weakly dependent random processes*, volume 80 of *Probability Theory and Stochastic Modelling*. Springer, Berlin, 2017. Translated from the 2000 French edition [MR2117923].

[29] S. van de Geer and J. Lederer. The Bernstein-Orlicz norm and deviation inequalities. *Probab. Theory Related Fields*, 157(1-2):225–250, 2013.

[30] A.W. van der Vaart and J.A. Wellner. *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer Series in Statistics. Springer New York, 1996.

Part II

Concentration inequalities for suprema of empirical processes

Chapter 3

Concentration inequalities for suprema of unbounded empirical processes

Using martingale methods, we obtain some Fuk-Nagaev type inequalities for suprema of unbounded empirical processes associated with independent and identically distributed random variables. We then derive weak and strong moment inequalities. Next, we apply our results to suprema of empirical processes which satisfy a power-type tail condition.

This Chapter is adapted from the work [13].

3.1 Introduction

Let us consider a sequence $X_1, X_2, \ldots$ of independent random variables valued in some measurable space $(\mathcal{X}, \mathcal{F})$. Let P_n denote for every integer n the empirical probability measure $P_n := n^{-1}(\delta_{X_1} + \ldots + \delta_{X_n})$. Let $\mathscr{F}$ be a countable class of measurable functions from $\mathcal{X}$ into $\mathbb{R}$ such that $\mathbb{E}[f(X_k)] = 0$ for all $f \in \mathscr{F}$ and all $k = 1, \ldots, n$. We assume that $\mathscr{F}$ has a square integrable envelope function Φ, that is

$$|f| \leq \Phi \text{ for any } f \in \mathscr{F}, \text{ and } \Phi \in \mathbb{L}^2. \tag{3.1.1}$$

As in Boucheron, Lugosi and Massart [5], we define the wimpy variance σ^2 and the weak variance Σ^2 by

$$\sigma^2 := \sup_{f \in \mathscr{F}} \frac{1}{n} \sum_{k=1}^{n} \mathbb{E}[f^2(X_k)], \text{ and } \Sigma^2 := \mathbb{E}\Big[\sup_{f \in \mathscr{F}} \frac{1}{n} \sum_{k=1}^{n} f^2(X_k)\Big]. \tag{3.1.2}$$

Let us also define

$$E_k := \mathbb{E} \sup_{f\in\mathscr{F}} P_k(f) \quad \text{for any } k = 1\ldots, n. \tag{3.1.3}$$

The purpose of this chapter is to provide concentration inequalities around its mean for the random variable

$$Z := \sup\{nP_n(f) : f \in \mathscr{F}\}, \tag{3.1.4}$$

involving σ^2, and under the additional assumption of identically distributed data. Our approach is based on a decomposition of Z into a sum of two martingales. Then, we control each martingale separately by Fuk-Nagaev type inequalities: in a first part, by one found by Courbot [8], which allows us to derive a strong moment inequality (following Petrov [16]), and in a second part, by one found recently by Rio [19], which allows us to derive a weak moment inequality. We stress out that we only require that the envelope function Φ has an ℓth weak or strong moment, while classical concentration inequalities for suprema of empirical processes assume uniform boundedness condition on the elements of $\mathscr{F}$. Let us recall a main result in this direction: the following Bennett type inequality obtained by Bousquet [6], which is an improvement of Theorem 1.1 in Rio [18]:

Theorem 3.1.1 ([6], Theorem 7.3)**.** *Let $X_1, \ldots, X_n$ be a sequence of independent random variables with values in $\mathcal{X}$ and distributed according to P. Assume that $P(f) = 0$ and $f \leq 1$ for all $f \in \mathscr{F}$. Let Z be defined by* (3.1.4) *and set $v_n := n\sigma^2 + 2\,\mathbb{E}[Z]$, where σ^2 is defined in* (3.1.2)*. Let h be the function defined, for any $u \geq -1$, by $h(u) := (1+u)\log(1+u) - u$. Then, for all $t \geq 0$,*

$$\begin{aligned}\mathbb{P}(Z - \mathbb{E}[Z] \geq t) &\leq \exp\Big(-v_n\, h\Big(\frac{t}{v_n}\Big)\Big) && \text{(Bousquet's inequality)}\\ &\leq \exp\Big(-\frac{t}{2}\log\Big(1+\frac{t}{v_n}\Big)\Big). && \text{(Rio's inequality)}\end{aligned}$$

We refer the reader to Section 12 of [5] for an overview of the bounded case. Here we are interested in unbounded functions. Few results in the literature concern concentration inequalities for suprema of unbounded empirical processes. Let us first mention the considerable work of Boucheron, Bousquet, Lugosi and Massart [4], concerning moment inequalities for general functions of independent random variables. Their methods are based on an extension of the entropy method proposed by Ledoux [12]. In particular, they establish the following generalized moment inequality for suprema of (possibly unbounded) empirical processes involving σ^2 and Σ^2:

Theorem 3.1.2 ([5], Theorems 15.14 and 15.5)**.** *Let $X_1, \ldots, X_n$ be a sequence of independent random variables with values in $\mathcal{X}$. Assume that for all $f \in \mathscr{F}$ and all $k = 1, \ldots, n$, $\mathbb{E}[f(X_k)] = 0$. Let Z be defined by*

$$Z := \sup_{f \in \mathcal{F}} \left| \sum_{k=1}^{n} f(X_k) \right|.$$

Let $M := \max_{k=1,\ldots,n} \Phi(X_k)$, where Φ is defined in (3.1.1). Then, for all $\ell \geq 2$,

$$\|(Z - \mathbb{E}[Z])_+\|_\ell \leq \sqrt{n\kappa(\ell-1)}(\Sigma + \sigma) + \kappa(\ell - 1)\Big(\|M\|_\ell + \sup_{\substack{f \in \mathscr{F} \\ k=1,\ldots,n}} \|f(X_k)\|_2\Big),$$

where σ^2, Σ^2 are defined in (3.1.2) and $\kappa := \sqrt{e}/(\sqrt{e} - 1)$.

For several reasons (see, for instance, the discussion after Theorem 3 in Adamczack [1]) one would like to express the variance factor in terms of σ^2 rather than Σ^2. First, observe that Σ^2 is greater than σ^2. In the bounded case, an application of the contraction principle gives $n\Sigma^2 \leq n\sigma^2 + 16\,\mathbb{E}[Z]$, when $|f| \leq 1$ for any $f \in \mathscr{F}$ (see Corollary 15 in Massart [15]). However, in the unbounded case, Σ^2 is more difficult to compare to σ^2. In the setting of Theorem 3.1.2, one can only prove the much less efficient inequality

$$n\Sigma^2 \leq n\sigma^2 + 32\sqrt{\mathbb{E}[M^2]}\mathbb{E}[Z] + 8\,\mathbb{E}[M^2],$$

(see Theorem 11.17 and Section 15 in [5]). Similarly to the bounded case, the bounds that we will obtain in this chapter will involve σ^2, and the expectations E_k rather than the weak variance Σ^2. Furthermore, we shall prove in a particular case that our bounds provide a much more accurate estimate of the variance.

Einmahl and Li [9] prove a Fuk-Nagaev type inequality for suprema of empirical processes involving σ^2 and the ℓth strong moment of the envelope function. They use an improvement of Bousquet's inequality for suprema of bounded empirical processes to nonnecessarily identically distributed random variables obtained by Klein and Rio [10], a truncation argument and the so-called Hoffman-Jørgensen inequality. Using similar techniques, Adamczak [1] provides a concentration inequality for suprema of empirical processes under a semi-exponential tail condition on the envelope function Φ of $\mathscr{F}$:

Theorem 3.1.3 ([1], Theorem 4)**.** *Let $X_1, \ldots, X_n$ be a sequence of independent random variables with values in $\mathcal{X}$. Assume that $\mathbb{E}[f(X_k)] = 0$ for all $f \in \mathscr{F}$ and all $k = 1, \ldots, n$. For all $\alpha \in]0, 1]$, let ψ_α be the function defined,* *for any $x > 0$, by $\psi_\alpha(x) := \exp(x^\alpha) - 1$ and let $\|.\|_{\psi_\alpha}$ denote the associated Orlicz norm, which is defined by*

$$\|X\|_{\psi_\alpha} := \inf\{\lambda > 0 : \mathbb{E}[\psi_\alpha(|X|/\lambda)] \leq 1\}, \text{ for all random variables } X.$$

Assume now that for some $\alpha \in]0, 1]$, $\|\Phi(X_k)\|_{\psi_\alpha} < \infty$ for all $k = 1, \dots n$, where Φ is defined in (3.1.1). Let Z be defined by $Z := \sup_{f\in\mathcal{F}} |\sum_{k=1}^n f(X_k)|$. Then, for all $0 < \eta < 1$ and $\delta > 0$, there exists a constant $C = C(\alpha, \eta, \delta)$, such that, for all $t \geq 0$,

$$\begin{aligned}&\mathbb{P}(Z - (1+\eta)\mathbb{E}[Z] \geq t)\\ &\quad\leq \exp\Big(-\frac{t^2}{2(1+\delta)n\sigma^2}\Big) + 3\exp\Big(-\Big(\frac{t}{C\|\max_{k=1,\dots,n}\Phi(X_k)\|_{\psi_\alpha}}\Big)^\alpha\Big).\end{aligned}$$

Let us point out that the upper bound in the inequality above (and also in [9]) do not involve $\mathbb{E}[Z]$ or the entropy of the class $\mathscr{F}$. The price to be paid is the additional factor $1 + \eta$ in front of $\mathbb{E}[Z]$ and the non explicit constant $C(\alpha, \eta, \delta)$. More recently, van de Geer and Lederer [20] introduce a new Orlicz norm (called Bernstein-Orlicz norm), and under some Bernstein conditions satisfied by the envelope function Φ, they derive exponential inequalities. Their upper bounds involve the constant K of the Bernstein conditions and $\mathbb{E}[Z]$ (which is bounded up in terms of the complexity of $\mathscr{F}$ and K). Next, the same authors in [11], require only that the envelope function Φ has an ℓth strong moment and obtained deviation and moment inequalities involving σ^2. However, it concerns the deviation of Z around $(1+\eta)\mathbb{E}[Z]$. Finally, Marchina [14] provides deviation inequalities around $\mathbb{E}[Z]$ for suprema of randomized unbounded empirical processes involving only the envelope function Φ, see for example the following proposition:

Proposition 3.1.4 ([14], Proposition 7.14)**.** *Let $X_1, \dots, X_n$ be a sequence of independent random variables with values in $\mathcal{X}$. Let $Y_1, \dots, Y_n$ be a sequence of independent real-valued symmetric random variables such that the two sequences are independent. Let $\mathscr{F}$ be a countable class of measurable functions $f : \mathcal{X} \to \mathbb{R}$ such that $-G \leq f \leq H$ for all $f \in \mathscr{F}$, where G and H are nonnegative functions. Define $Z = \sup_{f\in\mathscr{F}} \sum_{k=1}^n Y_k f(X_k)$. Let*

$$s_k^2 := \mathbb{E}[Y_k^2]\,\mathbb{E}[H^2(X_k) + G^2(X_k)] \text{ and } s^2 := \sum_{k=1}^n s_k^2.$$

Then, for any $2 \leq \ell \leq 4$,

$$\|(Z - \mathbb{E}[Z])_+\|_\ell^\ell \leq \frac{1}{2}\sum_{k=1}^n \mathbb{E}\Big[|Y_k|^\ell(H^\ell(X_k) + G^\ell(X_k))\Big] + \frac{1}{2}\,s^\ell\|g\|_\ell^\ell,$$

where g is a standard Gaussian random variable.

The results of [14] are based on martingale techniques. The purpose of the present chapter is to introduce the wimpy variance in the concentration inequalities derived from the martingale approach. We shall give deviation inequalities around $\mathbb{E}[Z]$, without extra centering term $\eta\mathbb{E}[Z]$ and with explicit constants.

The chapter is organized as follows: we first recall some definitions and notations in Section 3.2. In Section 3.3, we state Fuk-Nagaev type inequalities for $Z - \mathbb{E}[Z]$ and the resulting corollaries concerning the weak and strong moments of order $\ell > 2$. We also shall apply the Fuk-Nagaev inequalities to bound up the generalized moment $\mathbb{E}[(Z - \mathbb{E}[Z] - t)_+]$. Finally, in Section 3.4, we apply the main results to the special case $Z = \sup_{g\in\mathscr{G}} \sum_{k=1}^n Y_k\, g(X_k)$ where Y_k satisfies a power-type tail condition and $\mathscr{G}$ is a class of bounded functions.

3.2 Definitions and notations

In this section, we give the notations and definitions which we will use all along the chapter. Let us start with the classical notations $x_+ := \max(0, x)$ and $x_+^\alpha := (x_+)^\alpha$ for all reals x and α. Next, we define the tail function, the quantile function and the Conditional Value-at-Risk.

Definition 3.2.1. *Let X be a real-valued random variable.*
(i) The distribution function of X is denoted by F_X and the càglàd inverse of F_X is denoted by F_X^{-1}.
(ii) The quantile function of X, which is the càdlàg inverse of the tail function $t \mapsto 1 - F_X(t)$, is denoted by Q_X.
(iii) Assume that X is integrable. The integrated quantile function $\tilde{Q}_X$ of X, which is also known as the Conditional Value-at-Risk (CVaR for short), is defined by $\tilde{Q}_X(u) := u^{-1}\int_0^u Q_X(s)ds$.

We recall the following elementary properties of these quantities, which are given and proved by Pinelis [17].

Proposition 3.2.2. *Let X and Y be real-valued and integrable random variables. Then, for any $u \in]0, 1]$,*

(i) $\mathbb{P}(X > Q_X(u)) \leq u$,

(ii) $Q_X(u) \leq \tilde{Q}_X(u)$,

(iii) $\tilde{Q}_{X+Y}(u) \leq \tilde{Q}_X(u) + \tilde{Q}_Y(u)$.

Let us now define the following class of distribution functions.

Notation 3.2.3. *Let* $q \in [0,1]$*. Let* ψ *be a nonnegative random variable and set* $b_{\psi,q} := F_\psi^{-1}(1-q)$*. We denote by* $F_{\psi,q}$ *the distribution function defined by*

$$F_{\psi,q}(x) := (1-q)\mathbb{1}_{0 \le x < b_{\psi,q}} + F_\psi(x)\mathbb{1}_{x \ge b_{\psi,q}}. \tag{3.2.1}$$

These distribution functions will be used to bound up the generalized moments of nonnegative random variables which are dominated by ψ. Precisely, let X be a nonnegative random variable stochastically dominated by ψ, that is $\mathbb{P}(X > x) \le \mathbb{P}(\psi > x)$ for all $x > 0$. Let $\zeta_{\psi,q}$ be a random variable with distribution function $F_{\psi,q}$, where q is such that $\mathbb{E}[X] = \mathbb{E}[\zeta_{\psi,q}]$. Then Lemma 1 of Bentkus [3] (see also Lemma 2.1 of Marchina [14]) ensures that for any function $\varphi \in \mathcal{H}_+^1$, $\mathbb{E}[\varphi(X)] \le \mathbb{E}[\varphi(\zeta_{\psi,q})]$, where $\mathcal{H}_+^1$ is the class of numerical functions φ defined by

$$\mathcal{H}_+^1 := \Big\{\varphi : \varphi \text{ is convex, differentiable, and } \lim_{x\to-\infty} \varphi(x) = 0\Big\}. \tag{3.2.2}$$

Now, we recall the definitions of strong and weak norms of a real-valued random variable X. For all $r \ge 1$, let $\mathbb{L}^r$ be the space of real-valued random variables with a finite absolute moment of order r and we denote by $\|X\|_r$ the $\mathbb{L}^r$-norm of X. Let

$$\Lambda_r^+(X) := \sup_{t>0} t\,(\mathbb{P}(X > t))^{1/r}. \tag{3.2.3}$$

We say that X has a weak moment of order r if $\Lambda_r^+(|X|)$ is finite. Define also

$$\tilde{\Lambda}_r^+(X) := \sup_{u\in]0,1]} u^{(1/r)-1} \int_0^u Q_X(s)ds. \tag{3.2.4}$$

From the definition of Q_X, we have (see, for instance, Chapter 4 of Bennett and Sharpley [2])

$$\Lambda_r^+(X) = \sup_{u\in]0,1]} u^{1/r} Q_X(u). \tag{3.2.5}$$

Hence, we get that

$$\Lambda_r^+(X) \le \tilde{\Lambda}_r^+(X) \le \Big(\tfrac{r}{r-1}\Big)\Lambda_r^+(X). \tag{3.2.6}$$

Furthermore, from Proposition 3.2.2 (*iii*), $\tilde{\Lambda}_r^+(.)$ is subadditive.

3.3 Statement of results

Let us first recall the assumptions we work with. Let $X_1, \dots, X_n$ be a sequence of independent random variables valued in $\mathcal{X}$, with common distribution P. Let $\mathscr{F}$ be a countable class of measurable functions $f : \mathcal{X} \to \mathbb{R}$

such that $P(f) = 0$ for all $f \in \mathscr{F}$, and we suppose that $\mathscr{F}$ has a square integrable envelope function Φ defined in (3.1.1). In this situation, the wimpy variance is $\sigma^2 = \sup_{f\in\mathscr{F}} P(f^2)$. We consider the random variable

$$Z = \sup_{f\in\mathscr{F}} \sum_{k=1}^{n} f(X_k). \tag{3.3.1}$$

Throughout the rest of the chapter, ζ_k denotes a random variable with distribution function $F_{2\,\Phi(X_1),\,q_k}$ defined in (3.2.1) where q_k is the real in $[0,1]$ such that $\mathbb{E}[\zeta_k] = E_k$ (E_k is defined in (3.1.3)). We also set

$$V_n := \sum_{k=1}^{n} \mathbb{E}[\zeta_k^2]. \tag{3.3.2}$$

Remark 3.3.1. *If the class $\mathscr{F}$ satisfies the uniform law of large numbers, that is $\sup_{f\in\mathscr{F}} |P_n(f)|$ converges to 0 in probability, then E_n decreases to 0 (see, for instance, Section 2.4 of van der Vaart and Wellner [21]). Now, from the integrability of Φ^2 and (3.2.1), the convergence of E_n to 0 implies the convergence of $\mathbb{E}[\zeta_n^2]$ to 0, which ensures that V_n/n tends to 0. More precise estimates of V_n will be proved for particular cases in Section 3.4.*

We first derive a Fuk-Nagaev type inequality for $Z - \mathbb{E}[Z]$ from one obtained by Courbot [8] concerning martingales.

Theorem 3.3.2. *Let $x > 0$. For any $s > 0$, we have*

$$\begin{aligned}
\mathbb{P}(Z - \mathbb{E}[Z] \geq x) &\leq \mathbb{P}((Z - \mathbb{E}[Z])_+ \geq x) && (a)\\
&\leq \Big(1 + \frac{x^2}{4sn\sigma^2}\Big)^{-s/2} + \Big(1 + \frac{x^2}{4sV_n}\Big)^{-s/2} && (b)\\
&\qquad + 2n\,\mathbb{P}\Big(\Phi(X_1) \geq \frac{x}{4s}\Big).
\end{aligned}$$

Next, under weak moment conditions, we derive from a Fuk-Nagaev type inequality for martingales with efficient constants obtained recently by Rio [19], another Fuk-Nagaev type inequality for $Z - \mathbb{E}[Z]$.

Theorem 3.3.3. *Let $\ell > 2$. Assume that $\Phi(X_1)$ have a weak moment of order ℓ. Then for any $u \in]0,1[$,*

$$\begin{aligned}
Q_{Z-\mathbb{E}[Z]}(u) &\leq \tilde{Q}_{Z-\mathbb{E}[Z]}(u) && (a)\\
&\leq \sqrt{2\log(1/u)}\,(\sigma\sqrt{n} + \sqrt{V_n}\,) + 3\,n^{1/\ell}\mu_\ell\,\Lambda_\ell^+(\Phi(X_1))u^{-1/\ell}, && (b)
\end{aligned}$$

where $\mu_\ell := 2 + \max(4/3, \ell/3)$. Consequently,

$$\mathbb{P}\Big(Z - \mathbb{E}[Z] > \sqrt{2\log(1/u)}\,\Big(\sigma\sqrt{n} + \sqrt{V_n}\Big) + 3\,n^{1/\ell}\mu_\ell\,\Lambda_\ell^+(\Phi(X_1))u^{-1/\ell}\Big) \leq u. \qquad (c)$$

In the two following results, we derive from Theorems 3.3.2 and 3.3.3, strong and weak moment inequalities for $Z - \mathbb{E}[Z]$.

Corollary 3.3.4. *Let $\ell \geq 2$. Assume that $\Phi(X_1)$ is $\mathbb{L}^\ell$-integrable. Then*

$$\|(Z-\mathbb{E}[Z])_+\|_\ell \leq 2\beta_\ell\, \ell^{1/\ell}\sqrt{\ell+1}\left(\sigma\sqrt{n}+\sqrt{V_n}\right)+2^{2+1/\ell}\, n^{1/\ell}(\ell+1)\|\Phi(X_1)\|_\ell,$$

where $\beta_\ell := (\sqrt{\pi}/2)\,\Gamma(\ell/2)/\,\Gamma((\ell+1)/2)$.

Remark 3.3.5. *Note that $\ell^{1/\ell} \leq e^{1/e} \simeq 1.4447$. Furthermore, $\beta_2 = 1$ and $\ell \mapsto \beta_\ell$ decreases to 0 as ℓ tends to ∞.*

Remark 3.3.6. *By analyzing the proofs of Theorem 3.3.2 and Corollary 3.3.4, we can slightly improve the constant $2^{1+1/\ell}$ to $(1+2^\ell)^{1/\ell}$.*

Corollary 3.3.7. *Let $\ell > 2$. Assume that $\Phi(X_1)$ have a weak moment of order ℓ. Then*

$$\begin{aligned} \Lambda_\ell^+(Z-\mathbb{E}[Z]) &\leq \tilde{\Lambda}_\ell^+(Z-\mathbb{E}[Z]) && (a)\\ &\leq \sqrt{(\ell/e)}\,(\sigma\sqrt{n}+\sqrt{V_n}\,)+3\,n^{1/\ell}\mu_\ell\,\Lambda_\ell^+(\Phi(X_1)), && (b) \end{aligned}$$

where $\mu_\ell := 2+\max(4/3, \ell/3)$.

3.3.1 Bound of generalized moments of $Z - \mathbb{E}[Z]$

In this section, we apply Theorem 3.3.3 to bound up $\mathbb{E}[(Z-\mathbb{E}[Z]-t)_+]$ for every $t > 0$. We emphasize that it is of interest to obtain such bounds in various situations coming from statistical applications, such the study of rates of convergence for estimators (see, for instance, Comte and Lacour [7]).

Proposition 3.3.8. *Let Z, σ, V_n be defined as in Section 3.3. Let $\ell > 2$ and $\mu_\ell = 2+\max(4/3, \ell/3)$. Set also*

$$s_n := \sigma\sqrt{n}+\sqrt{V_n}, \text{ and } b_{n,\ell} := 3\,n^{1/\ell}\mu_\ell\,\Lambda_\ell^+(\Phi(X_1)).$$

Then, for any $t > 0$,

$$\mathbb{E}[(Z-\mathbb{E}[Z]-t)_+] \leq s_n\,\frac{e^{-\frac{1}{2}(1+t^2/s_n^2)}}{\sqrt{1+t^2/s_n^2}}+b_{n,\ell}.$$

Proof. Let us start by recalling the variational expression of $\mathbb{E}[(X-t)_+]$ involving $\tilde{Q}_X$. Since $x < Q_X(u)$ if and only if $1 - F_X(x) > u$, we get for any $t \in \mathbb{R}$,

$$\mathbb{E}[(X-t)_+] = \sup_{u \in]0,1]} u(\tilde{Q}_X(u) - t). \tag{3.3.3}$$

Now, Equality (3.3.3) and Theorem 3.3.3 (b) imply

$$\begin{aligned}\mathbb{E}[(Z - \mathbb{E}[Z] - t)_+] &\leq \sup_{u \in]0,1]} u\Big(s_n\sqrt{2\log(1/u)} + b_{n,\ell}\, u^{-1/\ell} - t\Big) \\ &\leq \sup_{u \in]0,1]} u\Big(s_n\sqrt{2\log(1/u)} - t\Big) + b_{n,\ell}, \end{aligned} \tag{3.3.4}$$

since $u^{1-1/\ell} \leq 1$. With the change of variables $y = \sqrt{2\log(1/u)} \in [0, \infty[$, clearly, the supremum is achieved at

$$y_0 := \frac{t}{2\, s_n} + \sqrt{1 + \frac{t^2}{4\, s_n^2}}. \tag{3.3.5}$$

Then, the supremum in (3.3.4) is equal to $s_n\, e^{-y_0^2/2}/y_0$. Observing now that $y_0 \geq \sqrt{1 + t^2/s_n^2}$, we finally get the desired inequality which concludes the proof. $\square$

Remark 3.3.9. *As starting point of the proof, in place of* (3.3.3)*, we can use the equality* $Z - \mathbb{E}[Z] \overset{\mathcal{D}}{=} Q_{Z-\mathbb{E}[Z]}(U)$*, where* U *is a random variable distributed uniformly on* $[0,1]$*. However, contrary to the above proof, we then need to integrate Inequality (b) of Theorem 3.3.3, which shows the interest of the CVaR.*

3.3.2 Proofs of the main results

Our method is based on a martingale decomposition of Z which we now recall. We suppose that $\mathscr{F}$ is a finite class of functions, that is $\mathscr{F} = \{f_i : i \in \{1, \dots, m\}\}$. The results in the countable case are derived from the finite case using the monotone convergence theorem. We define $\mathcal{F}_0 := \{\varnothing, \Omega\}$ and for all $k = 1, \dots, n$, $\mathcal{F}_k := \sigma(X_1, \dots, X_k)$ and $\mathcal{F}_n^k := \sigma(X_1, \dots, X_{k-1}, X_{k+1}, \dots, X_n)$. Let $\mathbb{E}_k$ (respectively $\mathbb{E}_n^k$) denote the conditional expectation operator associated with $\mathcal{F}_k$ (resp. $\mathcal{F}_n^k$). Set also

$$Z_k := \mathbb{E}_k[Z], \tag{3.3.6}$$

$$Z^{(k)} := \sup\{nP_n(f) - f(X_k) : f \in \mathscr{F}\}. \tag{3.3.7}$$

The sequence (Z_k) is an $(\mathcal{F}_k)$-adapted martingale (the Doob martingale associated with $Z - \mathbb{E}[Z]$) and

$$Z - \mathbb{E}[Z] = \sum_{k=1}^{n} \Delta_k, \qquad \text{where } \Delta_k := Z_k - Z_{k-1}. \tag{3.3.8}$$

Define now the random variables τ and τ_k, respectively $\mathcal{F}_n$-measurable and $\mathcal{F}_n^k$-measurable, by

$$\tau := \inf\{i \in \{1, \dots, m\} : nP_n(f_i) = Z\}, \tag{3.3.9}$$

$$\tau_k := \inf\{i \in \{1, \dots, m\} : nP_n(f_i) - f_i(X_k) = Z^{(k)}\}. \tag{3.3.10}$$

Notice first that

$$Z^{(k)} + f_{\tau_k}(X_k) \leq Z \leq Z^{(k)} + f_\tau(X_k).$$

From this, conditioning by $\mathcal{F}_k$ gives

$$\mathbb{E}_k[f_{\tau_k}(X_k)] \leq Z_k - \mathbb{E}_k[Z^{(k)}] \leq \mathbb{E}_k[f_\tau(X_k)]. \tag{3.3.11}$$

Set now $\xi_k := \mathbb{E}_k[f_{\tau_k}(X_k)]$ and let $\varepsilon_k \geq r_k \geq 0$ be random variables such that

$$\xi_k + r_k = Z_k - \mathbb{E}_k[Z^{(k)}] \qquad \text{and} \qquad \xi_k + \varepsilon_k = \mathbb{E}_k[f_\tau(X_k)].$$

Thus (3.3.11) becomes

$$\xi_k \leq \xi_k + r_k \leq \xi_k + \varepsilon_k. \tag{3.3.12}$$

Since the random variables τ_k is $\mathcal{F}_n^k$-measurable, we have by the centering assumption on the elements of $\mathscr{F}$,

$$\mathbb{E}_n^k[f_{\tau_k}(X_k)] = P(f_{\tau_k}) = 0, \tag{3.3.13}$$

which ensures that $\mathbb{E}_{k-1}[\xi_k] = 0$. Moreover, $\mathbb{E}_k[Z^{(k)}]$ is $\mathcal{F}_{k-1}$-measurable. Hence we get

$$\Delta_k = Z_k - \mathbb{E}_k[Z^{(k)}] - \mathbb{E}_{k-1}[Z_k - \mathbb{E}_k[Z^{(k)}]] = \xi_k + r_k - \mathbb{E}_{k-1}[r_k],$$

which, combined with (3.3.8), yields the decomposition of $Z - \mathbb{E}[Z]$ in a sum of two martingales:

$$Z - \mathbb{E}[Z] = \Xi_n + R_n, \tag{3.3.14}$$

where

$$\Xi_n := \sum_{k=1}^{n} \xi_k \text{ and } R_n := \sum_{k=1}^{n} (r_k - \mathbb{E}_{k-1}[r_k]). \tag{3.3.15}$$

Before proving the results, we provide bounds for their quadratic variations which will be needed in the proofs.

(i) Bound of $\sum_{k=1}^{n} \mathbb{E}_{k-1}[\xi_k^2]$.
Notice that the same argument as in (3.3.13) yields $\mathbb{E}_n^k[f_{\tau_k}^2(X_k)] = P(f_{\tau_k}^2)$. It follows from the conditional Jensen inequality that $\sum_{k=1}^{n} \mathbb{E}_{k-1}[\xi_k^2] \leq n\sigma^2$.

(ii) Bound of $\sum_{k=1}^{n} \mathbb{E}_{k-1}[(r_k - \mathbb{E}_{k-1}[r_k])^2]$.
First, we observe that $\mathbb{E}_{k-1}[r_k]$ is bounded by a deterministic constant. This is given by the following lemma of exchangeability of variables.

Lemma 3.3.10. *For any integer $j \geq k$, $\mathbb{E}_{k-1}[f_\tau(X_k)] = \mathbb{E}_{k-1}[f_\tau(X_j)]$.*

Proof. By the definition of the random variable τ, for every permutation on n elements σ, $\tau(X_1, \ldots, X_n) = \tau \circ \sigma(X_1, \ldots, X_n)$ almost surely. Applying now this fact to $\sigma = (k\, j)$ (the transposition which exchanges k and j), it suffices to use Fubini's theorem (recalling that $j \geq k$) to complete the proof. □

Hence,

$$\begin{aligned}
\mathbb{E}_{k-1}[\varepsilon_k] &= \mathbb{E}_{k-1}[f_\tau(X_k)] \\
&= \mathbb{E}_{k-1}[f_\tau(X_k) + \ldots + f_\tau(X_n)]/(n-k+1) \\
&\leq \mathbb{E}_{k-1} \sup_{f \in \mathscr{F}} \{f(X_k) + \ldots + f(X_n)\}/(n-k+1) = E_{n-k+1}. \quad (3.3.16)
\end{aligned}$$

Since $0 \leq r_k \leq \varepsilon_k$, we thus get that $0 \leq \mathbb{E}_{k-1}[r_k] \leq E_{n-k+1}$.

Moreover, (3.3.12) implies that $0 \leq r_k \leq 2\,\Phi(X_k)$. Then Lemma 1 of Bentkus [3] ensures that for any function $\varphi \in \mathcal{H}_+^1$, $\mathbb{E}_{k-1}[\varphi(r_k)] \leq \mathbb{E}[\varphi(\zeta_{n-k+1})]$, where $\mathcal{H}_+^1$ is defined in (3.2.2). Notice that $x \mapsto x_+^2$ belongs to $\mathcal{H}_+^1$ and $r_{k+} = r_k$, we get

$$\sum_{k=1}^{n} \mathbb{E}_{k-1}[(r_k - \mathbb{E}_{k-1}[r_k])^2] \leq \sum_{k=1}^{n} \mathbb{E}_{k-1}[r_{k+}^2] \leq \sum_{k=1}^{n} \mathbb{E}[\zeta_{n-k+1}^2] = \sum_{k=1}^{n} \mathbb{E}[\zeta_k^2].$$

We are now in a position to prove the main results.

Proof of Theorem 3.3.2. First, note that (a) is straighforward since $x \leq x_+$ for any $x \in \mathbb{R}$. Les us prove (b). The key result is the following Fuk-Nagaev inequality for martingales obtained by Courbot:

Theorem 3.3.11 ([8], Theorem 1)**.** *Let $M_n := \sum_{k=1}^{n} X_k$ be a martingale in $\mathbb{L}^2$ with respect to a nondecreasing filtration $(\mathcal{F}_k)$, such that $M_0 = 0$ and $\|\mathbb{E}[X_k^2 \mid \mathcal{F}_{k-1}]\|_\infty < \infty$. Define*

$$\langle M \rangle_n := \sum_{k=1}^{n} \mathbb{E}[X_k^2 \mid \mathcal{F}_{k-1}].$$

Then, for any $x, s, v > 0$,

$$\mathbb{P}(M_{n+} \geq x) \leq \sum_{k=1}^{n} \mathbb{P}(sX_{k+} > x) + \mathbb{P}(\langle M \rangle_n > v) + \exp\Big(- \frac{s^2 v}{x^2} h\Big(\frac{x^2}{sv}\Big)\Big),$$

where $h(x) = (1 + x)\log(1 + x) - x$.

We apply the above result to Ξ_n (respectively R_n) with $v = n\sigma^2$ (respectively $v = V_n$). Then, since $h(x) \geq x \log(1+x)/2$, we have for any $x > 0$ and any $s > 0$,

$$\mathbb{P}(\Xi_{n+} \geq x) \leq \sum_{k=1}^{n} \mathbb{P}(s\xi_{k+} > x) + \Big(1 + \frac{x^2}{sn\sigma^2}\Big)^{-s/2}, \tag{3.3.17}$$

$$\mathbb{P}(R_{n+} \geq x) \leq \sum_{k=1}^{n} \mathbb{P}(s(r_k - \mathbb{E}_{k-1}[r_k])_+ > x) + \Big(1 + \frac{x^2}{sV_n}\Big)^{-s/2}. \tag{3.3.18}$$

Furthermore, since $\xi_k \leq \Phi(X_k)$, $0 \leq r_k \leq 2\,\Phi(X_k)$ and the random variables X_k are i.i.d.,

$$\mathbb{P}(\xi_{k+} > x/s) \leq \mathbb{P}(\Phi(X_1) > x/s) \leq \mathbb{P}(\Phi(X_1) > x/2s), \tag{3.3.19}$$

and

$$\mathbb{P}((r_k - \mathbb{E}_{k-1}[r_k])_+ > x/s) \leq \mathbb{P}(\Phi(X_1) > x/2s). \tag{3.3.20}$$

Moreover, we derive from (3.3.14) and the sub-additivity of $x \mapsto x_+$, that for any $x > 0$,

$$\mathbb{P}((Z - \mathbb{E}[Z])_+ \geq x) \leq \inf_{t \in [0,1]} \Big\{\mathbb{P}(\Xi_{n+} \geq tx) + \mathbb{P}(R_{n+} \geq (1-t)x)\Big\}. \tag{3.3.21}$$

Since the optimization in t in the right-hand side of above inequality is quite complicated to calculate, we take $t = 1/2$ in the sequel. Now, combining (3.3.21) with (3.3.17)–(3.3.20) leads to inequality (b) of Theorem 3.3.2 which ends the proof. □

Proof of Theorem 3.3.3. First observe that (a) is the point (ii) of Proposition 3.2.2 and that (c) follows immediately from (b) by the point (i) of the same Proposition 3.2.2. Let us now prove (b). Recalling the decomposition (3.3.14), the point (iii) of Proposition 3.2.2 implies

$$\tilde{Q}_{Z-\mathbb{E}[Z]}(u) \leq \tilde{Q}_{\Xi_n}(u) + \tilde{Q}_{R_n}(u). \tag{3.3.22}$$

Next, to control the summands in the right-hand side, the key result is the following new Fuk-Nagaev inequality obtained by Rio:

Theorem 3.3.12 ([19], Theorem 4.1)**.** *Let $M_n := \sum_{k=1}^n X_k$ be a martingale in $\mathbb{L}^2$ with respect to a nondecreasing filtration $(\mathcal{F}_k)$, such that $M_0 = 0$ and for some constant $r > 2$,*

$$\|\mathbb{E}[X_k^2 \mid \mathcal{F}_{k-1}]\|_\infty < \infty \quad \textit{and} \quad \Big\| \sup_{t>0} \Big(t^r\, \mathbb{P}(X_{k+} > t \mid \mathcal{F}_{k-1}) \Big) \Big\|_\infty < \infty.$$

Define

$$\sigma = \Big\| \sum_{k=1}^n \mathbb{E}[X_k^2 \mid \mathcal{F}_{k-1}] \Big\|_\infty^{1/2} \textit{ and } C_r^w(M) = \Big\| \sup_{t>0} \Big(t^r \sum_{k=1}^n \mathbb{P}(X_{k+} > t \mid \mathcal{F}_{k-1}) \Big) \Big\|_\infty^{1/r}.$$

Then for any $u \in]0,1[$,

$$\tilde{Q}_{M_n}(u) \le \sigma \sqrt{2\log(1/u)} + C_r^w(M)\mu_r u^{-1/r},$$

where $\mu_r := 2 + \max(4/3, r/3)$.

As in the proof of Theorem 3.3.2, we bound up ξ_k and $r_k - \mathbb{E}_{k-1}[r_k]$ respectively by $\Phi(X_k)$ and $2\,\Phi(X_k)$ to get

$$C_\ell^w\Big(\Xi_n\Big) + C_\ell^w\Big(R_n\Big) \le 3\, n^{1/\ell} \Lambda_\ell^+(\Phi(X_1)). \tag{3.3.23}$$

Recalling the bounds of the quadratic variations of the two martingales that we found previously, we then conclude the proof by combining (3.3.22), Theorem 3.3.12 and (3.3.23). □

Proof of Corollary 3.3.4. First, we have (see Petrov [16], p.61–62 and Exercice 2.26) that for any $\ell \ge 1$,

$$\mathbb{E}[(Z - \mathbb{E}[Z])_+^\ell] = \ell \int_0^\infty \mathbb{P}((Z - \mathbb{E}[Z])_+ \ge x) x^{\ell-1} dx. \tag{3.3.24}$$

Hence, using Theorem 3.3.2, we get

$$\begin{aligned} \mathbb{E}[(Z - \mathbb{E}[Z])_+^\ell] \le 2^\ell \frac{\ell}{2} s^{\ell/2} B\Big(\frac{s-\ell}{2}, \frac{\ell}{2}\Big) \Big((n\,\sigma^2)^{\ell/2} + (V_n)^{\ell/2} \Big) \\ + 2\, n\ell \int_0^\infty x^{\ell-1}\, \mathbb{P}(\Phi(X_1) > x/4s) dx, \end{aligned} \tag{3.3.25}$$

where $B(x,y) = \int_0^1 t^{x-1}(1-t)^{y-1} dt$ is the usual Beta function. See now that for $\ell \ge 2$ and $s := \ell + 1$, we have

$$B\Big(\frac{s-\ell}{2}, \frac{\ell}{2}\Big) = \sqrt{\pi} \frac{\Gamma(\ell/2)}{\Gamma((\ell+1)/2)}.$$

Finally, we conclude the proof by the change of variables $x/4s = y$ in the integral term in (3.3.25) and the subadditivity of the function $x \mapsto x^{1/\ell}$. □

Proof of Corollary 3.3.7. First, observe that (a) follows directly from (3.2.6). Let us now prove (b). We proceed exactly as in Rio [19, Theorem 5.1]. Both (3.2.4) and Theorem 3.3.3 (b) imply

$$\tilde{\Lambda}_\ell^+(Z-\mathbb{E}[Z]) \le (\sigma\sqrt{n}+\sqrt{V_n}\,) \sup_{u\in]0,1]} \Big(u^{1/\ell}\sqrt{2\log(1/u)}\Big) + 3\,n^{1/\ell}\mu_\ell\,\Lambda_\ell^+(\Phi(X_1)). \tag{3.3.26}$$

Next, observe that $u^{1/\ell}\sqrt{2\log(1/u)} \le \sqrt{(\ell/e)}$, which concludes the proof. □

3.4 Application to power-type tail

Let $Y_1,\ldots,Y_n$ be a finite sequence of nonnegative, independent and identically distributed random variables and $X_1,\ldots,X_n$ a finite sequence of independent and identically distributed random variables with values in some measurable space $(\mathcal{X},\mathcal{F})$ such that the two sequences are independent. Let P denote the common distribution of the X_k. Let $\mathscr{G}$ be a countable class of measurable functions from $\mathcal{X}$ into $[-1,1]$ such that for all $g\in\mathscr{G}$,

$$P(g)=0 \quad \text{and} \quad P(g^2)<\delta^2 \;\text{ for some } \delta\in]0,1[. \tag{3.4.1}$$

Let G be a measurable envelope function of $\mathscr{G}$ that is

$$|g|\le G \text{ for any } g\in\mathscr{G}, \text{ and } G(x)\le 1 \text{ for all } x\in\mathcal{X}. \tag{3.4.2}$$

We suppose furthermore that for some constant $p>2$,

$$\mathbb{P}(Y_1>t)\le t^{-p} \qquad \text{for any } t>0. \tag{3.4.3}$$

Define now

$$Z := \sup_{g\in\mathscr{G}} \sum_{k=1}^n Y_k\, g(X_k). \tag{3.4.4}$$

Setting $\tilde{X}_k := (X_k,Y_k)$ and $\mathscr{F}$ the class of functions from $\mathcal{X}\times\mathbb{R}_+$ into $\mathbb{R}$ which verified that for any $f\in\mathscr{F}$ there exists a unique $g\in\mathscr{G}$ such that $f(x,y)=yg(x)$, we then have $Z=\sup_{f\in\mathscr{F}}\sum_{k=1}^n f(\tilde{X}_k)$. Hence, this allows us to apply results of the previous section. The envelope function of $\mathscr{F}$ is defined by $F(x,y):=y\,G(x)$. Moreover we can obtain a more precise bound for V_n. Indeed, we will use an upper bound for the mean of suprema of empirical processes, expressed in terms of the uniform entropy integral, proved by van der Vaart and Wellner [22]. Let us first recall some classical definitions.

Definition 3.4.1 (Covering number and uniform entropy integral)**.** *The covering number $N(\epsilon, \mathscr{G}, Q)$ is the minimal number of balls of radius ϵ in $\mathbb{L}^2(Q)$ needed to cover the set $\mathscr{G}$. The uniform entropy integral is defined by*

$$J(\delta, \mathscr{G}) := \sup_Q \int_0^\delta \sqrt{1 + \log N\Big(\epsilon \|G\|_{Q,2}, \mathscr{G}, Q\Big)}\, d\epsilon.$$

Here, the supremum is taken over all finitely discrete probability distributions Q on $(\mathcal{X}, \mathcal{F})$ and $\|f\|_{Q,2}$ denotes the norm of a function f in $\mathbb{L}^2(Q)$.

Throughout this section, K denotes an universal constant which may change from line to line.

Theorem 3.4.2. *Let Z be defined by* (3.4.4)*. Under conditions* (3.4.1) – (3.4.3)*, the following results hold :*
(i) If Y_1 is $\mathbb{L}^p$-integrable, then

$$\begin{aligned}
&\|(Z - \mathbb{E}[Z])_+\|_p \\
&\le 2\,\beta_p\, p^{1/p}\sqrt{p+1}\left(\sigma\sqrt{n} + K\sqrt{\frac{p}{p-2}}\left(n^{q/4}\sqrt{J(\delta,\mathscr{G})} + \sqrt{p}\, n^{1/p}\Big(\frac{J(\delta,\mathscr{G})}{\delta}\Big)^{1/q}\right)\right) \\
&\qquad + 2^{2+1/p}\, n^{1/p}(p+1)\|Y_1\|_p\, \|G(X_1)\|_p, \quad (a)
\end{aligned}$$

where $\beta_p = (\sqrt{\pi}/2)\,\Gamma(p/2)\,/\Gamma((p+1)/2)$.
(ii) Moreover,

$$\begin{aligned}
&\tilde{\Lambda}_p^+(Z - \mathbb{E}[Z]) \\
&\quad \le \sqrt{(p/e)}\left(\sigma\sqrt{n} + K\sqrt{\frac{p}{p-2}}\left(n^{q/4}\sqrt{J(\delta,\mathscr{G})} + \sqrt{p}\, n^{1/p}\Big(\frac{J(\delta,\mathscr{G})}{\delta}\Big)^{1/q}\right)\right) \\
&\qquad + 3\, n^{1/p}\mu_p, \quad (b)
\end{aligned}$$

where $q = p/(p-1)$ and $\mu_p = 2 + \max(4/3, p/3)$.

We now compare Inequality (a) above with results in the literature. Set $C_p := 2\,\beta_p\, p^{1/p}\sqrt{p+1}$. Consider first the bounded case : $Y_k \le 1$. Integrating the Rio inequality recalled in Theorem 3.1.1 and bounding up $\mathbb{E}[Z]$ by Proposition 3.4.5 on the present chapter, one obtains

$$\|(Z - \mathbb{E}[Z])_+\|_p \le C_p\left(\sigma\sqrt{n} + K\sqrt{\frac{p}{p-2}}\Big(n^{1/4}\sqrt{J(\delta,\mathscr{G})} + \sqrt{p}\,\frac{J(\delta,\mathscr{G})}{\delta}\Big)\right). \tag{3.4.5}$$

Remark 3.4.3. *Note that when p tends to infinity, q tends to* 1*. This allows us to see Inequality (a) of Theorem 3.4.2 as an extension of* (3.4.5) *to the unbounded case.*

Remark 3.4.4. *In the unbounded case, the result of Boucheron & al. [4] recalled in Theorem 3.1.2.*

$$\|(Z - \mathbb{E}[Z])_+\|_p \leq B_p\, \sigma\sqrt{n} + o(\sqrt{n}), \tag{3.4.6}$$

where $B_p := 2\,(1 - e^{-1/2})^{-1/2}\sqrt{p-1}$. *Remark that the constant* C_p *is always better than the constant* B_p. *For instance, for* $p = 4$, *we have* $B_4 \simeq 5.5225$ *and* $C_4 \simeq 4.2164$. *Furthermore, when* p *tends to infinity,* B_p *is equivalent to* $3.1884\,\sqrt{p}$ *while* C_p *is equivalent to* $\sqrt{p}$.

Proof of Theorem 3.4.2. First, we bound up the term $V_n = \sum_{k=1}^n \mathbb{E}[\zeta_k^2]$. We recall that ζ_k is a random variable with distribution function $F_{2\,Y_1 G(X_1), q_k}$ (defined in (3.2.1)) and q_k is such that $\mathbb{E}[\zeta_k] = E_k$. Let ψ be a random variable with tail function defined by $\mathbb{P}(\psi > t) = t^{-p}$ for all $t \geq 1$ and let $\tilde{\zeta}_k$ be a random variable with distribution function $F_{2\psi, \tilde{q}_k}$ where $\tilde{q}_k$ is the real in $[0,1]$ such that $\mathbb{E}[\tilde{\zeta}_k] = E_k$. Clearly,

$$F_{2\,Y_1 G(X_1), q_k}(x) \geq F_{2\psi, \tilde{q}_k}(x) \text{ for any } x \in \mathbb{R}.$$

Then Lemma 1 of Bentkus [3] ensures that $\mathbb{E}[\varphi(\zeta_k)] \leq \mathbb{E}[\varphi(\tilde{\zeta}_k)]$ for any $\varphi \in \mathcal{H}_+^1$. In particular, this implies $\mathbb{E}[\zeta_k^2] \leq \mathbb{E}[\tilde{\zeta}_k^2]$. Therefore, an elementary calculation yields

$$V_n \leq 2^{p/(p-1)} \frac{p}{p-2} \Big(\frac{p-1}{p}\Big)^{\frac{p-2}{p-1}} \sum_{k=1}^n E_k^{\frac{p-2}{p-1}}. \tag{3.4.7}$$

Next we show how we can obtain a bound for E_k in terms of uniform entropy integral.

Proposition 3.4.5. *There exists a universal constant* K *such that for any integer* $k \geq 1$,

$$\frac{1}{k}\, \mathbb{E} \sup_{g \in \mathscr{G}} \Big| \sum_{j=1}^k Y_j g(X_j) \Big| \leq K \frac{p}{p-2} \Big(k^{-1/2} J(\delta, \mathscr{G}) + p\, k^{(1/p)-1} \Big(\frac{J^2(\delta, \mathscr{G})}{\delta^2} \Big)^{1-1/p} \Big).$$

Proof of Proposition 3.4.5. Notice that $Q_\psi(u) = u^{-1/p}$ for any $u \in]0,1[$ and that (3.4.3) implies $Q_Y \leq Q_\psi$. Let also $\kappa \in \mathbb{R}$ such that

$$2^\kappa = \frac{k\,\delta^2}{J^2(\delta, \mathscr{G})}. \tag{3.4.8}$$

Let $U_1, \ldots, U_k$ be k independent copies of a random variable U distributed uniformly on $[0,1]$. Let us now define for every $j = 1, \ldots, \lceil \kappa \rceil$,

$$I_j := \{m \in \{1, \ldots, k\} : U_m \in]2^{-j}, 2^{1-j}]\},$$
$$J_\kappa := \{m \in \{1, \ldots, k\} : U_m \leq 2^{-\lceil \kappa \rceil}\}.$$

Here, $\lfloor . \rfloor$ and $\lceil . \rceil$ denote the classical floor and ceiling functions. We recall the basic property of the quantile function Q_X of a random variable X : $Q_X(U)$ has the same distribution as X for any random variable U with the uniform distribution over $[0,1]$. Then,

$$\mathbb{E} \sup_{g\in\mathscr{G}} \Big| \sum_{j=1}^{k} Y_j g(X_j) \Big| \leq \mathbf{E}_1 + \mathbf{E}_2, \tag{3.4.9}$$

where

$$\mathbf{E}_1 := \sum_{j=1}^{\lceil\kappa\rceil} \mathbb{E} \sup_{g\in\mathscr{G}} \Big| \sum_{i\in I_j} Q_\psi(U_i) g(X_i) \Big| \quad \text{and} \quad \mathbf{E}_2 := \mathbb{E} \sup_{g\in\mathscr{G}} \Big| \sum_{j\in J_\kappa} Q_\psi(U_j) g(X_j) \Big|.$$

Let us bound up $\mathbf{E}_2$. Since $G \leq 1$, a straightforward calculation gives

$$\mathbf{E}_2 \leq k \int_0^{2^{-\lceil\kappa\rceil}} Q_\psi(u)\, du \leq k\, \frac{p}{p-1}\, 2^{-\kappa(1-1/p)}. \tag{3.4.10}$$

To bound up $\mathbf{E}_1$, we first notice that, since Q_ψ is decreasing, for any $m \in I_j$, $|Y_m g(X_m)| \leq Q_\psi(2^{-j})$. We can then apply Theorem 2.1 of Van der Vaart and Wellner [22] which leads to

$$\mathbf{E}_1 \leq K\Big(J(\delta,\mathscr{G}) \sum_{j=1}^{\lceil\kappa\rceil} \mathbb{E}\Big[|I_j|^{\frac{1}{2}}\Big]\, Q_\psi(2^{-j}) + \frac{J^2(\delta,\mathscr{G})}{\delta^2} \sum_{j=1}^{\lceil\kappa\rceil} Q_\psi(2^{-j}) \Big). \tag{3.4.11}$$

By the definition of I_j, it is easy to see that

$$\mathbb{E}\Big[|I_j|\Big] = \sum_{i=1}^{k} i \binom{k}{i} (2^{-j})^i (1-2^{-j})^{k-i} = k\, 2^{-j}.$$

Then, Jensen's inequality yields $\mathbb{E}[|I_j|^{\frac{1}{2}}] \leq \sqrt{k\, 2^{-j}}$. Since $Q_\psi(u) = u^{-1/p}$, one has

$$\sum_{j=1}^{\lceil\kappa\rceil} 2^{-j/2} Q_\psi(2^{-j}) \leq \frac{2^{1/p-1/2}}{1-2^{1/p-1/2}} \leq \frac{2}{\log(2)}\, \frac{p}{p-2}. \tag{3.4.12}$$

Likewise,

$$\begin{aligned} \sum_{j=1}^{\lceil\kappa\rceil} Q_\psi(2^{-j}) &= 2^{\lfloor\kappa\rfloor/p}\Big(2^{1/p} + \sum_{j=0}^{\lfloor\kappa\rfloor-1} 2^{-j/p}\Big) \\ &\leq 2^{\lfloor\kappa\rfloor/p}\Big(2^{1/p} + \frac{1}{1-2^{-1/p}}\Big) \leq \frac{2^{\kappa/p}}{\log(2)}\, \frac{p^2}{p-2}. \end{aligned} \tag{3.4.13}$$

Hence, we derive from (3.4.11) – (3.4.13),

$$\mathbf{E}_1 \leq K \frac{p}{p-2}\Big(\sqrt{k}\, J(\delta, \mathscr{G}) + p\, \frac{J^2(\delta, \mathscr{G})}{\delta^2}\, 2^{\kappa/p}\Big). \tag{3.4.14}$$

Finally, (3.4.9), (3.4.14), (3.4.10) and the definition of κ imply Proposition 3.4.5. $\square$

Let us continue the proof of Theorem 3.4.2. Using the subadditivity of the functions $x \mapsto x^a$ for $0 < a < 1$, from (3.4.7) and Proposition 3.4.5 we obtain that

$$\sqrt{V_n} \leq K \sqrt{\frac{p}{p-2}}\Big(n^{q/4}\sqrt{J(\delta, \mathscr{G})} + \sqrt{p}\, n^{1/p}\Big(\frac{J(\delta, \mathscr{G})}{\delta}\Big)^{1/q}\Big). \tag{3.4.15}$$

Injecting this bound into the inequality of Corollary 3.3.4 gives (a). Similarly, injecting this bound in Inequality (b) of Corollary 3.3.7, we conclude the proof of (b) since $\Lambda_p^+(Y_1 G(X_1)) \leq \Lambda_p^+(\psi) = 1$. This ends the proof of Theorem 3.4.2. $\square$

References

[1] R. Adamczak. A tail inequality for suprema of unbounded empirical processes with applications to Markov chains. *Electron. J. Probab.*, 13:no. 34, 1000–1034, 2008.

[2] C. Bennett and R. Sharpley. *Interpolation of operators*, volume 129 of *Pure and Applied Mathematics*. Academic Press, Inc., Boston, MA, 1988.

[3] V. Bentkus. An extension of the Hoeffding inequality to unbounded random variables. *Lithuanian Mathematical Journal*, 48(2):137–157, 2008.

[4] S. Boucheron, O. Bousquet, G. Lugosi, and P. Massart. Moment inequalities for functions of independent random variables. *Ann. Probab.*, 33(2):514–560, 03 2005.

[5] S. Boucheron, G. Lugosi, and P. Massart. *Concentration Inequalities: A Nonasymptotic Theory of Independence*. OUP Oxford, 2013.

[6] O. Bousquet. Concentration inequalities for sub-additive functions using the entropy method. In *Stochastic inequalities and applications*, volume 56 of *Progr. Probab.*, pages 213–247. Birkhäuser, Basel, 2003.

[7] F. Comte and C. Lacour. Anisotropic adaptive kernel deconvolution. *Ann. Inst. Henri Poincaré Probab. Stat.*, 49(2):569–609, 2013.

[8] B. Courbot. Rates of convergence in the functional CLT for martingales. *Academie des Sciences Paris Comptes Rendus Serie Sciences Mathematiques*, 328:509–513, March 1999.

[9] U. Einmahl and D. Li. Characterization of LIL behavior in Banach space. *Trans. Amer. Math. Soc.*, 360(12):6677–6693, 2008.

[10] T. Klein and E. Rio. Concentration around the mean for maxima of empirical processes. *Ann. Probab.*, 33(3):1060–1077, 05 2005.

[11] J. Lederer and S. van de Geer. New concentration inequalities for suprema of empirical processes. *Bernoulli*, 20(4):2020–2038, 11 2014.

[12] M. Ledoux. On Talagrand's deviation inequalities for product measures. *ESAIM: Probability and Statistics*, 1:63–87, 1997.

[13] A. Marchina. Concentration inequalities for suprema of unbounded empirical processes. Preprint on <hal-01545101>, 2017.

[14] A. Marchina. Concentration inequalities for separately convex functions. *To appear in Bernoulli Journal*, 2018+. Preprint on <hal-01344861>.

[15] P. Massart. About the constants in Talagrand's concentration inequalities for empirical processes. *Ann. Probab.*, 28(2):863–884, 2000.

[16] V.V. Petrov. *Limit Theorems of Probability Theory: Sequences of Independent Random Variables.* Oxford science publications. Clarendon Press, 1995.

[17] I. Pinelis. An optimal three-way stable and monotonic spectrum of bounds on quantiles: a spectrum of coherent measures of financial risk and economic inequality. *Risks*, 2(3):349–392, 2014.

[18] E. Rio. Une inégalité de Bennett pour les maxima de processus empiriques. *Ann. Inst. H. Poincaré Probab. Statist.*, 38(6):1053–1057, 2002. En l'honneur de J. Bretagnolle, D. Dacunha-Castelle, I. Ibragimov.

[19] E. Rio. About the constants in the Fuk-Nagaev inequalities. *Electron. Commun. Probab.*, 22:12 pp., 2017.

[20] S. van de Geer and J. Lederer. The Bernstein-Orlicz norm and deviation inequalities. *Probab. Theory Related Fields*, 157(1-2):225–250, 2013.

[21] A.W. van der Vaart and J.A. Wellner. *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer Series in Statistics. Springer New York, 1996.

[22] A.W. van der Vaart and J.A. Wellner. A local maximal inequality under uniform entropy. *Electron. J. Statist.*, 5:192–203, 2011.

Chapter 4

About the rate function in the concentration inequalities for suprema of bounded empirical processes

We provide new deviation inequalities in the large deviations bandwidth for suprema of empirical processes indexed by classes of uniformly bounded functions associated with independent and identically distributed random variables. The improvements we get concern the rate function which is, as expected, the Legendre transform of suprema of the log-Laplace transform of the pushforward measure by the functions of the considered class (up to an additional corrective term). Our approach is based on a decomposition in martingale together with some comparison inequalities.

4.1 Introduction

Let $X_1, \dots, X_n$ be a sequence of independent random variables valued in some measurable space $(\mathcal{X}, \mathcal{F})$ with common distribution P. Let P_n denote the empirical probability measure $P_n := n^{-1}(\delta_{X_1} + \ldots + \delta_{X_n})$. Let $\mathscr{F}$ be a countable class of measurable functions $f : \mathcal{X} \to \mathbb{R}$ such that $P(f) = 0$ and $|f(x)| \leq 1$ for all $x \in \mathcal{X}$ and all $f \in \mathscr{F}$. In this Chapter we are interested in exponential deviation inequalities with precise rate functions in the large deviations bandwidth for the random variable

$$Z := \sup\{nP_n(f) : f \in \mathscr{F}\}, \tag{4.1.1}$$

around its mean. First, let us briefly recall known results on concentration of Z around its mean for uniform bounded classes $\mathscr{F}$. Talagrand [22] obtains a Bennett-type inequality by means of isoperimetric inequalities for product measures. Ledoux [12] introduces a new method based on entropic inequalities to recover more directly Talagrand's inequalities. This method, which allows to bound up the Laplace transform of Z, is the starting point of a series of papers, mainly to reach optimal constants in Talagrand's inequalities. Let us cite among others, Massart [14], Rio [17, 18, 19], Bousquet [6], Klein [10], Klein and Rio [11]. In the large deviations bandwidth, as rate function, we expect the Legendre transform of $t \mapsto \sup_{f \in \mathscr{F}} \ell_f(t)$, denoted by $\ell^*_{\mathscr{F}}$, where ℓ_f is the log-Laplace transform $\ell_f(t) := \log P(e^{tf})$ for all $t \geq 0$ and all $f \in \mathscr{F}$. Indeed, one has

$$\frac{1}{n} \log \mathbb{E}[e^{tZ}] \geq \sup_{f \in \mathscr{F}} \ell_f(t) =: \ell_{\mathscr{F}}(t), \tag{4.1.2}$$

which implies

$$\frac{1}{n} \log \mathbb{E}[e^{t(Z - \mathbb{E}[Z])}] \geq \ell_{\mathscr{F}}(t) - t\, \frac{\mathbb{E}[Z]}{n}. \tag{4.1.3}$$

Now, if $\mathbb{E}[Z]/n$ tends to 0 (for example, this condition is satisfied when $\mathscr{F}$ is a Glivenko-Cantelli class), then

$$\liminf_{n \to \infty} \frac{1}{n} \log \mathbb{E}[e^{t(Z - \mathbb{E}[Z])}] \geq \ell_{\mathscr{F}}(t). \tag{4.1.4}$$

This elementary lower bound shows that the large deviations rate function $\ell^*_{\mathscr{F}}$ cannot be improved. To the best of our knowledge, the only result in this direction is obtained in Rio [17] and concerns the particular case of set-indexed empirical processes. Rio get as rate function, for the right-hand side deviations for sets with large measure under P and for the left-hand side deviations, that of a Bernoulli random variable which actually corresponds to $\ell^*_{\mathscr{F}}$. In this Chapter, we obtain as rate function for the general case, the function $\ell^*_{\mathscr{F}}$ with an additional corrective term which tends to 0 as n tends to infinity as soon as $\mathscr{F}$ is a weak Glivenko-Cantelli class (see Remark 4.3.3). Our methods are only based on martingale techniques and comparison inequalities.

The Chapter is organized as follows. First, in Section 4.2 we recall some definitions and preliminary results on the conditional value-at-risk and some comparison inequalities. In Section 4.3 we state the main results of this Chapter. We study the rate function $\ell^*_{\mathscr{F}}$ appearing in the main result in Section 4.4. Finally, we provide detailed proofs in Section 4.5.

4.2 Notations and preliminary results

In this section, we give the notations and definitions which we will use all along the Chapter. Let us start by the definition of the Conditional Value-at-Risk (CVaR for short).

Definition 4.2.1. *Let X be a real-valued integrable random variable. Let the function Q_X be the càdlàg inverse of $x \mapsto \mathbb{P}(X > x)$. The Conditional Value-at-Risk is defined by*

$$\tilde{Q}_X(u) := u^{-1} \int_0^u Q_X(s) ds \quad \text{for any } u \in]0,1]. \tag{4.2.1}$$

Let us now recall the definition of the Legendre transform of a convex function.

Definition 4.2.2. *Let $\phi : [0, \infty[\to [0, \infty]$ be a convex, nondecreasing and càdlàg function such that $\phi(0) = 0$. The Legendre transform ϕ^* of the function ϕ is defined by*

$$\phi^*(\lambda) := \sup\{\lambda t - \phi(t) : t > 0\} \quad \text{for any } \lambda \geq 0. \tag{4.2.2}$$

The inverse function of ϕ^* admits the following variational expression (see, for instance, Rio [20, Lemma A.2]).

$$\phi^{*-1}(x) = \inf\{t^{-1}(\phi(t) + x) : t > 0\} \quad \text{for any } x \geq 0. \tag{4.2.3}$$

A particular function ϕ satisfying conditions in Definition 4.2.2 is the log-Laplace transform of a random variable:

Notation 4.2.3. *Let X be a real-valued integrable random variable with a finite Laplace transform on right neighborhood of 0. The log-Laplace transform of X, denoted by ℓ_X, is defined by*

$$\ell_X(t) := \log \mathbb{E}[\exp(tX)] \quad \text{for any } t \geq 0. \tag{4.2.4}$$

The function Q_X and the CVaR satisfy the following elementary properties, which are given and proved in Pinelis [15, Theorem 3.4].

Proposition 4.2.4. *Let X and Y be real-valued and integrable random variables. Then, for any $u \in]0,1]$,*

(i) $\mathbb{P}(X > Q_X(u)) \leq u$,

(ii) $Q_X(u) \leq \tilde{Q}_X(u)$,

(iii) $\tilde{Q}_{X+Y}(u) \leq \tilde{Q}_X(u) + \tilde{Q}_Y(u)$.

(iv) Assume that X has a finite Laplace transform on a right neighborhood of 0. Then $\tilde{Q}_X(u) \leq \ell_X^{*-1}(\log(1/u))$.

Remark 4.2.5. *Since we use different notations from those of Pinelis, let us mention that his notations $Q_0(X;u)$, $Q_1(X;u)$ and $Q_\infty(X;u)$ correspond respectively to $Q_X(u)$, $\tilde{Q}_X(u)$ and $\ell_X^{*-1}(\log(1/u))$.*

We now recall some comparison inequalities which will be used in the proof of the main result. Let us first give a notation for a family of distribution probability.

Notation 4.2.6. *Let α, β be two reals such that $\alpha < \beta$. We say that a random variable θ follows a Bernoulli distribution if it assumes exactly two values and we write $\theta \sim \mathcal{B}_m(\alpha, \beta)$ if*

$$\mathbb{P}(\theta = \beta) = 1 - \mathbb{P}(\theta = \alpha) \in]0,1[, \text{ and } \mathbb{E}[\theta] = m. \tag{4.2.5}$$

Notice that

$$\operatorname{Var}(\theta) = (m - \alpha)(\beta - m). \tag{4.2.6}$$

The following classical convex comparison inequality between a bounded random variable X and a Bernoulli random variable with values the bounds of X was first proved by Hoeffding (see Inequalities (4.1) and (4.2) in [9]); it straight follows by the property of convexity.

Proposition 4.2.7. *Let a, b two positive reals and let X be a bounded random variable $-a \leq X \leq b$ with mean m. Let $\theta \sim \mathcal{B}_m(-a, b)$ (defined by (4.2.5)). Then, for any convex function $\varphi : \mathbb{R} \to \mathbb{R}$,*

$$\mathbb{E}[\varphi(X)] \leq \mathbb{E}[\varphi(\theta)].$$

In particular, since $\mathbb{E}[\theta] = m$, $\operatorname{Var}(X) \leq \operatorname{Var}(\theta) = (a+m)(b-m)$.

Next, Bentkus (see Lemmas 4.4 and 4.5 in [3]) proved that a martingale with bounded from above increments is more concentrate with respect to a certain class of convex functions than a sum of independent and identically distributed Bernoulli random variables.

Proposition 4.2.8. *Let $b, s_1^2, \ldots, s_n^2$ be positive reals. Let $M_n := \sum_{k=1}^n X_k$ be a martingale with respect to a nondecreasing filtration $(\mathcal{F}_k)$ such that $M_0 = 0$,*

$$X_k \leq b, \text{ and } \mathbb{E}[X_k^2 \mid \mathcal{F}_{k-1}] \leq s_k^2 \quad a.s. \tag{4.2.7}$$

Let $s^2 = n^{-1}(s_1^2 + \ldots + s_n^2)$ and $S_n := \vartheta_1 + \ldots + \vartheta_n$ be a sum of n independent copies of a random variable ϑ with distribution $\mathcal{B}_0(-s^2/b, b)$ (defined by (4.2.5)). Then, for any convex nondecreasing function $\varphi : \mathbb{R} \to \mathbb{R}$, differentiable and with convex derivative,

$$\mathbb{E}[\varphi(M_n)] \leq \mathbb{E}[\varphi(S_n)].$$

Remark 4.2.9. *Actually Bentkus obtains the above inequality in a smaller class of functions. This generalization is due to Pinelis (see Corollary 5.8 in [16]).*

4.3 Main results

Let us first introduce one more notation. We denote for any $k = 1, \dots, n$ the expectations

$$E_k := \mathbb{E} \sup_{f \in \mathscr{F}} P_k(f). \tag{4.3.1}$$

The main result of the Chapter is the following theorem:

Theorem 4.3.1. *Let $\mathscr{F}$ be a countable class of measurable functions from $\mathcal{X}$ into $[-1,1]$ such that $P(f) = 0$ for all $f \in \mathscr{F}$. Let Z be defined by (4.1.1). For any $f \in \mathscr{F}$, let ℓ_f and $\ell_{\mathscr{F}}$ be the functions defined by*

$$\ell_f(t) := \log P(e^{tf}) \text{ and } \ell_{\mathscr{F}}(t) := \sup_{f \in \mathscr{F}} \ell_f(t) \quad \text{for any } t \geq 0. \tag{4.3.2}$$

Set $\bar{E} := n^{-1}(E_1 + \ldots + E_n)$ and define

$$v_n := \frac{\bar{E}}{2}\Big(1 - \frac{\bar{E}}{2}\Big). \tag{4.3.3}$$

Let $\theta^{(n)}$ be a Bernoulli random variable with distribution $\mathcal{B}_0(-v_n, 1)$ (defined by (4.2.5)). We denote by ℓ_{v_n} the log-Laplace transform of $\theta^{(n)}$ (as defined by (4.2.4)). Then, for any $x \geq 0$,

$$n^{-1}\tilde{Q}_{Z-\mathbb{E}[Z]}(e^{-nx}) \leq \ell_{\mathscr{F}}^{*-1}(x) + 2\,\ell_{v_n}^{*-1}(x). \tag{a}$$

Consequently, for any $x \geq 0$,

$$\mathbb{P}(Z - \mathbb{E}[Z] > n(\ell_{\mathscr{F}}^{*-1}(x) + 2\,\ell_{v_n}^{*-1}(x))) \leq e^{-nx}. \tag{b}$$

The inverse function of $\ell_{v_n}^*$ cannot be explicitly computed. For this reason we provide below a more tractable bound.

Corollary 4.3.2. *Let ψ be the function defined by*

$$\psi(x) := \frac{\sqrt{2x} + 4x/3}{\log(1 + x/3 + \sqrt{2x})} - 1 \quad \text{for any } x \geq 0. \tag{4.3.4}$$

Then

$$\ell_{v_n}^{*-1}(x) \leq v_n \psi\Big(\frac{x}{v_n}\Big) \quad \text{for any } x \geq 0. \tag{a}$$

Consequently, for any $x \geq 0$,

$$\mathbb{P}\Big(Z - \mathbb{E}[Z] > n\Big(\ell_{\mathscr{F}}^{*-1}(x) + 2\,v_n\psi\Big(\frac{x}{v_n}\Big)\Big)\Big) \leq e^{-nx}. \tag{b}$$

Remark 4.3.3. *If the class $\mathscr{F}$ is a weak Glivenko-Cantelli class, that is $\sup_{f\in\mathscr{F}} |P_n(f)|$ converges to 0 in probability, then E_n decreases to 0 (see, for instance, Section 2.4 of van der Vaart and Wellner [23]) and so v_n also decreases to* 0 *(we recall that $\bar{E} \leq 1$). Consequently, we assert by the variational formula* (4.2.3)*, that*

$$\lim_{n\to\infty} \ell_{v_n}^{*-1}(x) = 0 \quad \text{for all } x \geq 0. \tag{4.3.5}$$

Therefrom it allows us to see that $2\,\ell_{v_n}^{*-1}(x)$ is just a correctional term. Moreover, note that $\psi(x)/x$ tends to 0 *as x tends to infinity and thus,*

$$\lim_{n\to\infty} v_n \psi\Big(\frac{x}{v_n}\Big) = 0 \quad \text{for all } x \geq 0. \tag{4.3.6}$$

Hence, the bound given in Corollary 4.3.2 is still relevant in the large deviations bandwidth.
Consider now the classical bound $\ell_{v_n}(t) \leq v_n t^2/(2-2t/3)$ for any $t \in [0,3[$, which follows from the domination by a centered Poisson distribution. This leads to

$$\ell_{v_n}^{*-1}(x) \leq \sqrt{2v_n x} + \frac{x}{3}. \tag{4.3.7}$$

Note that the right-hand side does not tend to 0 contrary to the other bounds which makes (4.3.7) *non efficient in the large deviations bandwidth.*

Remark 4.3.4 (On the large deviations on the left)**.** *Assume that $\mathscr{F}$ is a Glivenko-Cantelli class and that the identically zero function belongs to $\mathscr{F}$. Then $\mathbb{E}[Z]$ is small with respect to n and $Z \geq 0$. Thus for any $x > 0$, $\mathbb{P}(Z - \mathbb{E}[Z] \leq -nx) = 0$ for n large enough.*

Remark 4.3.5 (Explicit bound for v_n)**.** *In view of* (4.3.3)*, since the function $x \mapsto x(1-x)$ is increasing between 0 and 1/2, in order to provide a more explicit bound for v_n, we only have to provide a bound for $\bar{E}$ (which is lower than* 1 *and tends to 0 as n tends to infinity). To this end, we shall use the recent results of Baraud [1] who provides (see his Theorems 2.1 and 2.2) upper bounds with explicit constants for the expectations of suprema of empirical processes, under the hypothesis that $\mathscr{F}$ is a weak VC-major class.*

Assume then that $\mathscr{F}$ is a weak VC-major class with dimension d. Let $\sigma^2 := \sup_{f\in\mathscr{F}} P(f^2)$ denote the wimpy variance. Then Inequality (2.8) in [1] implies the following proposition (the proof is postponed to Section 4.5).

Proposition 4.3.6. *Assume that $n \geq d$. Then*

$$\bar{E} \leq 2\sqrt{2}\,\sigma\log(e/\sigma)n^{-1/2}\Big(\sqrt{C_1(d)} + \sqrt{C_2(n,d)}\,\Big) + 8\,n^{-1}\Big(C_1(d) + C_2(n,d)\Big), \tag{a}$$

where

$$C_1(d) := \log(2) \sum_{k=1}^{d} (1 + 1/k) \quad \textit{and}$$
$$C_2(n, d) := (d/2) \log\left(\frac{n + 1/2}{d + 1/2}\right) \log\left(4 \frac{e^2}{d^2} (n + 1/2)(d + 1/2)\right).$$

As n tends to infinity, the right-hand side of (a) admits the following behavior

$$2\sqrt{2}\, \sigma \log(e/\sigma) n^{-1/2} \log(n) + 4\, d\, n^{-1} \log^2(n). \tag{b}$$

We end this section by giving a simple example where the function $\ell_{\mathscr{F}}$ is explicit.

Example 4.3.7. *Let $\mathscr{S}$ be a countable class of sets. Let $\varepsilon_1, \ldots, \varepsilon_n$ be a sequence of independent Rademacher random variables and independent of $X_1, \ldots, X_n$. Define*

$$Z := \sup_{S \in \mathscr{S}} \sum_{k=1}^{n} \varepsilon_k \mathbb{1}_S(X_k). \tag{4.3.8}$$

For any $S \in \mathscr{S}$ and any $k = 1, \ldots, n$, we get by a straightforward calculation

$$\ell_S(t) := \log \mathbb{E}[e^{t\varepsilon_k \mathbb{1}_S(X_k)}] = \log(1 + P(S)(\cosh(t) - 1)) \quad \textit{for any } t \geq 0. \tag{4.3.9}$$

Clearly the right-hand side is increasing with respect to $P(S)$. Then

$$\ell_{\mathscr{S}}(t) := \sup_{S \in \mathscr{S}} \ell_S(t) = \log(1 + p(\cosh(t) - 1)) \quad \textit{for any } t \geq 0, \tag{4.3.10}$$

where $p := \sup\{P(S) : S \in \mathscr{S}\}$. By (4.2.3), $\ell^{*-1}_{\mathscr{S}}$ is then given by the variational formula

$$\ell^{*-1}_{\mathscr{S}}(x) = \inf_{t>0} \Big\{ t^{-1}\Big(x + \log(1 + p(\cosh(t) - 1))\Big)\Big\} \quad \textit{for any } x \geq 0. \tag{4.3.11}$$

We also refer the reader to Bennett [2], p. 532, for an explicit formula for $\ell^*_{\mathscr{S}}$.

4.4 About the rate function $\ell^*_{\mathscr{F}}$

4.4.1 Comments on Large Deviation Principle

In this subsection we explain how the rate function $\ell^*_{\mathscr{F}}$ arises in the large deviations theory for suprema of bounded empirical processes.

Throughout this section, we assume that for all $f \in \mathscr{F}$, $0 \leq f \leq 1$. We denote by $l_\infty(\mathscr{F})$ the space of all bounded real functions on $\mathscr{F}$ equipped with the norm $\|F\|_{\mathscr{F}} := \sup_{f\in\mathscr{F}} |F(f)|$, making $(l_\infty(\mathscr{F}), \|.\|_\infty)$ a Banach space. For each finite measure ν on $(\mathcal{X}, \mathcal{F})$ corresponds to an element $\nu^{\mathscr{F}} \in l_\infty(\mathscr{F})$ defined by $\nu^{\mathscr{F}}(f) := \nu(f) = \int f d\nu$ for any $f \in \mathscr{F}$. With a slight abuse of notation, we will keep the notation ν instead of $\nu^{\mathscr{F}}$. Wu [24] gives necessary and sufficient conditions with respect to $\mathscr{F}$ which ensure that P_n satisfies the Large Deviation Principle (LDP for short) in $l_\infty(\mathscr{F})$. We refer the reader to the paper of Wu for these conditions (for example, if $\mathscr{F}$ is a Donsker class then the required conditions are satisfied). The (good) rate function is given by

$$h_{\mathscr{F}}(F) := \inf\{H(\nu \mid P) : \nu \text{ is a probability and } \nu = F \text{ on } \mathscr{F}\}, \tag{4.4.1}$$

where $H(\nu \mid P)$ is the relative entropy of ν with respect to P. From there, a direct application of the contraction principle (see, for instance, Theorem 4.2.1 in Dembo and Zeitouni [8]) ensures that $\|P_n\|_{\mathscr{F}}$ satisfies the LDP with rate function given by

$$J(y) := \inf\{H(\nu \mid P) : \nu \text{ is a probability and } \|\nu\|_{\mathscr{F}} = y\} \text{ for any } y \in [0,1]. \tag{4.4.2}$$

We prove the following lemma which gives a better understanding of the rate function J:

Lemma 4.4.1. *$J(y) = \inf_{f\in\mathscr{F}} \ell_f^*(y)$, where $\ell_f(t) := \log P(e^{tf})$ for any $t \geq 0$.*

The important remark is that if we can invert the infimum and the supremum in $\inf_{f\in\mathscr{F}} \sup_{y>0}\{ty - \ell_f(y)\}$, we get that $\inf_{f\in\mathscr{F}} \ell_f^*(y) = \ell_{\mathscr{F}}^*(y)$. It seems not possible to invert the infimum and the supremum in general. However, note that we always have the inequality $\inf_{f\in\mathscr{F}} \ell_f^*(y) \geq \ell_{\mathscr{F}}^*(y)$. In the following proposition, we describe a particular case in which the inversion is valid, which then simplifies the calculation of $\ell_{\mathscr{F}}^*$. Since it directly follows from a minimax theorem (see, for instance, Corollary 3.3 in Sion [21]), we omit the proof.

Proposition 4.4.2. *Let X be a random variable valued in $(\mathcal{X}, \mathcal{F})$ with distribution P. Let $\mathscr{F}$ be a countable class of measurable functions from $\mathcal{X}$ into $[-1,1]$ such that $P(f) = 0$ for all $f \in \mathscr{F}$. Let Θ be a convex compact subset of a vector space. Let $\{\mu_\theta : \theta \in \Theta\}$ be a family of probability distribution on $[-1,1]$ such that, for any $t \geq 0$, $\theta \mapsto \ell_{\mu_\theta}(t) := \log \int e^{tz} \mu_\theta(dz)$ is concave and upper semi-continuous. We assume that for all $f \in \mathscr{F}$, there exists $\theta \in \Theta$ such that $f(X)$ has the distribution μ_θ. Then,*

$$\ell_{\mathscr{F}}^*(x) \geq \inf_{\theta\in\Theta} \ell_{\mu_\theta}^*(x) \quad \textit{for any } x \geq 0.$$

Example 4.4.3 (Set-indexed empirical processes)**.** *Let X be a random variable valued in $(\mathcal{X}, \mathcal{F})$ with distribution P and let $\mathscr{S}$ be a countable class of measurable sets of $\mathcal{X}$. We consider the class $\mathscr{F} := \{\mathbb{1}_S - P(S) : S \in \mathscr{S}\}$. Define $p := \sup\{P(S) : S \in \mathscr{S}\}$ and assume that $p < 1/2$. For any $\theta \in [0, p]$, let us define the function $\ell_\theta(t) := \log(1 + \theta(e^t - 1)) - \theta t$ for any $t \geq 0$. Then Proposition 4.4.2 yields*

$$\ell^*_{\mathscr{F}}(x) \geq \inf_{\theta \in [0,p]} \ell^*_\theta(x) \quad \textit{for any } x \geq 0. \tag{4.4.3}$$

The computation of the right-hand side of (4.4.3) *is perfomed by Rio (see page 175 in [17]): for any $x \leq 1 - 2p$,*

$$\inf_{\theta \in [0,p]} \ell^*_\theta(x) = \ell^*_p(x) = (p + x)\log(1 + x/p) + (1 - p - x)\log(1 - x/(1 - p)). \tag{4.4.4}$$

Furthermore, for any $x \geq 1 - 2p$,

$$\begin{aligned}\inf_{\theta \in [0,p]} \ell^*_\theta(x) &\geq \ell^*_p(1 - 2p) + \int_{1-2p}^x (\ell^*_\theta)'(y)dy \\ &= 2(1 + x)\log\ (1 + x) + 2(1 - x)\log(1 - x) \qquad (4.4.5) \\ &\qquad - (1 + 2p)\log(2p) - (3 - 2p)\log(2 - 2p).\end{aligned}$$

Remark 4.4.4. *Proceeding as in the proof of Theorem 4.2 in Rio [17], one can derive from Theorem 6.3 in Bousquet [6] that, if $p_0 := p + E_n$ satisfies $p_0 < 1/2$, then for any $t > 0$ such that $p_0 < (te^t - e^t + 1)(e^t - 1)^{-2}$,*

$$n^{-1}\log \mathbb{E}[\exp(t(Z - \mathbb{E}[Z]))] \leq \log(1 + p_0(e^t - 1)) - tp_0. \tag{4.4.6}$$

From (4.4.6) *and the usual Cramér-Chernoff calculation, one can derive that $\mathbb{P}(Z - \mathbb{E}[Z] \geq nx) \leq \exp(-n\ell^*_{p_0}(x))$, for any $x > 0$ such that*

$$x \leq (x + p_0)(1 - x - p_0)\log\Big(\frac{(1 - p_0)}{p_0}\frac{(t + p_0)}{(1 - t - p_0)}\Big). \tag{4.4.7}$$

Bousquet [6] tells without proof that (4.4.7) *holds for any $x \leq (3/4)(1-2p_0)$. If $x = x_0 := 1 - 2p_0$,* (4.4.7) *is equivalent to*

$$p_0(1 - p_0) \geq (1 - 2p_0)/2\log(1/p_0 - 1),$$

which is wrong (see Hoeffding [9] p. 19). Recall now that Bousquet's results are derived from the entropy method introduced by Ledoux [12] on the context of concentration inequalities. It appears here that this method does not provide the exact rate function for large values of x, including $x = 1 - 2p_0$.

4.4.2 The case of nondecreasing 1-Lipschitz functions

Here we study the special case of $\mathscr{F}$ included in the set of nondecreasing 1-Lipschitz functions. We can then bound up $\ell_{\mathscr{F}}^{*-1}$ by a more tractable quantity.

Corollary 4.4.5. *Let X be a random variable valued in $(\mathcal{X}, \mathcal{F})$ with distribution P and $X_1, \ldots, X_n$ be n independent copies of X. Let $\mathscr{F}$ be a countable class of measurable functions from $\mathcal{X}$ into $[-1,1]$, nondecreasing, 1-Lipschitz and such that $P(f) = 0$ for all $f \in \mathscr{F}$. Let Z be defined by* (4.1.1)*. Moreover, we assume that the distribution P satisfies that for any $t \in \mathbb{R}$, $\int e^{tx} P(dx) < \infty$. Then*

$$\ell_{\mathscr{F}}^{*-1}(x) \leq \ell_{X-\mathbb{E}[X]}^{*-1}(x) \quad \textit{for any } x \geq 0. \tag{a}$$

Consequently, for any $x \geq 0$,

$$\mathbb{P}\Big(Z - \mathbb{E}[Z] > n(\ell_{X-\mathbb{E}[X]}^{*-1}(x) + 2\,\ell_{v_n}^{*-1}(x))\Big) \leq e^{-nx}. \tag{b}$$

Example 4.4.6. *Let P be the uniform distribution on $[-1,1]$. Then, by* (4.2.3)*, ℓ_X^{*-1} is given by the following variational formula whose values in every point is computable:*

$$\ell_X^{*-1}(x) = \inf_{t>0} \left\{ \frac{1}{t}\left(x + \log\left(\frac{\sinh(t)}{t} \right) \right) \right\} \quad \textit{for any } x \geq 0. \tag{4.4.8}$$

Let us also provide a bound of ℓ_X^{*-1} which is relevant for large values of x. Since $\sinh(t) \leq e^t/2$ *for any $t > 0$, one has*

$$\ell_X^{*-1}(x) \leq 1 + \inf_{t>0} \left\{ \frac{1}{t}(x - \log(2t)) \right\} \quad \textit{for any } x \geq 0. \tag{4.4.9}$$

Then, for each $x \geq 0$, the infimum in (4.4.9) *is reached at $t_x := e^{x+1}/2$, which leads to*

$$\ell_X^{*-1}(x) \leq 1 - \frac{2}{e}\, e^{-x} \quad \textit{for any } x \geq 0. \tag{4.4.10}$$

Furthermore, one can prove that $\ell_X^{*-1}(x)$ is equivalent to $1 - \frac{2}{e}\, e^{-x}$ as x tends to infinity.

4.5 Proofs

4.5.1 Proofs of Section 4.3

Proof of Theorem 4.3.1. First, notice that (b) follows immediately from (a) by Proposition 4.2.4 (i). Let us now prove (a). Our method is based on a martingale decomposition of Z which we now recall. We suppose that $\mathscr{F}$ is a finite class of functions, that is $\mathscr{F} = \{f_i : i \in \{1, \ldots, m\}\}$. The results in the countable case are derived from the finite case using the monotone convergence theorem. Set $\mathcal{F}_0 := \{\varnothing, \Omega\}$ and for all $k = 1, \ldots, n$, $\mathcal{F}_k := \sigma(X_1, \ldots, X_k)$ and $\mathcal{F}_n^k := \sigma(X_1, \ldots, X_{k-1}, X_{k+1}, \ldots, X_n)$. Let $\mathbb{E}_k$ (respectively $\mathbb{E}_n^k$) denote the conditional expectation operator associated with $\mathcal{F}_k$ (resp. $\mathcal{F}_n^k$). Set also

$$Z_k := \mathbb{E}_k[Z], \tag{4.5.1}$$

$$Z^{(k)} := \sup\{nP_n(f) - f(X_k) : f \in \mathscr{F}\}. \tag{4.5.2}$$

The sequence (Z_k) is an $(\mathcal{F}_k)$-adapted martingale and

$$Z - \mathbb{E}[Z] = \sum_{k=1}^{n} \Delta_k, \qquad \text{where } \Delta_k := Z_k - Z_{k-1}. \tag{4.5.3}$$

Define now the random variables τ and τ_k, respectively $\mathcal{F}_n$-measurable and $\mathcal{F}_n^k$-measurable, by

$$\tau := \inf\{i \in \{1, \ldots, m\} : nP_n(f_i) = Z\}, \tag{4.5.4}$$

$$\tau_k := \inf\{i \in \{1, \ldots, m\} : nP_n(f_i) - f_i(X_k) = Z^{(k)}\}. \tag{4.5.5}$$

Notice first that

$$Z^{(k)} + f_{\tau_k}(X_k) \leq Z \leq Z^{(k)} + f_\tau(X_k).$$

From this, conditioning by $\mathcal{F}_k$ gives

$$\mathbb{E}_k[f_{\tau_k}(X_k)] \leq Z_k - \mathbb{E}_k[Z^{(k)}] \leq \mathbb{E}_k[f_\tau(X_k)]. \tag{4.5.6}$$

Set now $\xi_k := \mathbb{E}_k[f_{\tau_k}(X_k)]$ and let $\varepsilon_k \geq r_k \geq 0$ be random variables such that

$$\xi_k + r_k = Z_k - \mathbb{E}_k[Z^{(k)}] \quad \text{and} \quad \xi_k + \varepsilon_k = \mathbb{E}_k[f_\tau(X_k)].$$

Thus (4.5.6) becomes

$$\xi_k \leq \xi_k + r_k \leq \xi_k + \varepsilon_k. \tag{4.5.7}$$

Since the random variable τ_k is $\mathcal{F}_n^k$-measurable, we have by the centering assumption on the elements of $\mathscr{F}$,

$$\mathbb{E}_n^k[f_{\tau_k}(X_k)] = P(f_{\tau_k}) = 0 \quad \text{a.s.}, \tag{4.5.8}$$

which ensures that $\mathbb{E}_{k-1}[\xi_k] = 0$. Moreover, $\mathbb{E}_k[Z^{(k)}]$ is $\mathcal{F}_{k-1}$-measurable. Hence we get

$$\Delta_k = Z_k - \mathbb{E}_k[Z^{(k)}] - \mathbb{E}_{k-1}[Z_k - \mathbb{E}_k[Z^{(k)}]] = \xi_k + r_k - \mathbb{E}_{k-1}[r_k],$$

which, combined with (4.5.3), yields the decomposition of $Z-\mathbb{E}[Z]$ in a sum of two martingales:

$$Z-\mathbb{E}[Z]=\Xi_n+R_n, \tag{4.5.9}$$

where

$$\Xi_n:=\sum_{k=1}^{n}\xi_k \text{ and } R_n:=\sum_{k=1}^{n}(r_k-\mathbb{E}_{k-1}[r_k]). \tag{4.5.10}$$

Now, we bound up separately the log-Laplace transforms of Ξ_n and R_n.

Lemma 4.5.1. *We have*

$$\log\mathbb{E}\Big[\exp\Big(t\,\Xi_n\Big)\Big]\le n\,\ell_{\mathscr{F}}(t)\quad \textit{for any } t\ge 0.$$

Proof of Lemma 4.5.1. The $\mathscr{F}_n^k$-measurability of τ_k gives

$$\mathbb{E}_n^k[\exp(tf_{\tau_k}(X_k))]=P(e^{tf_{\tau_k}}). \tag{4.5.11}$$

This ensures, with an application of the conditional Jensen inequality, that

$$\mathbb{E}_{k-1}[e^{t\xi_k}]=\mathbb{E}_{k-1}\mathbb{E}_n^k[e^{t\xi_k}]\le\mathbb{E}_{k-1}[P(e^{tf_{\tau_k}})]\le\sup_{f\in\mathscr{F}}P(e^{tf}), \tag{4.5.12}$$

almost surely. Then Lemma 4.5.1 follows by an immediate induction on n. □

Lemma 4.5.2. *We have*

$$\log\mathbb{E}\Big[\exp\Big(t\,R_n\Big)\Big]\le n\,\ell_{v_n}(2\,t)\quad \textit{for any } t\ge 0.$$

Proof of Lemma 4.5.2. Actually, the inequality follows by taking $\varphi(x)=e^{tx}$ with $t\ge 0$ in the more general comparison inequality below:

Lemma 4.5.3. *Let $\theta_1^{(n)},\dots,\theta_n^{(n)}$ be a sequence of n independent copies of $\theta^{(n)}$. Then, for any convex nondecreasing function φ from $\mathbb{R}$ into $\mathbb{R}$, differentiable and with convex derivative,*

$$\mathbb{E}[\varphi(R_n)]\le\mathbb{E}\Big[\varphi\Big(2\sum_{k=1}^{n}\theta_k^{(n)}\Big)\Big].$$

Proof of Lemma 4.5.3. We start the proof by showing that

$$r_k-\mathbb{E}_{k-1}[r_k]\le 2,\ \text{ and }\ \mathrm{Var}(r_k\mid\mathcal{F}_{k-1})\le E_{n-k+1}(2-E_{n-k+1})\quad\text{a.s.} \tag{4.5.13}$$

The first inequality above is straightforward by (4.5.7) and the uniform boundedness condition on $\mathscr{F}$. Let us prove now the second inequality. We start by bounding up $\mathrm{Var}(r_k\mid\mathcal{F}_{k-1})$ in terms of $\mathbb{E}_{k-1}[r_k]$. Since $0\le r_k\le 2$, Proposition 4.2.7, applied conditionally to $\mathcal{F}_{k-1}$, yields

$$\mathrm{Var}(r_k\mid\mathcal{F}_{k-1})\le\mathbb{E}_{k-1}[r_k](2-\mathbb{E}_{k-1}[r_k])\quad\text{a.s.} \tag{4.5.14}$$

Next, we prove that $E_{k-1}[r_k]$ is bounded up by a deterministic constant.

Lemma 4.5.4. *We have* $0 \le \mathbb{E}_{k-1}[r_k] \le E_{n-k+1} \le 1$ *a.s.*

Proof of Lemma 4.5.4. The proof is based on the following result on exchangeability of variables, proved in Marchina [13]. Since it is the fundamental tool of the Chapter, we give again the proof for sake of completeness.

Lemma 4.5.5. *For any integer* $j \ge k$, $\mathbb{E}_{k-1}[f_\tau(X_k)] = \mathbb{E}_{k-1}[f_\tau(X_j)]$ *a.s.*

Proof of Lemma 4.5.5. By the definition of the random variable τ, for every permutation on n elements σ, $\tau(X_1, \ldots, X_n) = \tau \circ \sigma(X_1, \ldots, X_n)$ almost surely. Applying now this fact to $\sigma = (k\, j)$ (the transposition which exchanges k and j), it suffices to use Fubini's theorem (recalling that $j \ge k$) to complete the proof. ☐

Then,

$$\begin{aligned}\mathbb{E}_{k-1}[\varepsilon_k] &= \mathbb{E}_{k-1}[f_\tau(X_k)] \\ &= \mathbb{E}_{k-1}[f_\tau(X_k) + \ldots + f_\tau(X_n)]/(n-k+1) \\ &\le \mathbb{E}_{k-1} \sup_{f \in \mathscr{F}} \{f(X_k) + \ldots + f(X_n)\}/(n-k+1) = E_{n-k+1}. \quad (4.5.15)\end{aligned}$$

Recalling that $0 \le r_k \le \varepsilon_k$, we get $0 \le \mathbb{E}_{k-1}[r_k] \le E_{n-k+1}$. The bound $E_{n-k+1} \le 1$ is straightforward by the uniform boundedness condition on the elements of $\mathscr{F}$, which ends the proof of Lemma 4.5.4. ☐

Finally, (4.5.14) together with Lemma 4.5.4 and the fact that $x \mapsto x(2-x)$ is increasing between 0 and 1 imply

$$\mathrm{Var}(r_k \mid \mathcal{F}_{k-1}) \le E_{n-k+1}(2 - E_{n-k+1}) \quad \text{a.s..} \tag{4.5.16}$$

Now Proposition 4.2.8 yields that for any convex, nondecreasing function φ differentiable with convex derivative,

$$\mathbb{E}[\varphi(R_n)] \le \mathbb{E}\Big[\varphi\Big(\sum_{k=1}^{n} \vartheta_k^{(n)}\Big)\Big], \tag{4.5.17}$$

where $\vartheta_1^{(n)}, \ldots, \vartheta_n^{(n)}$ is a sequence of i.i.d. random variables such that $\vartheta_k^{(n)}$ has the distribution $\mathcal{B}_0(-\tilde{v}_n, 2)$ (defined by (4.2.5)) and $\tilde{v}_n := \sum_{k=1}^{n} E_k(2 - E_k)$. Moreover, since $x \mapsto x(2-x)$ is concave, $\tilde{v}_n \le \bar{E}(2 - \bar{E})$. Finally a classical result due to Hoeffding [9] (see his Inequalities (4.1) and (4.2)) yields that for any convex, nondecreasing function φ differentiable with convex derivative,

$$\mathbb{E}\Big[\varphi\Big(\sum_{k=1}^{n} \vartheta_k^{(n)}\Big)\Big] \le \mathbb{E}\Big[\varphi\Big(2\sum_{k=1}^{n} \theta_k^{(n)}\Big)\Big], \tag{4.5.18}$$

This inequality associated with (4.5.17) conclude the proof of Lemma 4.5.3. ☐

As mentioned at the beginning of the proof, this also concludes the proof of Lemma 4.5.2 by taking $\varphi(x) = e^{tx}$ with $t \geq 0$. $\square$

Let us now complete the proof of Theorem 4.3.1. From (4.2.3) and Lemmas 4.5.1–4.5.2 we derive for any $x \geq 0$,

$$\ell_{\Xi_n}^{*-1}(nx) \leq n\,\ell_{\mathscr{F}}^{*-1}(x) \text{ and } \ell_{R_n}^{*-1}(nx) \leq 2n\,\ell_{v_n}^{*-1}(x). \tag{4.5.19}$$

Furthermore, from Proposition 4.2.4 (*iii*), (*iv*) and (4.5.9)

$$\begin{aligned}\tilde{Q}_{Z-\mathbb{E}[Z]}(e^{-nx}) &\leq \tilde{Q}_{\Xi_n}(e^{-nx}) + \tilde{Q}_{R_n}(e^{-nx}) \\ &\leq \ell_{\Xi_n}^{*-1}(nx) + \ell_{R_n}^{*-1}(nx).\end{aligned} \tag{4.5.20}$$

Finally, both (4.5.20) and (4.5.19) conclude the proof of Theorem 4.3.1 (*a*). $\square$

Proof of Corollary 4.3.2. Let Π_n be a random variable with Poisson distribution with parameter v_n and let $\tilde{\Pi}_n := \Pi_n - v_n$. A classical result gives $\ell_{v_n}(t) \leq \ell_{\tilde{\Pi}_n}(t)$ for any $t \geq 0$ (see, for instance, Theorem 2.9 of [5]). Therefore, for any $x \geq 0$,

$$\ell_{v_n}^{*-1}(x) \leq \ell_{\tilde{\Pi}_n}^{*-1}(x) = v_n\, h^{-1}\Big(\frac{x}{v_n}\Big), \tag{4.5.21}$$

where $h(u) := (1+u)\log(1+u) - u$ for any $u \geq 0$. Next, a Newton algorithm performed in Del Moral and Rio [7] (see their Appendix A.6, especially (A.15)) allows to derive the bound $h^{-1}(x) \leq \psi(x)$, which concludes the proof of Corollary 4.3.2. $\square$

Proof of Proposition 4.3.6. Inequality (2.8) in Baraud [1] implies that for any $k = 1, \ldots, n$,

$$E_k \leq 2\sqrt{2}\,\sigma \log(e/\sigma)\sqrt{\tilde{\Gamma}_k(d)} + 8\,\tilde{\Gamma}_k(d), \tag{4.5.22}$$

where

$$\tilde{\Gamma}_k(d) := k^{-1}\log\left(2\sum_{j=0}^{d\wedge k}\binom{k}{j}\right). \tag{4.5.23}$$

We recall (see p. 1714 in [1]) that, for any $d \geq k$, $\tilde{\Gamma}_k(d) = \log(2)(k+1)/k$ and for any $d < k$,

$$\tilde{\Gamma}_k(d) \leq \frac{d}{k}\log\left(\frac{2e}{d}k\right). \tag{4.5.24}$$

Moreover, observe that by the concavity of $x \mapsto \sqrt{x}$, one has

$$\frac{1}{n}\sum_{k=1}^{n}\sqrt{\tilde{\Gamma}_k(d)} \leq \sqrt{\frac{1}{n}\sum_{k=1}^{n}\tilde{\Gamma}_k(d)}. \tag{4.5.25}$$

Then, since $\bar{E} = n^{-1}(E_1 + \ldots + E_n)$, the previous facts together with the sub-additivity of $x \mapsto \sqrt{x}$ yield

$$\bar{E} \leq 2\sqrt{2}\,\sigma \log(e/\sigma) n^{-1/2}\Big(\sqrt{C_1(d)} + \Big(\sum_{k=d+1}^{n} \frac{d}{k} \log\Big(\frac{2e}{d}k\Big)\Big)^{1/2}\Big)$$
$$+ 8\,n^{-1}\Big(C_1(d) + \sum_{k=d+1}^{n} \frac{d}{k} \log\Big(\frac{2e}{d}k\Big)\Big). \quad (4.5.26)$$

Observe now that the function h defined by $h(x) := x^{-1}\log((2e/d)x)$ is convex (at least) on $[d, +\infty[$. Thus for any integer $k > d$, $h(k) \leq \int_{k-1/2}^{k+1/2} h(x)dx$. Summing then this inequality from $d+1$ to n gives

$$\sum_{k=d+1}^{n} h(k) \leq \frac{1}{2}\log\Big(\frac{n+1/2}{d+1/2}\Big)\log\Big(\frac{4e^2}{d^2}(n+1/2)(d+1/2)\Big). \quad (4.5.27)$$

Finally injecting (4.5.27) in (4.5.26) concludes the proof. □

4.5.2 Proofs of Section 4.4

Proof of Lemma 4.4.1. We use the notation $I(y) := \inf_{f\in\mathscr{F}} \ell_f^*(y)$ throughout the proof.

(***i***) Proof of $J(y) \leq I(y)$.
Let $y \in [0,1]$ and let $\varepsilon > 0$. There exists a function $f \in \mathscr{F}$ such that $\ell_f^*(y) \leq I(y) + \varepsilon$. Now, Cramér's Theorem ensures that

$$\lim_{n\to\infty} -n^{-1}\log \mathbb{P}(P_n(f) \geq y) = \ell_f^*(y). \quad (4.5.28)$$

Since $\|P_n\|_{\mathscr{F}}$ satisfies the LDP with rate function J and since $P_n(f) \leq \|P_n\|_{\mathscr{F}}$ for all $f \in \mathscr{F}$, we get

$$J(y) \leq \limsup_{n\to\infty} -n^{-1}\log \mathbb{P}(\|P_n\|_{\mathscr{F}} \geq y)$$
$$\leq \lim_{n\to\infty} -n^{-1}\log \mathbb{P}(P_n(f) \geq y) = \ell_f^*(y). \quad (4.5.29)$$

Therefrom $J(y) \leq I(y) + \varepsilon$. Since $\varepsilon > 0$ is arbitrary, we conclude the proof of "$\leq$" by letting ε tend to 0.

(***ii***) Proof of $J(y) \geq I(y)$.
Since the infima may be written as the limit of a sequence of infima taken over finit subsets, it is enough to prove the inequality for a finite class of functions $\mathscr{F}$. Let $y \in [0,1]$ and $t > 0$. Let ν be a probability measure absolutely continuous with respect to P such that $\|\nu\|_{\mathscr{F}} = y$. Let $d := (d\nu/dP)$ be the

Radon-Nikodym derivative of ν with respect to P and set $g_f := tf - \log P(e^{tf})$ for any $f \in \mathscr{F}$. Young's inequality (see, for instance, Equation (A.2) in Rio [20]) implies that

$$t\nu(f) - \log P(e^{tf}) = \int dg_f \, dP \leq \int e^{g_f} dP + \int (d \log d - d) \, dP. \tag{4.5.30}$$

Since $\int e^{g_f} dP = 1$, (4.5.30) leads to

$$t\nu(f) - \log P(e^{tf}) \leq H(\nu \mid P). \tag{4.5.31}$$

In particular, (4.5.31) is valid for the function $\tilde{f} \in \mathscr{F}$ which satisfies $y = \nu(\tilde{f})$ (recall that $\mathscr{F}$ is finite) and for any $t > 0$. Then we have

$$\ell^*_{\tilde{f}}(y) \leq H(\nu \mid P), \tag{4.5.32}$$

which implies $I(y) \leq J(y)$ and ends the proof. □

Proof of Corollary 4.4.5. Let X be a random variable with distribution P. Recalling that $P(f) = 0$ for any $f \in \mathscr{F}$, Lemma 2 of Bobkov [4] states that for any convex function $\varphi : \mathbb{R} \to \mathbb{R}$ and for any $f \in \mathscr{F}$,

$$\mathbb{E}[\varphi(f(X))] \leq \mathbb{E}[\varphi(X - \mathbb{E}[X])].$$

In particular with $\varphi(x) = e^{tx}$, $t \geq 0$,

$$\ell_{\mathscr{F}}(t) = \sup_{f \in \mathscr{F}} \ell_f(t) \leq \log \mathbb{E}[e^{t(X - \mathbb{E}[X])}] = \ell_{X - \mathbb{E}[X]}(t). \tag{4.5.33}$$

Thus the variational formula (4.2.3) implies $\ell^{*-1}_{\mathscr{F}}(x) \leq \ell^{*-1}_{X-\mathbb{E}[X]}(x)$ for all $x \geq 0$. An application of Theorem 4.3.1 completes the proof. □

References

[1] Y. Baraud. Bounding the expectation of the supremum of an empirical process over a (weak) VC-major class. *Electron. J. Stat.*, 10(2):1709–1728, 2016.

[2] G. Bennett. On the probability of large deviations from the expectation for sums of bounded, independent random variables. *Biometrika*, 50:528–535, 1963.

[3] V. Bentkus. On Hoeffding's inequalities. *Ann. Probab.*, 32(2):1650–1673, 04 2004.

[4] S.G. Bobkov. Localization proof of the Bakry–Ledoux isoperimetric inequality and some applications. *Theory of Probability & Its Applications*, 47(2):308–314, 2003.

[5] S. Boucheron, G. Lugosi, and P. Massart. *Concentration Inequalities: A Nonasymptotic Theory of Independence*. OUP Oxford, 2013.

[6] O. Bousquet. Concentration inequalities for sub-additive functions using the entropy method. In *Stochastic inequalities and applications*, volume 56 of *Progr. Probab.*, pages 213–247. Birkhäuser, Basel, 2003.

[7] P. Del Moral and E. Rio. Concentration inequalities for mean field particle models. *Ann. Appl. Probab.*, 21(3):1017–1052, 2011.

[8] A. Dembo and O. Zeitouni. *Large deviations techniques and applications*, volume 38 of *Stochastic Modelling and Applied Probability*. Springer-Verlag, Berlin, 2010. Corrected reprint of the second (1998) edition.

[9] W. Hoeffding. Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.*, 58:13–30, 1963.

[10] T. Klein. Une inégalité de concentration à gauche pour les processus empiriques. *C. R. Math. Acad. Sci. Paris*, 334(6):501–504, 2002.

[11] T. Klein and E. Rio. Concentration around the mean for maxima of empirical processes. *Ann. Probab.*, 33(3):1060–1077, 05 2005.

[12] M. Ledoux. On Talagrand's deviation inequalities for product measures. *ESAIM: Probability and Statistics*, 1:63–87, 1997.

[13] A. Marchina. Concentration inequalities for suprema of unbounded empirical processes. Preprint on <hal-01545101>, 2017.

[14] P. Massart. About the constants in Talagrand's concentration inequalities for empirical processes. *Ann. Probab.*, 28(2):863–884, 2000.

[15] I. Pinelis. An optimal three-way stable and monotonic spectrum of bounds on quantiles: a spectrum of coherent measures of financial risk and economic inequality. *Risks*, 2(3):349–392, 2014.

[16] I. Pinelis. Convex cones of generalized multiply monotone functions and the dual cones. *Banach J. Math. Anal.*, 10(4):864–897, 10 2016.

[17] E. Rio. Inégalités de concentration pour les processus empiriques de classes de parties. *Probab. Theory Related Fields*, 119(2):163–175, 2001.

[18] E. Rio. Une inégalité de Bennett pour les maxima de processus empiriques. *Ann. Inst. H. Poincaré Probab. Statist.*, 38(6):1053–1057, 2002. En l'honneur de J. Bretagnolle, D. Dacunha-Castelle, I. Ibragimov.

[19] E. Rio. Sur la fonction de taux dans les inégalités de Talagrand pour les processus empiriques. *C. R. Math. Acad. Sci. Paris*, 350(5-6):303–305, 2012.

[20] E. Rio. *Asymptotic theory of weakly dependent random processes*, volume 80 of *Probability Theory and Stochastic Modelling*. Springer, Berlin, 2017. Translated from the 2000 French edition [MR2117923].

[21] M. Sion. On general minimax theorems. *Pacific J. Math.*, 8:171–176, 1958.

[22] M. Talagrand. New concentration inequalities in product spaces. *Invent. Math.*, 126(3):505–563, 1996.

[23] A.W. van der Vaart and J.A. Wellner. *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer Series in Statistics. Springer New York, 1996.

[24] L. Wu. Large deviations, moderate deviations and LIL for empirical processes. *Ann. Probab.*, 22(1):17–27, 1994.

Chapter 5

Comparison inequalities for suprema of bounded empirical processes

In this Chapter we provide comparison moment inequalities for suprema of bounded empirical processes. Our methods are only based on a decomposition in martingale and on comparison results concerning martingales proved by Bentkus and Pinelis.

5.1 Introduction

Let $X_1, \ldots, X_n$ be a sequence of independent random variables valued in some measurable space $(\mathcal{X}, \mathcal{F})$ with common distribution P. Let P_n denote the empirical probability measure $P_n := n^{-1}(\delta_{X_1} + \ldots + \delta_{X_n})$. Let $\mathscr{F}$ be a countable class of measurable functions $f : \mathcal{X} \to \mathbb{R}$ such that $P(f) = 0$ for all $f \in \mathscr{F}$. In this Chapter we are concerned with concentration properties around the mean of the random variable

$$Z := \sup\{nP_n(f) : f \in \mathscr{F}\}, \tag{5.1.1}$$

when $\mathscr{F}$ satisfied a two-sided or a one-sided (from above) boundedness conditions. Our approach is based on a decomposition of $Z - \mathbb{E}[Z]$ into a sum of martingale increments together with comparison inequalities for martingales with (two-sided or one-sided) bounded increments proved by Bentkus [1] and Pinelis [7]. Before going further, let us introduce some notations.

Definition 5.1.1. *(i) Let α, β be two reals such that $\alpha < \beta$. We say that a random variable θ follows a Bernoulli distribution if it assumes exactly two* *values and we write* $\theta \sim \mathcal{B}_m(\alpha, \beta)$ *if*

$$\mathbb{P}(\theta = \beta) = 1 - \mathbb{P}(\theta = \alpha) \in]0, 1[, \text{ and } \mathbb{E}[\theta] = m. \tag{5.1.2}$$

Notice that

$$\mathrm{Var}(\theta) = (m - \alpha)(\beta - m). \tag{5.1.3}$$

(ii) For any $a \geq 0$, Γ_a *stands for any centered Gaussian random variable with variance equals to* a.
(iii) For any $a > 0$, Π_a *stands for any Poisson random variable with parameter* a. *We also denote by* $\tilde{\Pi}_a := \Pi_a - a$ *the centered Poisson random variable.*

Let us introduce the class of convex functions in which the comparison inequalities, stated in this Chapter, are valid.

Definition 5.1.2. *Let* $k \in \mathbb{N}^*$. *As usual, we denote by* $\mathcal{C}^k$ *the space of* k*-times continuously diffenrentiable functions from* $\mathbb{R}$ *to* $\mathbb{R}$. *We define the following class of functions:*

$$\mathcal{G}^k := \{\varphi \in \mathcal{C}^{k-1} : \varphi,\, \varphi', \ldots, \varphi^{(k-1)} \text{ are convex}\}. \tag{5.1.4}$$

We now recall the two following comparison results, proved by Bentkus [1] and Pinelis [7], which are the main tools in our proofs.

Proposition 5.1.3 (Lemmas 4.4 and 4.5 in [1] and Theorem 2.1 and Remark 2.3 in [7])**.** *Let* $s_1^2, \ldots, s_n^2, \beta_1, \ldots, \beta_n$ *be positive reals. Let* $M_n := \sum_{k=1}^n X_k$ *be a martingale with respect to a nondecreasing filtration* $(\mathcal{F}_k)$ *such that* $M_0 = 0$,

$$X_k \leq 1, \quad \text{and} \quad \mathbb{E}[X_k^2 \mid \mathcal{F}_{k-1}] \leq s_k^2 \quad \text{a.s.} \tag{5.1.5}$$

(i) Let $s^2 := n^{-1}(s_1^2 + \ldots + s_n^2)$ *and* $S_n := \theta_1 + \ldots + \theta_n$ *be a sum of* n *independent copies of a random variable* θ *with distribution* $\mathcal{B}_0(-s^2, 1)$ *(defined by* (5.1.2)*). Then for any* $\varphi \in \mathcal{G}^2$,

$$\mathbb{E}[\varphi(M_n)] \leq \mathbb{E}[\varphi(S_n)]. \tag{a}$$

(ii) Additionally to (5.1.5), *assuming that*

$$\mathbb{E}_{k-1}[X_{k+}^3] \leq \beta_k \text{ a.s., and } \beta := \frac{1}{n}\sum_{k=1}^n \beta_k \leq s^2, \tag{5.1.6}$$

we have for any $\varphi \in \mathcal{G}^3$,

$$\mathbb{E}[\varphi(M_n)] \leq \mathbb{E}[\varphi(\Gamma_{n(s^2-\beta)} + \tilde{\Pi}_{n\beta})], \tag{b}$$

where $\Gamma_{n(s^2-\beta)}$ *and* $\tilde{\Pi}_{n\beta}$ *are independent and respectively defined by Definition 5.1.1 (ii) and (iii).*

Remark 5.1.4. *In fact, the results in the original papers are stated in the following slightly smaller class of functions*

$$\mathcal{H}_+^\alpha := \Big\{\varphi : \varphi(u) = \int_{-\infty}^{\infty} (u-t)_+^\alpha \, \mu(dt) \text{ for some Borel measure } \mu \geq 0 \text{ on } \mathbb{R} \\ \text{and all } u \in \mathbb{R}\Big\},$$

for $\alpha \in \{2,3\}$. *The extensions to* $\mathcal{G}^\alpha$ *follows from a result of Pinelis [8, Corollary 5.8] (see also [6, Section 2]).*

Remark 5.1.5. *From moment comparison inequalities in* $\mathcal{H}_+^\alpha$*, such as in the above Proposition, one can derive tail comparison inequalities. We refer the reader to [4, 5] for the statements of these results and for some more details.*

Finally, we use the notations:

$$E_k := \mathbb{E} \sup_{f \in \mathscr{F}} P_k(f) \text{ for any } k = 1, \ldots, n \text{ and } \bar{E} := n^{-1}(E_1 + \ldots + E_n). \tag{5.1.7}$$

5.2 Results

5.2.1 Two-sided boundedness condition

Here, by two-sided boundedness condition, we mean that $\mathscr{F}$ is a countable class of measurable functions with values in $[-a, 1]$ for some positive real a. Let ψ be the function defined on $[0,1]$ by

$$\psi(x) = x(1-x) \text{ if } x \in [0, 1/2] \text{ and } \psi(x) = 1/4 \text{ if } x \in]1/2, 1]. \tag{5.2.1}$$

Theorem 5.2.1. *Let* $\mathscr{F}$ *be a countable class of measurable functions from* $\mathcal{X}$ *into* $[-a,1]$ *such that* $P(f) = 0$ *for all* $f \in \mathscr{F}$. *Let* Z *be defined by* (5.1.1).
(i) Case $a \geq 1$.
Let θ *be a Bernoulli random variable with distribution* $\mathcal{B}_0(-a, 1)$ *(defined by* (5.1.2)*). Let* $\theta_1, \ldots, \theta_n$ *be* n *independent copies of* θ *and let* $S_n := \theta_1 + \ldots + \theta_n$. *Then for any function* $\varphi \in \mathcal{G}^2$,

$$\mathbb{E}[\varphi(Z - \mathbb{E}Z)] \leq \mathbb{E}[\varphi(S_n)]. \tag{a}$$

(ii) Case $a < 1$.
Let ϑ *be a Bernoulli random variable with distribution*

$$\mathcal{B}_0\Big(-(a+1)^2\, \psi\Big(\frac{a+\bar{E}}{a+1}\Big), 1\Big)$$

(defined by (5.1.2)*), where* ψ *is defined by* (5.2.1)*, and* $\bar{E}$ *is defined by* (5.1.7)*. Let* $\vartheta_1, \dots, \vartheta_n$ *be* n *independent copies of* ϑ *and let* $T_n := \vartheta_1 + \dots + \vartheta_n$*. Then for any function* $\varphi \in \mathcal{G}^2$*,*

$$\mathbb{E}[\varphi(Z - \mathbb{E}Z)] \leq \mathbb{E}[\varphi(T_n)]. \qquad (b)$$

Remark 5.2.2. *If the class* $\mathscr{F}$ *satisfies the uniform law of large numbers, that is* $\sup_{f \in \mathscr{F}} |P_n(f)|$ *converges to 0 in probability, then* E_n *decreases to 0 (see, for instance, Section 2.4 of van der Vaart and Wellner [10]) and so* $\bar{E}$ *also decreases to* 0*. This ensures that* $n \mapsto (a+1)^2 \psi((a+\bar{E})/(a+1))$ *(which is also the variance of* ϑ*) is nonincreasing and tends to* a *as* n *tends to infinity.*

Example 5.2.3 (Set-indexed empirical processes)**.** *Let* $\mathscr{S}$ *be a countable class of measurable sets of* $\mathcal{X}$*. We consider the class of functions*

$$\mathscr{F} = \left\{ \frac{\mathbb{1}_S - P(S)}{1 - P(S)} : S \in \mathscr{S} \right\}. \qquad (5.2.2)$$

Let $p := \sup\{P(S) : S \in \mathscr{S}\}$ *and we assume that* $p < 1$*. Since* $x \mapsto x/(1-x)$ *is increasing on* $[0,1[$*, we can apply Theorem 5.2.1 with* $a = p/(1-p)$*. Hence with the notations of Theorem 5.2.1:*
(i) If $p > 1/2$*, then for any* $\varphi \in \mathcal{G}^2$*,* $\mathbb{E}[\varphi(Z - \mathbb{E}[Z])] \leq \mathbb{E}[\varphi(S_n)]$*, and* θ *has the distribution* $\mathcal{B}_0(-p/(1-p), 1)$*.*
(ii) If $p < 1/2$*, then for any* $\varphi \in \mathcal{G}^2$*,* $\mathbb{E}[\varphi(Z - \mathbb{E}[Z])] \leq \mathbb{E}[\varphi(T_n)]$*, and* ϑ *has the distribution* $\mathcal{B}_0(-(1-p)^{-2}\psi(\bar{E} + p(1-\bar{E})), 1)$*.*

Theorem 3.1 of Rio [9], when applied to Z *(see also his Theorem 4.2 (a)), provides a Bennett-type inequality for class of sets with small measures under* P*. Precisely the condition is* $E_n + p(1 - E_n) \leq 1/2$*. Hence, since* $\mathcal{G}^2$ *contains all increasing exponential functions* $x \mapsto e^{tx}$*,* $t > 0$*, the Case (ii) above completes Rio's result in this situation.*

5.2.2 One-sided boundedness condition

Here, by one-sided boundedness condition, we mean that $\mathscr{F}$ is a countable class of measurable functions with values in $]-\infty, 1]$. Let ρ be the function defined on $[0,1]$ by

$$\rho(x) = x(1-x)^2 \text{ if } x \in [0, 1/3] \text{ and } \rho(x) = 4/27 \text{ if } x \in]1/3, 1]. \qquad (5.2.3)$$

Theorem 5.2.4. *Let* $\mathscr{F}$ *be a countable class of measurable functions from* $\mathcal{X}$ *into* $]-\infty, 1]$ *such that* $P(f) = 0$ *for any* $f \in \mathscr{F}$*. Let* Z *be defined by* (5.1.1)*. Define also*

$$\sigma^2 := \sup\{P(f^2) : f \in \mathscr{F}\} \text{ and } m_+^3 := \sup\{P(f_+^3) : f \in \mathscr{F}\}, \qquad (5.2.4)$$

and

$$v^2 := \sigma^2 + 2\psi(\bar{E}) \quad \textit{and} \quad \beta_+^3 := m_+^3 + 3\rho(\bar{E}). \tag{5.2.5}$$

(i) Let θ be a Bernoulli random variable with distribution $\mathcal{B}_0(-v^2, 1)$ (defined by (5.1.2)). Let $\theta_1, \dots, \theta_n$ be n independent copies of θ and let $S_n := \theta_1 + \dots + \theta_n$. Then for any function $\varphi \in \mathcal{G}^2$,

$$\mathbb{E}[\varphi(Z - \mathbb{E}[Z])] \leq \mathbb{E}[\varphi(S_n)]. \tag{a}$$

(ii) Let $\beta := \min\{\beta_+^3, v^2\}$. Then for any function $\varphi \in \mathcal{G}^3$,

$$\mathbb{E}[\varphi(Z - \mathbb{E}[Z])] \leq \mathbb{E}[\varphi(\Gamma_{n(v^2-\beta)} + \tilde{\Pi}_{n\beta})], \tag{b}$$

where $\Gamma_{n(v^2-\beta)}$ and $\tilde{\Pi}_{n\beta}$ are independent and respectively defined by Definition 5.1.1 (ii) and (iii).

Remark 5.2.5. *Inequalities (a) and (b) of Theorem 5.2.4 are not easy to compare directly. To this end, we start by observing that Inequality (a) yields the following (see [1, Theorem 1.1]):*

$$\mathbb{E}[\varphi(Z - \mathbb{E}[Z])] \leq \mathbb{E}[\varphi(\tilde{\Pi}_{n\beta})] \quad \textit{for any } \varphi \in \mathcal{G}^2. \tag{a'}$$

One can find in Pinelis [6] a thorough study of the comparison between the right-hand sides of (b) and (a′) as well as with other classical bounds. Let us mention here some facts.
(i) If $\beta = v^2$, then the right-hand side of (b) is equal to the right-hand side of (a′). Thus, since $\mathcal{G}^3 \subset \mathcal{G}^2$, Inequality (b) is relevant with respect to (a) if and only if $\beta_+^3 < v^2$ (see also the point (ii) below).
(ii) For any $v^2 > 0$ and any $\varepsilon \in]0, 1[$,

$$\mathbb{E}[\varphi(\Gamma_{(1-\varepsilon)v^2} + \tilde{\Pi}_{\varepsilon v^2})] \leq \mathbb{E}[\varphi(\tilde{\Pi}_{v^2})] \quad \textit{for any } \varphi \in \mathcal{G}^2. \tag{5.2.6}$$

5.3 Proofs

The starting point of the proofs is a martingale decomposition of Z which we briefly recall. Firstly by virtue of the monotone convergence theorem, we can suppose that $\mathscr{F}$ is a finite class of functions. Set $\mathcal{F}_0 := \{\varnothing, \Omega\}$ and for all $k = 1, \dots, n$, $\mathcal{F}_k := \sigma(X_1, \dots, X_k)$ and $\mathcal{F}_n^k := \sigma(X_1, \dots, X_{k-1}, X_{k+1}, \dots, X_n)$. Let $\mathbb{E}_k$ (respectively $\mathbb{E}_n^k$) denote the conditional expectation operator associated with $\mathcal{F}_k$ (resp. $\mathcal{F}_n^k$). Set also $Z^{(k)} := \sup\{nP_n(f) - f(X_k) : f \in \mathscr{F}\}$ and $Z_k := \mathbb{E}_k[Z]$. Let us number the functions of the class $\mathscr{F}$ and consider the random variables

$$\tau := \inf\{i > 0 : nP_n(f_i) = Z\} \text{ and } \tau_k := \inf\{i > 0 : nP_n(f_i) - f_i(X_k) = Z^{(k)}\}.$$

We notice that

$$f_{\tau_k}(X_k) \leq Z - Z^{(k)} \leq f_\tau(X_k). \tag{5.3.1}$$

Set now $\xi_k := \mathbb{E}_k[f_{\tau_k}(X_k)]$ and let $r_k \geq 0$ be the random variable such that $\xi_k + r_k = Z_k - \mathbb{E}_k[Z^{(k)}]$. Since $\mathbb{E}_n^k[f_{\tau_k}(X_k)] = P(f_{\tau_k})$ (see, for instance, Section 3.2 in [3]), the centering assumption on the elements of $\mathscr{F}$ ensures that

$$Z - \mathbb{E}[Z] = \sum_{k=1}^{n} \Delta_k, \quad \text{where } \Delta_k := \xi_k + r_k - \mathbb{E}_{k-1}[r_k]. \tag{5.3.2}$$

The important point is that $\mathbb{E}_{k-1}[r_k]$ is a corrective term which is essentially small. This is the statement of the following lemma:

Lemma 5.3.1. *We have* $0 \leq \mathbb{E}_{k-1}[r_k] \leq E_{n-k+1} \leq 1$ *a.s.*

We refer the reader to [3] for a proof which rests on a property of exchangeability of variables (see Lemma 3.10 in [3]). We are now in position to prove Theorem 5.2.1.

Proof of Theorem 5.2.1. Observe first that $\Delta_k \leq 1$ by (5.3.1) and the uniform boundedness condition on $\mathscr{F}$. Then, in view of Proposition 5.1.3 (i), it remains us to bound up the conditional variance with respect to $\mathcal{F}_{k-1}$ of Δ_k. This is the subject of the following lemma:

Lemma 5.3.2. *For any* $k = 1, \ldots, n$,

$$\mathbb{E}_{k-1}[\Delta_k^2] \leq \sup_{m \in [0, E_{n-k+1}]} (m + a)(1 - m) \quad \text{a.s.}$$

Proof of Lemma 5.3.2. A classical result due to Hoeffding [2] (see his Inequalities (4.1) and (4.2)) states that any bounded random variable X such that $a \leq X \leq b$ for some reals a and b, satisfies

$$\mathbb{E}[\varphi(X)] \leq \mathbb{E}[\varphi(\theta)] \quad \text{for any convex function } \varphi, \tag{5.3.3}$$

where θ is a Bernoulli random variable with distribution $\mathcal{B}_{\mathbb{E}[X]}(a, b)$ (defined by (5.1.2)). In particular, $\mathrm{Var}(X)$ is lower than $\mathrm{Var}(\theta)$. We apply this result conditionally to $\mathcal{F}_{k-1}$ to the variable $\xi_k + r_k$ which has its values in $[-a, 1]$ by (5.3.1). Recalling now (5.1.3), one immediately obtains

$$\mathrm{Var}(\xi_k + r_k \mid \mathcal{F}_{k-1}) \leq (\mathbb{E}_{k-1}[r_k] + a)(1 - \mathbb{E}_{k-1}[r_k]) \text{ a.s.}, \tag{5.3.4}$$

which combined with Lemma 5.3.1 conclude the proof of Lemma 5.3.2. $\square$

Let us now complete the proof of Theorem 5.2.1. Define the function V by

$$V(m) := (m+a)(1-m) \quad \text{for any } m \in [0, E_{n-k+1}]. \tag{5.3.5}$$

Case (i): $a \geq 1$. Since $(1-a)/2 \leq 0$, V is decreasing on $[0, E_{n-k+1}]$ and then by Lemma 5.3.2, $\mathbb{E}_{k-1}[\Delta_k^2] \leq a$. Thus Inequality (a) follows by Proposition 5.1.3 (i).

Case (ii): $a < 1$. Here the maximum of V is reached at

$$m = \min(E_{n-k+1}, (1-a)/2).$$

Thus Lemma 5.3.2 implies

$$\mathbb{E}_{k-1}[\Delta_k^2] \leq (a+1)^2\, \psi\Big(\frac{a + E_{n-k+1}}{a+1}\Big). \tag{5.3.6}$$

Furthermore, since ψ is concave,

$$\frac{1}{n}\sum_{k=1}^{n} \psi\Big(\frac{a + E_{n-k+1}}{a+1}\Big) \leq \psi\Big(\frac{a + \bar{E}}{a+1}\Big). \tag{5.3.7}$$

Hence Inequality (b) follows again from Proposition 5.1.3 (i) together with (5.3.6)–(5.3.7). The proof of Theorem 5.2.1 is now complete. □

Proof of Theorem 5.2.4. Case (i). We start from (5.3.2). Since $\mathbb{E}_{k-1}[\xi_k] = 0$, we get

$$\mathbb{E}_{k-1}[\Delta_k^2] = \mathbb{E}_{k-1}[(\xi_k + r_k)^2] - (\mathbb{E}_{k-1}[r_k])^2\,. \tag{5.3.8}$$

Since $\mathbb{E}_n^k[f_{\tau_k}^2(X_k)] = P(f_{\tau_k}^2)$, the conditional Jensen inequality implies that $\mathbb{E}_{k-1}[\xi_k^2] \leq \sigma^2$. Therefore, from (5.3.8) and the fact that (5.3.1) implies $\xi_k + r_k \leq 1$, we get

$$\begin{aligned}
\mathbb{E}_{k-1}[\Delta_k^2] &\leq \sigma^2 + \mathbb{E}_{k-1}[2\xi_k r_k + r_k^2] - (\mathbb{E}_{k-1}[r_k])^2 \\
&\leq \sigma^2 + 2\mathbb{E}_{k-1}[r_k] - (\mathbb{E}_{k-1}[r_k])^2 \\
&\leq \sigma^2 + 2\,\psi(\mathbb{E}_{k-1}[r_k]) \\
&\leq \sigma^2 + 2\,\psi(E_{n-k+1}),
\end{aligned} \tag{5.3.9}$$

where the last inequality follows from Lemma 5.3.1 and ψ is the nondecreasing function already defined in (5.2.1). Then the same conclusion as in the proof of the Case (ii) of Theorem 5.2.1 allows us to conclude the proof of Inequality (a) of Theorem 5.2.4.

Case (ii). Since $x \mapsto x_+^3$ is a convex function, we have

$$\Delta_{k+}^3 \leq (\xi_k - \mathbb{E}_{k-1}[r_k])_+^3 + 3\, r_k\, (\xi_k + r_k - \mathbb{E}_{k-1}[r_k])_+^2\,. \tag{5.3.10}$$

In the same way as previously, since $\mathbb{E}_n^k[f_{\tau_k+}^3(X_k)] = P(f_{\tau_k+}^3)$, we observe that $\mathbb{E}_{k-1}[\xi_{k+}^3] \le m_+^3$. Therefrom, recalling that $r_k \ge 0$ and $\xi_k + r_k \le 1$, we get

$$\mathbb{E}_{k-1}[\Delta_{k+}^3] \le m_+^3 + 3\,\rho(\mathbb{E}_{k-1}[r_k]) \le m_+^3 + 3\,\rho(E_{n-k+1}), \tag{5.3.11}$$

where the last inequality follows from Lemma 5.3.1 and ρ is the nondecreasing function defined in (5.2.3). Since ρ is concave, we complete the proof in the same way as the Case (i) by using Proposition 5.1.3 (ii) in place of Proposition 5.1.3 (i).

□

References

[1] V. Bentkus. On Hoeffding's inequalities. *Ann. Probab.*, 32(2):1650–1673, 04 2004.

[2] W. Hoeffding. Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.*, 58:13–30, 1963.

[3] A. Marchina. Concentration inequalities for suprema of unbounded empirical processes. Preprint on <hal-01545101>, 2017.

[4] I. Pinelis. Optimal tail comparison based on comparison of moments. In *High dimensional probability*, pages 297–314. Springer, 1998.

[5] I. Pinelis. Fractional sums and integrals of r-concave tails and applications to comparison probability inequalities. *Advances in stochastic inequalities (Atlanta, GA, 1997)*, 234:149–168, 1999.

[6] I. Pinelis. On the Bennett–Hoeffding inequality. Preprint. Available at arXiv:0902.4058 [math.PR], 2009.

[7] I. Pinelis. On the Bennett–Hoeffding inequality. *Ann. Inst. H. Poincaré Probab. Statist.*, 50(1):15–27, 02 2014.

[8] I. Pinelis. Convex cones of generalized multiply monotone functions and the dual cones. *Banach J. Math. Anal.*, 10(4):864–897, 10 2016.

[9] E. Rio. Inégalités de concentration pour les processus empiriques de classes de parties. *Probab. Theory Related Fields*, 119(2):163–175, 2001.

[10] A.W. van der Vaart and J.A. Wellner. *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer Series in Statistics. Springer New York, 1996.

Chapter 6

An exponential inequality for suprema of empirical processes with heavy-tails on the left

In this Chapter, we give exponential inequalities around the mean for suprema of empirical processes indexed by classes of functions which are bounded from above and heavy-tail like on the left. Our approach is based on a martingale decomposition together with comparison results proved in Section 3 of Chapter 2. Furthermore, the constants in the inequalities are explicit.

6.1 Introduction

Let $X_1, \ldots, X_n$ be a sequence of independent random variables valued in some measurable space $(\mathcal{X}, \mathcal{F})$ with common distribution P. Let $\mathscr{F}$ be a class of measurable functions from $\mathcal{X}$ to $]-\infty, 1]$ such that $P(f) = 0$ for all $f \in \mathscr{F}$. Let $1 < p < 2$. We suppose that for all $f \in \mathscr{F}$, $f(X_1)$ satisfies the following behavior on the left:

$$\mathbb{P}(f(X_1) \leq -t) \leq t^{-p} \wedge 1 \quad \text{for any } t > 0. \tag{6.1.1}$$

In particular (6.1.1) admits that $\mathrm{Var}(f(X_1))$ is infinite. We are concerned with concentration properties around the mean of the random variable

$$Z := \sup_{f \in \mathscr{F}} \sum_{k=1}^{n} f(X_k). \tag{6.1.2}$$

Classical concentration inequalities assume that the functions of $\mathscr{F}$ are square integrable under P, and the aim is then to establish extensions of Hoeffding's, Bernstein's and Bennett's inequalities for suprema of empirical processes. We refer the reader to Chapter 12 of the book of Boucheron, Lugosi and Massart [1] for an overview of this subject. To the best of our knowledge, the only result where the square integrability of the functions of $\mathscr{F}$ is not required, is provided in Rio [2, Theorem 2]. Rio gives a bound of the log-Laplace transform of $Z - \mathbb{E}[Z]$ involving the squares of positive parts and truncated negative parts of $f(X_1)$ for all $f \in \mathscr{F}$. His proofs relies on a martingale decomposition of $Z - \mathbb{E}[Z]$ and on an exponential inequality for positive self-bounded functions proved in Rio [3]. The starting point of the proof of our main result is the same martingale decomposition. However we control the increments by comparison inequalities proved in Chapter 2.

Roughly speaking, we shall establish (see Inequality (a) of Theorem 6.2.1) that for small enough $x > 0$,

$$\mathbb{P}(Z - \mathbb{E}[Z] \geq n^{1/p} x) \leq \exp\Big(- K_p\, x^q \Big(1 + O\Big(x^{2/(p-1)} n^{-1/q}\Big)\Big)\Big), \qquad (6.1.3)$$

where $q = p/(p-1)$ is the Hölder exponent conjugate of p and K_p is a constant depending only on p. This result must be related to the following known result when the functions of $\mathscr{F}$ are square integrable under P (see, for instance, Rio [4, Theorem 1.1]):

$$\mathbb{P}(Z - \mathbb{E}[Z] \geq \sqrt{n} x) \leq \exp\Big(- \frac{x^2}{2v}\Big(1 + O\Big(x n^{-1/2}\Big)\Big)\Big), \qquad (6.1.4)$$

where $v := \sigma^2 + n^{-1}\mathbb{E}[Z]$ and $\sigma^2 := \sup_{f \in \mathscr{F}} P(f^2)$.

6.2 Result

Theorem 6.2.1. *Let $1 < p < 2$. Let c be a positive real satisfying*

$$c \geq 1 \ \ or \ \ \frac{p-1}{2p-1} \leq c^p < 1. \qquad (6.2.1)$$

Let q_0 be a real in $]0,1[$ such that

$$q_0 - c\,\frac{p}{p-1}(1-q_0)^{1-1/p} = 0. \qquad (6.2.2)$$

Let η be a random variable such that for any $t > 0$,

$$F_\eta(-t) = \Big(\frac{c}{t}\Big)^p \wedge 1 \ \ and \ \ F_\eta(t) = 1. \qquad (6.2.3)$$

Let $\mathscr{F}$ be a countable class of measurable functions from $\mathcal{X}$ into $]-\infty, 1]$ such that for any $f \in \mathscr{F}$ and any $t > 0$, $P(f) = 0$ and

$$\mathbb{E}[(-t - f(X))_+] \leq \mathbb{E}[(-t - \eta)_+]. \tag{6.2.4}$$

Let Z be defined by (6.1.2)*. Then for any $x \leq q_0\,\Gamma(2-p)$,*

$$\mathbb{P}(Z - \mathbb{E}Z \geq nx) \leq \exp\left(-n\gamma_p\, x^{p/(p-1)}(1 - \varepsilon_p(x))\right), \tag{a}$$

where

$$\gamma_p = (p\alpha_p)^{-1/(p-1)}\frac{p-1}{p} \quad \textit{and} \quad \alpha_p = \frac{c^p}{p-1}\Gamma(2-p), \tag{6.2.5}$$

$$\varepsilon_p(x) = \frac{p}{p-1}\sum_{k=2}^{\infty}\frac{q_0}{k!}x^{(k-p)/(p-1)}(p\alpha_p)^{-(k-1)/(p-1)}. \tag{6.2.6}$$

Moreover, for any $x > q_0\,\Gamma(2-p)$,

$$\mathbb{P}(Z - \mathbb{E}Z \geq nx) \leq \exp\Big(-n(x(1-q_0)^{1/p}c^{-1} - \beta_p)\Big), \tag{b}$$

where

$$\begin{aligned}\beta_p = {} & \frac{-\Gamma(2-p)}{p-1} \\ & - q_0\left(-\frac{\Gamma(2-p)}{p-1} - 1 + \exp((1-q_0)^{1/p}c^{-1}) - (1-q_0)^{1/p}c^{-1}\right).\end{aligned} \tag{6.2.7}$$

Remark 6.2.2. *Notice that if for any $f \in \mathscr{F}$ and any $t > 0$, we have*

$$\mathbb{P}(f(X_1) \leq -t) \leq \left(\frac{c}{t}\right)^p \wedge 1, \tag{6.2.8}$$

then the domination hypothesis (6.2.4) *is satisfied.*

Proof of Theorem 6.2.1. We start in the same way as in the proof of main results in Chapters 4 and 5 by a martingale decomposition of $Z - \mathbb{E}[Z]$. Let us recall the main points. Firstly by virtue of the monotone convergence theorem, we can suppose that $\mathscr{F}$ is a finite class of functions. Set $\mathcal{F}_0 := \{\varnothing, \Omega\}$ and for all $k = 1, \ldots, n$, $\mathcal{F}_k := \sigma(X_1, \ldots, X_k)$ and $\mathcal{F}_n^k := \sigma(X_1, \ldots, X_{k-1}, X_{k+1}, \ldots, X_n)$. Let $\mathbb{E}_k$ (respectively $\mathbb{E}_n^k$) denote the conditional expectation operator associated with $\mathcal{F}_k$ (resp. $\mathcal{F}_n^k$). Set also $Z^{(k)} := \sup_{f\in\mathscr{F}}\sum_{j\neq k} f(X_j)$ and $Z_k := \mathbb{E}_k[Z]$. Let us number the functions of the class $\mathscr{F}$ and consider the random variables

$$\tau := \inf\{i > 0 : nP_n(f_i) = Z\} \text{ and } \tau_k := \inf\{i > 0 : nP_n(f_i) - f_i(X_k) = Z^{(k)}\}.$$

Define $\xi_k := \mathbb{E}_k[f_{\tau_k}(X_k)]$ and $r_k := (Z_k - \mathbb{E}_k[Z^{(k)}]) - \xi_k \geq 0$. Note that we have

$$\xi_k \leq \xi_k + r_k \leq \mathbb{E}_k[f_\tau(X_k)]. \tag{6.2.9}$$

Then we have by the centering assumption on the element of $\mathscr{F}$,

$$Z - \mathbb{E}[Z] = \sum_{k=1}^{n} \Delta_k \quad \text{where } \Delta_k := \xi_k + r_k - \mathbb{E}_{k-1}[r_k]. \tag{6.2.10}$$

The strategy of the proof is the following one :

1. Using results of Section 3 of Chapter 2, we compare generalized (conditionnal) moments of Δ_k with a random variable ζ_{q_0}, with known distribution, built from the random variable η. The class of generalized moments on which we obtain a comparison inequality contains increasing exponential functions $x \mapsto e^{tx}$ for every $t \geq 0$.
2. We analyze the exponential moments of ζ_{q_0}.
3. We conclude the proof by the usual Cramér-Charnoff calculation.

Part 1 : Comparison inequality.

Let us denote by ζ_q, for any $q \in [0,1]$, a random variable with distribution function

$$F_q(x) := \begin{cases} F_\eta(x) & \text{if } x < a_q, \\ 1-q & \text{if } a_q \leq x < 1, \\ 1 & \text{if } x \geq 1, \end{cases} \tag{6.2.11}$$

where $a_q := F_\eta^{-1}(1-q)$. Notice that

$$\mathbb{E}[\zeta_q] = q - c\frac{p}{p-1}(1-q)^{1-1/p}, \tag{6.2.12}$$

which ensures that (6.2.2) is equivalent to $\mathbb{E}[\zeta_{q_0}] = 0$. Let us also define

$$\mathcal{H}_+^2 := \Big\{\varphi : \varphi(u) = \int_{-\infty}^{\infty} (u-t)_+^2 \, \mu(dt) \text{ for some Borel measure } \mu \geq 0 \text{ on } \mathbb{R} \text{ and all } u \in \mathbb{R}\Big\}.$$

We get the following lemma :

Lemma 6.2.3. *For any $\varphi \in \mathcal{H}_+^2$ and any $k = 1, \ldots, n$, we have*

$$\mathbb{E}_{k-1}[\varphi(\Delta_k)] \leq \mathbb{E}[\varphi(\zeta_{q_0})]. \tag{a}$$

Consequently, for any $t \geq 0$,

$$\log \mathbb{E}[\exp(t(Z - \mathbb{E}[Z]))] \leq n \log \mathbb{E}[\exp(t\,\zeta_{q_0})]. \tag{b}$$

Proof of Lemma 6.2.3. Since $\mathcal{H}^2_+$ contains all increasing exponential functions, taking $\varphi(x) = e^{tx}$ with $t \geq 0$ in (a) leads to (b) by an induction on n. Let us now prove (a). Note that (6.2.9) and (6.2.4) imply that, for any $t > 0$,

$$\mathbb{E}_{k-1}[(-t - (\xi_k + r_k))_+] \leq \mathbb{E}[(-t - \eta)_+]. \tag{6.2.13}$$

Furthermore, we can directly verify that the inequality (6.2.13) holds for $t \leq 0$. Since $f \leq 1$ for any $f \in \mathscr{F}$, the same inequality (6.2.9) implies that $\xi_k + r_k \leq 1$. Hence, conditionally to $\mathcal{F}_{k-1}$, the hypotheses (2.4.2) of Lemma 2.4.3 are satisfied with $X = \xi_k + r_k$, $\psi = 1$ and η the random variable defined by (6.2.3). Thus we get for any convex function φ,

$$\mathbb{E}_{k-1}[\varphi(\Delta_k)] \leq \mathbb{E}_{k-1}[\varphi(\zeta_{\hat{q}} - \mathbb{E}[\zeta_{\hat{q}}])], \tag{6.2.14}$$

where $\hat{q} \in [0,1]$ is such that $\mathbb{E}_{k-1}[r_k] = \mathbb{E}_{k-1}[\zeta_{\hat{q}}]$. Now, in order to control the right-hand side of (6.2.14), we shall use Lemma 2.4.6 (i) of Chapter 2. Let us define $\tilde{q} := \inf\{q \geq 1/2 : 1 + F_\eta^{-1}(1-q) \leq 2\mathbb{E}[\zeta_q]\}$ as in Lemma 2.4.6 (i) (with $\psi = 1$ and η the random variable defined by (6.2.3)). We shall prove the following :

Lemma 6.2.4. *We have*

(i) $1 + F_\eta^{-1}(1-q_0) \leq 0 = 2\,\mathbb{E}[\zeta_{q_0}]$. *Consequently,* $\tilde{q} \leq q_0$.

(ii) $q_0 \leq \hat{q}$.

Proof of Lemma 6.2.4. Let us first prove (i). Define $q_1 := 1 - \min(c^p, 1)$. One can verify that the conditions (6.2.1) on c imply $\mathbb{E}[\zeta_{q_1}] \leq 0$. Furthermore, observe that $\mathbb{E}[\zeta_q]$ is nondecreasing in q. Then, since $\mathbb{E}[\zeta_{q_0}] = 0$, one has $q_1 \leq q_0$. Next, since $F_\eta(-1) = 1 - q_1$, we have $F_\eta^{-1}(1-q_1) \leq -1$. Therefrom, since F_η^{-1} is nondecreasing,

$$1 + F_\eta^{-1}(1-q_0) \leq 1 + F_\eta^{-1}(1-q_1) \leq 0.$$

Thus, since $2\mathbb{E}[\zeta_{q_0}] = \mathbb{E}[\zeta_{q_0}] = 0$, by the definition of $\tilde{q}$, one has $\tilde{q} \leq q_0$ which ends the proof of (i). Let us now turn to the proof of (ii). Since $\mathbb{E}_{k-1}[\zeta_{\hat{q}}] = \mathbb{E}_{k-1}[r_k] \geq 0$ and, again, that $\mathbb{E}[\zeta_q]$ is nondecreasing in q and $\mathbb{E}[\zeta_{q_0}] = 0$, it follows that $q_0 \leq \hat{q}$ which concludes the proof of Lemma 6.2.4. □

Lemma 2.4.6 (i) of Chapter 2 guarantees us that $q \mapsto \mathbb{E}[(\zeta_q - \mathbb{E}[\zeta_q] - t)^2_+]$ is nonincreasing on $[\tilde{q}, 1]$. Then Lemma 6.2.4 and (6.2.14) yield for any $t \in \mathbb{R}$,

$$\mathbb{E}_{k-1}[(\Delta_k - t)^2_+] \leq \mathbb{E}_{k-1}[(\zeta_{\hat{q}} - \mathbb{E}[\zeta_{\hat{q}}] - t)^2_+] \leq \mathbb{E}[(\zeta_{q_0} - \mathbb{E}[\zeta_{q_0}] - t)^2_+],$$

which then implies Inequality (a) of Lemma 6.2.3 by the definition of $\mathcal{H}^2_+$, and finishes the proof. □

Part 2 : Analysis of exponential moments of ζ_{q_0}.

In this part, we bound up $\mathbb{E}[\exp(t\zeta_{q_0})]$ for any $t \geq 0$. First we recall the notation $a_q = F_\eta^{-1}(1-q) \leq 0$. We have

Lemma 6.2.5. *Let $t \geq 0$ such that $-ta_{q_0} \leq 1$. Then*

$$\log \mathbb{E}[e^{t\zeta_{q_0}}] \leq q_0(e^t - t - 1) + \alpha_p t^p.$$

Proof of Lemma 6.2.5. Let $t > 0$. Starting from the definition of the random variables ζ_q, one has

$$\begin{aligned}\mathbb{E}[e^{t\zeta_{q_0}}] &= q_0 e^t + p(tc)^p \int_{-ta_{q_0}}^{\infty} e^{-u} u^{-(p+1)} du \\ &= q_0 e^t + p(tc)^p \Big(\int_0^\infty u^{-(p+1)}(e^{-u} - 1 + u) du \\ &\qquad + \int_{-ta_{q_0}}^{\infty} u^{-(p+1)}(1-u) du - \int_0^{-ta_{q_0}} u^{-(p+1)}(e^{-u} - 1 + u) du \Big). \end{aligned} \tag{6.2.15}$$

Now for any $1 < p < 2$,

$$\int_0^\infty u^{-(p+1)}(e^{-u} - 1 + u) du = \Gamma(-p), \tag{6.2.16}$$

where $\Gamma(-p) = \frac{1}{p(p-1)}\Gamma(2-p)$ and $\Gamma(.)$ is the usual Gamma function. Moreover, using the expansion $e^{-u} = \sum_{k=0}^{\infty}(-u)^k/k!$, we can prove that

$$\begin{aligned}\int_{-ta_{q_0}}^{\infty} u^{-(p+1)}(1-u) du &- \int_0^{-ta_{q_0}} u^{-(p+1)}(e^{-u} - 1 + u) du \\ &= -\sum_{k=0}^{\infty} \frac{(-1)^k}{k!} \frac{(-ta_{q_0})^{k-p}}{k-p}. \end{aligned} \tag{6.2.17}$$

Thus, since $q_0 = 1 - (c/-a_{q_0})^p$, (6.2.15) becomes

$$\begin{aligned}\mathbb{E}[e^{t\zeta_{q_0}}] &= q_0 e^t + p(tc)^p \Gamma(-p) \\ &\quad + (1-q_0)\Big(1 - ta_{q_0}\frac{p}{1-p} + \sum_{k=2}^{\infty}(-1)^{k+1}\frac{(-ta_{q_0})^k}{k!}\frac{p}{k-p}\Big). \end{aligned} \tag{6.2.18}$$

Remark 6.2.6. *In fact, we can prove in the same way that* (6.2.18) *is valid for any noninteger $p > 1$.*

Next, we observe that under the hypothesis $-ta_{q_0} \leq 1$, the sum in (6.2.18) is an alternating series with the the absolute value of the general term decreasing to 0. Thus the sum is of the sign of the term corresponding to $k = 2$, which is negative. Hence,

$$\begin{aligned}\mathbb{E}[e^{t\zeta_{q_0}}] &\leq q_0 e^t + p(tc)^p \Gamma(-p) + (1-q_0)\Big(1 - ta_{q_0}\frac{p}{1-p}\Big) \\ &\leq 1 + q_0(e^t - t - 1) + p(tc)^p\Gamma(-p) + t\Big(q_0 - (1-q_0)a_{q_0}\frac{p}{1-p}\Big).\end{aligned} \tag{6.2.19}$$

Observe that, in view of (6.2.12),

$$q_0 - (1-q_0)a_{q_0}\frac{p}{1-p} = \mathbb{E}[\zeta_{q_0}] = 0.$$

Hence taking the logarithm and using the inequality $\log(1+x) \leq x$ for any $x > 0$ conclude the proof of Lemma 6.2.5. □

Part 3 : Conclusion by the Cramér-Chernoff calculation.

We now complete the proof of Theorem 6.2.1. From Lemma 6.2.3 (b) and Lemma 6.2.5, by the usual Cramér-Chernoff calculation, we get

$$\mathbb{P}(Z - \mathbb{E}Z \geq nx) \leq \exp\left(-n\,\phi^*_{\zeta_{q_0}}(x)\right), \tag{6.2.20}$$

where

$$\phi^*_{\zeta_{q_0}}(x) = \sup_{t\in]0,1/a_{q_0}]}\left\{tx - \alpha_p t^p - q_0(e^t - t - 1)\right\}. \tag{6.2.21}$$

In order to prove Inequality (a) of the Theorem, we bound from below $\phi^*_{\zeta_{q_0}}(x)$ by taking the real $t_x \in]0, 1/a_{q_0}]$ which maximizes $t \mapsto tx - \alpha_p t^p$. A straightforward calculation yields

$$t_x = x^{1/(p-1)}(p\alpha_p)^{-1/(p-1)}, \tag{6.2.22}$$

and $t_x a_{q_0} \leq 1$ is equivalent to $x \leq q_0\Gamma(2-p)$ (recall (6.2.2)). We then have

$$t_x x - \alpha_p t_x^p = \gamma_p\, x^{p/(p-1)} \text{ and } q_0(e^{t_x} - t_x - 1) = \gamma_p\,\varepsilon_p(x) x^{p/(p-1)}, \tag{6.2.23}$$

which concludes the proof of Inequality (a). The Inequality (b) follows directly by putting $t = a_{q_0}^{-1}$ in the right-hand side of (6.2.21). □

References

[1] S. Boucheron, G. Lugosi, and P. Massart. *Concentration Inequalities: A Nonasymptotic Theory of Independence*. OUP Oxford, 2013.

[2] E. Rio. Inégalités exponentielles pour les processus empiriques. *Comptes Rendus de l'Académie des Sciences - Series I - Mathematics*, 330(7):597 – 600, 2000.

[3] E. Rio. Inégalités de concentration pour les processus empiriques de classes de parties. *Probab. Theory Related Fields*, 119(2):163–175, 2001.

[4] E. Rio. Une inégalité de Bennett pour les maxima de processus empiriques. *Ann. Inst. H. Poincaré Probab. Statist.*, 38(6):1053–1057, 2002. En l'honneur de J. Bretagnolle, D. Dacunha-Castelle, I. Ibragimov.

École doctorale
de mathématiques
Hadamard (EDMH)

Titre : Inégalités de concentration pour des fonctions de variables aléatoires indépendantes

Mots Clefs : Inégalités de concentration, Indépendance, Inégalités de comparaison, Martingales, Processus empiriques.

Résumé : Cette thèse porte sur l'étude de la concentration autour de la moyenne de fonctions de variables aléatoires indépendantes à l'aide de techniques de martingales et d'inégalités de comparaison.
Dans une première partie, nous prouvons des inégalités de comparaison pour des fonctions générales séparément convexes de variables aléatoires indépendantes non nécessairement bornées. Ces résultats sont établis à partir de nouvelles inégalités de comparaison dans des classes de fonctions convexes (contenant, en particulier, les fonctions exponentielles croissantes) pour des variables aléatoires réelles uniquement dominées stochastiquement.
Dans la seconde partie, nous nous intéressons aux suprema de processus empiriques associés à des observations i.i.d. Le point clé de cette partie est un résultat d'échangeabilité des variables. Nous montrons d'abord des inégalités de type Fuk-Nagaev avec constantes explicites lorsque les fonctions de la classe ne sont pas bornées. Ensuite, nous prouvons de nouvelles inégalités de déviations avec une meilleure fonction de taux dans les bandes de grandes déviations dans le cas des classes de fonctions uniformément bornées. Nous donnons également des inégalités de comparaison de moments généralisés dans les cas uniformément borné et uniformément majoré. Enfin, les résultats de la première partie nous permettent d'obtenir une inégalité de concentration lorsque les fonctions de la classe n'ont plus de variance.

Title : Concentration inequalities for functions of independent random variables

Keys words : Concentration inequalities, Independent data, Comparison inequalities, Martingales, Empirical processes.

Abstract : This thesis deals with concentration properties around the mean of functions of independent random variables using martingales techniques and comparison inequalities.
In the first part, we prove comparison inequalities for general separately convex functions of independent and non necessarily bounded random variables. These results are based on new comparison inequalities in convex classes of functions (including, in particular, the increasing exponential functions) for real-valued random variables which are only stochastically dominated.
In the second part, we are interested in suprema of empirical processes associated to i.i.d. random variables. The key point of this part is a result of exchangeability of variables. We first give Fuk-Nagaev type inequalities with explicit constants when the functions of the considered class are unbounded. Next, we provide new deviation inequalities with an improved rate function in the large deviations bandwidth in the case of classes of uniformly bounded functions. We also provide generalized moment comparison inequalities in uniformly bounded and uniformly bounded from above cases. Finally, results from the first part allow us to prove a concentration inequality when the functions of the class have an infinite variance.