Víctor H. de la Peña Evarist Giné

Decoupling

From Dependence to Independence

Randomly Stopped Processes
U-Statistics and Processes
Martingales and Beyond

Springer

Preface

This book presents the theory and several applications of the decoupling principle, which provides a general approach for handling complex problems involving dependent variables. Its main tools consist of inequalities used for breaking (decoupling) the dependence structure in a broad class of problems by introducing enough independence so that they can be analyzed by means of standard tools from the theory of independent random variables.

Since decoupling reduces problems on dependent variables to problems on related (conditionally) independent variables, we begin with the presentation of a series of results on sums of independent random variables and (infinite-dimensional) vectors, which will be useful for analyzing the decoupled problems and which at the same time are tools in developing the decoupling inequalities. These include several recent definitive results, such as an extension of Lévy's maximal inequalities to independent and identically distributed but not necessarily symmetric random vectors, the Khinchin–Kahane inequality (Khinchin for random vectors) with best constants, and sharp decompositions of the L_p norm of a sum of independent random variables into functions that depend on their marginals only. A consequence of the latter consists of the first decoupling result we present, namely, comparing the L_p norms of sums of arbitrary positive random variables or of martingale differences with the L_p norms of sums of independent random variables with the same (one-dimensional) marginal distributions. With a few subjects, such as Hoffmann-Jørgensen's inequality, we compromise between sharpness and expediency and take a middle, practical road.

Concerning decoupling itself, we choose to introduce it by developing in great detail two of the main areas where it has been most successfully applied. These are 1) randomly stopped sums of independent random variables and processes with independent increments and 2) U-statistics and U-processes. There are two main reasons for starting with these "particular cases" rather than with the general theory: on the one hand, these examples motivate very clearly the general theory and, on the other hand, the general theory does not and cannot imply the strongest results in these two important areas.

The effect of decoupling on randomly stopped sums of independent random variables (and processes with independent increments) consists in creating independence between the stopping time and the variables so that the stopped sum can be treated conditionally as a sum of a fixed number of independent random variables. This is done for Banach space valued variables (sums) and processes. These results constitute striking generalizations of Wald's equation. A special case of the result on stopped processes consists of an extension of the Burkholder–Gundy inequality for randomly stopped Brownian motion to Banach valued processes with independent increments. An advantage of having these results in Banach spaces is that they apply also to some real valued processes that do not have independent increments, such as Bessel processes, by realizing them as functionals of Banach valued independent increments processes. Another advantage is that the constants involved in the approximations are independent of dimension. The decoupling results are then applied to the study of the first passage time for the absolute value (or the norm, in the Banach case) of a process with independent increments by developing a natural, yet surprising connection with boundary crossing by non-random functions, with applications that highlight the relevance of working in a Banach setting.

An important area of applications of the decoupling principle is the theory of U-statistics. Such statistics arise in the definition of unbiased estimators, including the sample variance of a sequence of i.i.d. random variables, and as higher-order terms in von Mises expansions of smooth statistical functionals. A U-statistic is the average value over the sample $X_1, \ldots, X_n$ of a function of several variables h; so, it involves a multiple sum of $m!\binom{n}{m}$ terms $h(X_{i_1}, \ldots, X_{i_m})$. Decoupling reduces the U-statistic to an average of terms of the form $h(X^1_{i_1}, \ldots, X^m_{i_m})$, where each entry of h (say, the jth entry) is filled up with the terms of a different, independent copy of the original sequence of random variables (say, $\{X^j_i : i = 1, \ldots, n\}$). This produces enough independence so that this average, conditioned on all but one of the independent sequences, becomes a sum of independent random variables. Decoupling is very different from Hoeffding's decompositions and constitutes an additional extremely powerful tool for analyzing U-statistics and processes. For instance, if the kernel has conditional mean zero, then decoupling allows for symmetrization and randomization, which are tools used to freeze the variables involved and/or reduce the problems to ones involving a weighted sum of Bernoulli random variables. Decoupling has played a central role in recent advances in the asymptotic theory of U-statistics, and has produced, among others, optimal results on the central limit theorem and very sharp results on the law of the iterated logarithm and on exponential inequalities. It has had even a more pivotal role in the development of the theory of U-processes, the analogue of empirical processes for U-statistics. We present a rather complete account of the asymptotic theory of U-statistics and U-processes, as well as a few statistical applications of the latter, e.g., to multidimensional M-estimators (including analogues of the sample median in several dimensions), and to the analysis of truncated data. As part of this study, we give a unified account of the construction of the chaos decomposition of the L_2 space of a Gaussian process and the proof of the central limit theorem for degenerate U-statistics.

In fact, we consider decoupling of a generalized form of U-statistics, with the kernel h varying with the multiindex $(i_1, \ldots, i_m)$. In this generality, the results presented apply as well to multilinear forms in independent random variables, which constitute one of the first objects to which decoupling was applied, motivated by multiple stochastic integration. A generalization of another historically important decoupling result comparing tail probabilities of Gaussian polynomials also follows in a straightforward way by combining decoupling of U-statistics with the central limit theorem.

The latter part of the book is devoted to the general theory of decoupling. More specifically, consider an arbitrary sequence $\{d_i\}$ of real random variables and let $\mathcal{F}_i$ be an increasing sequence of σ-fields to which it is adapted (we can take $\mathcal{F}_i$ to be the σ-field generated by $d_1, \ldots, d_i$). Let $\mathcal{G}$ be another σ-field contained in $\mathcal{F}_\infty$. Then, a sequence of random variables $\{e_i\}$ is a decoupled version of $\{d_i\}$ (with respect to $\{\mathcal{F}_i\}$ and $\mathcal{G}$), if i) $\mathcal{L}(e_i|\mathcal{F}_{i-1}) = \mathcal{L}(d_i|\mathcal{F}_{i-1})$, ii) the sequence $\{e_i\}$ is conditionally independent given $\mathcal{G}$, and iii) $\mathcal{L}(e_i|\mathcal{F}_{i-1}) = \mathcal{L}(e_i|\mathcal{G})$. Condition i) indicates proximity between the two sequences and allows for comparison of some of their characteristics such as moments of sums or maxima, etc. Sequences related by condition i) are said to be tangent to each other. Condition ii) expresses the fact that the sequence $\{e_i\}$ is more independent than the sequence $\{d_i\}$, and iii) allows for transfer to the original sequence $\{d_i\}$ of properties of the sequence $\{e_i\}$ related to its conditional independence given $\mathcal{G}$. Conditions ii) and iii) together are known as the C.I. (conditional independence) condition. Then, a general decoupling result is simply an inequality relating $\mathbb{E}\Phi(\sum d_i)$ and $\mathbb{E}\Phi(\sum e_i)$, where $\Phi(x)$ could be $|x|^p$, $\exp|x|^\alpha$, or $I_{x>t}$, in order to compare moments, exponential moments, or even the distributions of the sums; also, sums can be replaced by other functionals, such as maxima. It is important to stress the fact that decoupled sequences always exist and therefore, decoupling inequalities in this general context have a broad appeal as they always apply (the drawback is that they are not always useful). Moment inequalities hold in great generality and exponential moment inequalities hold also quite generally, but tail probability inequalities, which do hold in the case of Banach space valued U-statistics, do not hold in general. There are other types of inequalities such as, e.g., comparison of weak moments.

There is a very close connection between decoupling inequalities in this general setting and martingale inequalities. For instance, it can be shown that the square function martingale inequality of Burkholder–Davis–Gundy is equivalent to a decoupling inequality for martingales. There is also interplay between martingale and decoupling inequalities: for example, the Burkholder–Rosenthal inequality for martingale differences implies a decoupling inequality, and a sharp decoupling inequality paired with a sharp version of Rosenthal's inequality for sums of independent variables implies the Burkholder–Rosenthal inequality with best constants. Also, it is possible to give unified proofs of martingale inequalities and decoupling inequalities for conditionally symmetric sequences. These relationships are explored in depth in the chapters on decoupling of tangent sequences.

We also present the principle of conditioning, which is a general method for obtaining limit (in distribution and almost everywhere) results for sums of dependent variables based on analogous results for sums of independent variables. As applications we give a proof of the Brown–Eagleson central limit theorem for martingales by applying the principle of conditioning along with the Lindeberg–Feller central limit theorem for sums of independent random variables. As another consequence of this result we provide a proof of the central limit theorem for a sequence of (arbitrarily dependent) two by two tables. This result is relevant in the theory of biostatistics and provides a situation in which martingale methods do not seem to apply but the decoupling approach succeeds. Other applications of the general theory of decoupling that we present in detail include a general method for extending exponential inequalities for sums of independent variables to the ratio of a martingale over its conditional variance, an extension of Wald's equation to U-statistics, estimation of moments of randomly stopped U-statistics and an extension to U-statistics of Anscombe's theorem, convergence of moments included.

The decoupling approach to handling problems with dependent random variables can be traced back to a result of Burkholder and McConnell included in Burkholder (1983) which represents a step in extending the theory of martingales to Banach spaces. Therefore, it can be said that decoupling (for tangent sequences) was born as a natural continuation to the martingale approach in order to handle problems that traditionally could not be handled by means of martingale tools. A typical inequality for martingales due to Burkholder and Gundy compares a martingale to its square function, which is the square root of the sum of squares of its martingale difference sequence. This in effect transforms a problem involving martingales into one involving sums of non-negative variables and provides sufficient advantage in developing solutions to the problem in case. The idea of replacing the square function of a martingale by a decoupled (conditionally independent) version of the martingale was proposed in order to avoid problems with the definition of a square function in a Banach space. The first general decoupling inequality for tangent sequences was obtained by Zinn (1985) and extended by Hitczenko (1988).

A turning point in the theory of decoupling for tangent sequences has been Kwapień and Woyczynski (1991) (available as a preprint in 1986). It is shown in this paper that one can always obtain a decoupled tangent sequence to any adapted sequence, hence making general decoupling inequalities widely applicable. This work also develops the theory of semimartingale integrals via decoupling inequalities. Decoupling has been quite effective in treating problems involving U-statistics, multilinear forms and randomly stopped sums of independent random variables. Decoupling of multilinear forms was first considered by McConnell and Taqqu (1986) with a view towards the development of a theory of multiple stochastic integration. Their article, which precedes Zinn's and Kwapień and Woyczynski's, provided great impetus to the theory. Concerning randomly stopped sums, Klass (1988, 1990) obtained definitive decoupling results for variables in general Banach spaces. Kwapień and Woyczynski (1992) contains the development of the theory up to that point, including several L_p and tail probability inequalities, and uses decoupling to develop a general theory of stochastic integration. Building upon Kwapień's (1987) extension of McConnell and Taqqu's result, de la Peña (1992) further extended decoupling to a general class of random variables that contain both multilinear forms and U-statistics with values in general Banach spaces, hence providing a springboard to a wealth of results, initiated by Arcones and Giné (1993), on the general theory of U-statistics and more generally U-processes, the latter introduced by Nolan and Pollard (1987). Kwapień (1987) and Kwapień and Woyczynski (1992) proved the first tail probability decoupling inequalities for quadratic forms and multilinear forms of independent random variables. Giné and Zinn (1994) obtained a decoupling and symmetrization inequality for U-statistics. The definitive decoupling result along this line of work is de la Peña and Montgomery-Smith (1994), which provides a tail probability decoupling inequality for generalized U-statistics. Concerning the general theory of decoupling, recent developments include work of de la Peña (1994), with the first general exponential decoupling inequality, and of Hitczenko (1994), who extended this result by providing L_p inequalities with constants independent of p. A more detailed account of the history of the developments of decoupling can be found at the end of each chapter.

This book is addressed to researchers in Probability and Statistics and to advanced graduate students. Thus, the exposition is at the level of a second graduate course. For instance, we do not include a proof of Doob's maximal inequality, but we do include one for the Burkholder–Davis–Gundy inequality. This text contains as well a self-contained section on weak convergence of processes, sufficient for the study of U-processes. Except for relying on material from standard first year graduate courses, we only occasionally refer the reader to material not presented in this book. We have successfully incorporated some of the material from this book in our first year graduate probability courses, including Levy's and Hoffmann-Jørgensen's inequalities, the development of Wald's equations, exponential inequalities, some of the decoupling inequalities for U-statistics, which require only basic facts about conditional expectation and conditional Jensen's inequality, and several applications to the asymptotic theory of U-statistics.

The content of the book divides naturally into four parts. 1) Chapter 1, on sums of independent random variables; 2) Chapter 2, on randomly stopped sums and processes; 3) Chapters 3–4–5 on U-statistics and U-processes; and 4) Chapters 6–7–8 on the general theory of decoupling, with applications. The last three parts can be read independently of each other. On the other hand, the material in Chapter 1 is used in each of the other parts, occassionally in Chapters 2, 6, 7, and 8, but more often in Chapters 3, 4, and 5.

In the process of writing this book we benefited from past and present discussions with several people, including D. Alemayehu, M. Arcones, D. Burkholder, R. Dudley, N. Eisenbaum, J. Engel, P. Hitczenko, Z. Jin, M. Klass, S. Kou, S. Kwapień, R. Latała, S. Montgomery-Smith, D. Pollard, J. Szulga, J. Wellner, M. Yang, C. H. Zhang, and J. Zinn.

We would like to thank Doodmatie Kalicharan for excellent typing of part of this book.

New York
Storrs
April 1998

Víctor Hugo de la Peña Díaz Infante
Evarist Giné i Masdéu

Contents

1

Sums of Independent Random Variables

The theory of decoupling aims at reducing the level of dependence in certain problems by means of inequalities that compare the original sequence to one involving independent random variables. It is therefore important to have information on results dealing with functionals of independent random variables.

In this chapter we collect several types of results on sums of independent random variables that will be used throughout. We consider aspects of estimation of tail probabilities and moments that are relevant to the theory of decoupling and develop them to the extent needed, and, in a few instances, a little more.

We begin with the classical Lévy maximal inequalities, bounding the tail probabilities of the maximum of the norm of a sum of independent symmetric random vectors by the tail probabilities of the norm of the last sum, that is, the reflection principle for symmetric random walk extended to random variables taking values in a Banach space. Then, we also present analogous maximal inequalities for sums of arbitrary independent identically distributed random vectors. The proofs in the Banach space case are not more difficult than for the real case.

A way to prove integrability for (classes of) random variables is to obtain bounds for tail probabilities in terms of the squares of these same probabilities at lower levels. This is illustrated by the Hoffmann-Jørgensen type inequalities that we present in Section 2, which bound the pth moment of a sum of independent centered random vectors by a constant times the same moment of their maximum plus the pth power of a quantile. They are important as a means of upgrading stochastic boundedness (or weak convergence) of sequences of variables to uniform boundedness (or to convergence) of their moments. This very useful type of inequality originates with Kolmogorov's converse to his maximal inequality. As an application, we present a kind of reversed Jensen's inequality for exponentials. Other instances of the use of these inequalities can be found in subsequent chapters.

Next, we come to estimation of moments, starting with Khinchin's inequalities (Section 3). In their crudest form these inequalities assert that on the span of a Rademacher sequence all the L_p norms are equivalent. The inequality comparing the L_1 and L_2 norms is proved here for Rademacher linear combinations with coefficients in a Banach space (Khinchin–Kahane inequality) and with best constant. Extension to all moments and to Rademacher chaos is done in Chapter 3. This will be a basic ingredient in the asymptotic theory of U-statistics.

Finally, we consider the question of finding two-sided bounds for the L_p norm of a sum of independent random variables in terms of quantities that involve only one-dimensional integrals with respect to the probability laws of the individual summands. This is the subject of the last two sections, where several approaches are developed. Hoffmann-Jørgensen's inequality together with a quite precise estimate of the moments of the maximum of independent variables is used in one of the approaches (which carries to infinite dimensions), whereas the L function bounds, which constitute the most precise approach, do imply Rosenthal's and Hoffmann-Jørgensen's inequalities in $\mathbb{R}$ with essentially best constants. The K function approach, which was chronologically the first, is also briefly discussed. We present three applications of these inequalities. One, already mentioned, is Rosenthal's and Hoffmann-Jørgensen's inequalities for real variables with constants of the best order. Another, computes, up to multiplicative constants independent of p, the L_p norm of linear combinations of Rademacher variables. The third compares moments of sums of arbitrary positive random variables and martingales to moments of sums of independent positive and/or centered variables with the same individual distributions; these inequalities constitute an example of decoupling inequalities.

The developments just described require certain lemmas on truncation, randomization, etc., that are elementary but quite useful.

1.1 Lévy-Type Maximal Inequalities

This section is devoted to the extension of the classical Lévy inequalities for sums of independent symmetric random vectors to sums of not necessarily symmetric, but i.i.d., random vectors, possibly with different constants. These inequalities hold in great generality but, in order to avoid measurability considerations we assume the variables take values in a *separable* Banach space B.

For completeness sake, we begin with Lévy's inequalities, that are used all over this text.

THEOREM 1.1.1. *Let X_i, $1 \le i \le n$, be independent symmetric B-valued random variables. Then, for every $t > 0$,*

$$\Pr\Big\{\max_{1\le k\le n}\Big\|\sum_{i=1}^{k} X_i\Big\| > t\Big\} \le 2\Pr\Big\{\Big\|\sum_{i=1}^{n} X_i\Big\| > t\Big\} \tag{1.1.1}$$

and

$$\Pr\Big\{\max_{1\le k\le n}\|X_k\| > t\Big\} \le 2\Pr\Big\{\Big\|\sum_{i=1}^{n} X_i\Big\| > t\Big\}. \tag{1.1.2}$$

In particular,

$$\mathbb{E}\Big(\max_{1\le k\le n}\Big\|\sum_{i=1}^{k} X_i\Big\|\Big)^p \le 2\mathbb{E}\Big\|\sum_{i=1}^{n} X_i\Big\|^p, \qquad \mathbb{E}\Big(\max_{1\le k\le n}\|X_k\|\Big)^p \le 2\mathbb{E}\Big\|\sum_{i=1}^{n} X_i\Big\|^p$$

for all $p > 0$.

PROOF. We set $S_k = \sum_{i=1}^{k} X_i$, $k = 1, \dots, n$. The sets

$$A_k := \big\{\|S_i\| \le t \text{ for } 1 \le i \le k-1,\ \|S_k\| > t\big\}$$

are disjoint and

$$\Big\{\max_{1\le k\le n}\|S_k\| > t\Big\} = \cup_{k=1}^{n} A_k.$$

(A_k is the event "the random walk S_i exceeds the level t for the first time at time k.") For each $k \le n$ we define

$$S_n^{(k)} := S_k - X_{k+1} - \cdots - X_n$$

and note that, by symmetry and independence, the joint probability law of the n variables $(X_1, \dots, X_n)$ is the same as that of $(X_1, \dots, X_k, -X_{k+1}, \dots, -X_n)$, so that S_n and $S_n^{(k)}$ both have the same law. On the one hand, we obviously have that

$$\Pr\big[A_k \cap \{\|S_n\| > t\}\big] = \Pr\big[A_k \cap \{\|S_n^{(k)}\| > t\}\big],$$

and on the other hand,

$$A_k = \big[A_k \cap \{\|S_n\| > t\}\big] \cup \big[A_k \cap \{\|S_n^{(k)}\| > t\}\big]$$

since otherwise there would exist $\omega \in A_k$ such that

$$2\|S_k(\omega)\| = \|S_n(\omega) + S_n^{(k)}(\omega)\| \le 2t,$$

a contradiction with the definition of A_k. The last two identities imply that

$$\Pr A_k \le 2\Pr\big[A_k \cap \{\|S_n\| > t\}\big],\ k = 1, \dots, n,$$

and therefore,

$$\Pr\Big\{\max_{1\le k\le n}\|S_k\|>t\Big\}=\sum_{k=1}^{n}\Pr A_k\le 2\sum_{k=1}^{n}\Pr\big[A_k\cap\{\|S_n\|>t\}\big]$$
$$\le 2\Pr\big\{\|S_n\|>t\big\},$$

which gives inequality (1.1.1). The second inequality is proved in the same way if we redefine A_k as

$$A_k := \{\|X_i\|\le t \text{ for } 1\le i\le k-1,\ \|X_k\|>t\}$$

and $S_n^{(k)}$ as

$$S_n^{(k)} := -X_1-\cdots-X_{k-1}+X_k-X_{k+1}-\cdots-X_n.$$

The statements about expected values follow from (1.1.1) and (1.1.2) by integation by parts ($\int|\xi|^p dP = p\int t^{p-1}\Pr\{|\xi|>t\}dt$). □

If the random vectors are not symmetric, we have the following weaker inequality:

PROPOSITION 1.1.2. *Let X_i, $i\le n$, be independent B-valued random variables. Then, for all $t\ge 0$,*

$$\Pr\Big\{\max_{1\le k\le n}\Big\|\sum_{i=1}^{k}X_i\Big\|>t\Big\}\le 3\max_{k\le n}\Pr\Big\{\Big\|\sum_{i=1}^{k}X_i\Big\|>\frac{t}{3}\Big\}. \tag{1.1.3}$$

PROOF. Almost as in the previous proof, we define, for all $u, v\ge 0$ and $1\le k\le n$, $A_k=\big\{\|S_i\|\le u+v \text{ for } i<k, \text{ and } \|S_k\|>u+v\big\}$. The sets A_k are disjoint and their union is $\big\{\max_{1\le k\le n}\|S_k\|>u+v\big\}$. Therefore,

$$\begin{aligned}\Pr\{\|S_n\|>u\} &\ge \Pr\Big\{\|S_n\|>u,\ \max_{1\le k\le n}\|S_k\|>u+v\Big\}\\ &\ge \sum_{k=1}^{n}\Pr\big\{A_k\cap\{\|S_n-S_k\|\le v\}\big\}\\ &= \sum_{k=1}^{n}\Pr\{A_k\}\Pr\{\|S_n-S_k\|\le v\}\\ &\ge \Big[1-\max_{k\le n}\Pr\{\|S_n-S_k\|>v\}\Big]\Pr\Big\{\max_{1\le k\le n}\|S_k\|>u+v\Big\}.\end{aligned}$$

This is the well-known Lévy–Ottaviani inequality. Taking $u=t/3$ and $v=2t/3$ in this inequality gives

$$\begin{aligned}\Pr\Big\{\max_{1\le k\le n}\|S_k\|>t\Big\} &\le \frac{\Pr\{\|S_n\|>t/3\}}{1-\max_{k\le n}\Pr\{\|S_n-S_k\|>2t/3\}}\\ &\le \frac{\max_{k\le n}\Pr\{\|S_k\|>t/3\}}{1-2\max_{k\le n}\Pr\{\|S_k\|>t/3\}}.\end{aligned}$$

This proves inequality (1.1.3) if $\max_{k\le n}\Pr\{\|S_k\|>t/3\}<1/3$. But (1.1.3) is trivially satisfied otherwise. □

The above two inequalities are classical. Next we will extend Lévy's inequality to sequences of random vectors which are not necessarily symmetric, but which are i.i.d. The crucial point for these extension consists of the following theorem.

THEOREM 1.1.3. *If X_i, $i \in \mathbb{N}$, are independent identically distributed B-valued random variables, then, for $1 \le j \le k < \infty$,*

$$\Pr\Big\{\Big\|\sum_{i=1}^{j} X_i\Big\| > t\Big\} \le 3\Pr\Big\{\Big\|\sum_{i=1}^{k} X_i\Big\| > t/10\Big\}. \tag{1.1.4}$$

For $k = 2$ this theorem has a surprisingly simple proof: Let X, Y, Z, be i.i.d. Then,

$$\begin{aligned}\Pr\big\{\|X\| > t\big\} &= \Pr\big\{\|(X+Y)+(X+Z)-(Y+Z)\| > 2t\big\}\\ &\le 3\Pr\big\{\|X+Y\| > 2t/3\big\}.\end{aligned} \tag{1.1.4$'$}$$

The general case is more delicate. Its proof rests on the lemma that follows. First, an auxiliary definition: we say that x is a *t-concentration point* for the random vector X if $\Pr\{\|X - x\| > t\} < 1/3$.

LEMMA 1.1.4. *Let X_i, $i \in \mathbb{N}$, be i.i.d. random vectors. If $S_j = \sum_{i=1}^{j} X_i$ has a t-concentration point s_j for $1 \le j \le k$, then*

$$\|ks_j - js_k\| \le 3(k+j)t. \tag{1.1.5}$$

PROOF. First we observe that for X and Y arbitrary, if x is a t-concentration point for X, y is a t-concentration point for Y and z is a t-concentration point for $X + Y$ then

$$\|x + y - z\| \le 3t. \tag{1.1.6}$$

To see this just note

$$\begin{aligned}\Pr\{\|x+y-z\| > 3t\} &= \Pr\{\|X - x + Y - y - (X+Y-z)\| > 3t\}\\ &\le \Pr\{\|X-x\| > t\} + \Pr\{\|Y-y\| > t\} + \Pr\{\|X+Y-z\| > t\} < 1,\end{aligned}$$

so that $\Pr\{\|x+y-z\| \le 3t\} > 0$ and therefore (1.1.6) holds since x, y, and z are nonrandom. To prove the lemma we now proceed by induction. The lemma obviously holds for $j = k$, and (1.1.6) gives it for $k = 2$. Hence, it suffices to show that if the lemma holds for $1 \le j < r$ for all $r < k$, then it also holds for $1 \le j < k$. Now,

$$js_k - ks_j = js_k - (k-j)s_j - js_j = (js_{k-j} - (k-j)s_j) + j(s_k - s_j - s_{k-j}).$$

Hence, applying (1.1.6) and the induction hypothesis, we obtain

$$\begin{aligned}\|js_k - ks_j\| &\le \|js_{k-j} - (k-j)s_j\| + j\|s_k - s_{k-j} - s_j\|\\ &\le 3(k-j+j)t + 3jt = 3(k+j)t.\end{aligned}$$

□

PROOF OF THEOREM 1.1.3. We distinguish three cases. Suppose first that $\Pr\{\|S_{k-j}\| > 9t/10\} \le 1/3$. Then, independence of S_j and $S_k - S_j$, together with the fact that S_{k-j} and $S_k - S_j$ have the same distribution, give

$$\Pr\big\{\|S_j\| > t\big\} \le \Pr\big\{\|S_j\| > t,\ \|S_k - S_j\| \le 9t/10\big\} + \frac{1}{3}\Pr\big\{\|S_j\| > t\big\},$$

and therefore

$$\Pr\big\{\|S_j\| > t\big\} \le \frac{3}{2}\Pr\big\{\|S_j\| > t,\ \|S_k - S_j\| \le 9t/10\big\} \le \frac{3}{2}\Pr\big\{\|S_k\| > t/10\big\}.$$

Next we asume that there exists some $1 \le i \le k$ such that S_i does not have any $(t/10)$-concentration points. Then

$$\Pr\Big\{\|S_i + X_{i+1} + \cdots + X_k\| > \frac{t}{10} \,\Big|\, X_{i+1}, \ldots, X_k\Big\} \ge \frac{1}{3},$$

and therefore

$$\Pr\big\{\|S_k\| > t/10\big\} \ge \frac{1}{3} \ge \frac{1}{3}\Pr\{\|S_j\| > t\}$$

for all $1 \le j \le k$.

Suppose, finally, that $\Pr\big\{\|S_{k-j}\| > 9t/10\big\} > 1/3$ and that S_i has a $(t/10)$-concentration point s_i for all $1 \le i \le k$. Then

$$\{\|S_{k-j}\| > 9t/10\} \cap \{\|S_{k-j} - s_{k-j}\| \le t/10\} \ne \emptyset$$

and therefore, $\|s_{k-j}\| \ge 4t/5$. Hence, by Lemma 1.1.4,

$$\|s_k\| \ge \frac{k}{k-j}\|s_{k-j}\| - 3\frac{2k-j}{k-j}\frac{t}{10} \ge \frac{4kt}{5(k-j)} - \frac{6kt}{10(k-j)} > \frac{t}{5}.$$

This gives

$$\Pr\big\{\|S_k\| \ge t/10\big\} \ge \Pr\big\{\|S_k - s_k\| \le t/10\big\} \ge \frac{2}{3} \ge \frac{2}{3}\Pr\{\|S_j\| > t\}. \qquad \square$$

Combining Theorem 1.1.3 and Proposition 1.1.2 we readily obtain an analog of Lévy's inequality for sums of i.i.d random variables.

THEOREM 1.1.5. *If X_i, $i \in \mathbb{N}$, are independent identically distributed B-valued random variables, then, for $1 \le k \le n < \infty$ and all $t > 0$,*

$$\Pr\Big\{\max_{1\le k\le n} \Big\|\sum_{i=1}^{k} X_i\Big\| > t\Big\} \le 9\Pr\Big\{\Big\|\sum_{i=1}^{n} X_i\Big\| > \frac{t}{30}\Big\}. \qquad (1.1.7)$$

Theorems 1.1.3 and 1.1.5, which are quite recent (see the notes at the end of this chapter), constitute an important addition to the theory of sums of i.i.d. variables. They have consequences for decoupling, to be seen in Chapter 3. To immediately illustrate their usefulness, we end this section with an application to a *contraction principle* and to *randomization* for sums of i.i.d. random vectors. For this result (and only for it) we assume that the Banach space B where the variables lie is over the complex numbers.

COROLLARY 1.1.6. *There are universal constants $0 < c_1, c_2 < \infty$ such that if X_i, $i \in \mathbb{N}$, are either i.i.d. or independent and symmetric, and if α_i, $1 \le i \le n$, are any complex numbers such that $|\alpha_i| \le 1$, then, for all $n \in \mathbb{N}$ and $t \ge 0$,*

$$\Pr\Big\{ \Big\| \sum_{i=1}^{n} \alpha_i X_i \Big\| > t \Big\} \le c_1 \Pr\Big\{ \Big\| \sum_{i=1}^{n} X_i \Big\| > c_2 t \Big\}. \tag{1.1.8}$$

PROOF. Let k, $1 \le k \le n$, be fixed. The constants α_i, $1 \le i \le k$, can be assumed to be real and in decreasing order $-1 \le \alpha_k \le \cdots \le \alpha_1 \le 1$. So, we can write $\alpha_j = -1 + \sum_{i=j}^{k} \sigma_i$ for all $j \le k$, where $\sigma_i \ge 0$ for all $i \le k$ and $\sum_{i=1}^{k} \sigma_i \le 2$. Then,

$$\begin{aligned}
\Big\| \sum_{j=1}^{k} \alpha_j X_j \Big\| &= \Big\| \sum_{j=1}^{k} \Big(\sum_{i=j}^{k} \sigma_i - 1 \Big) X_j \Big\| \\
&= \Big\| \sum_{j=1}^{k} \sigma_j \sum_{i=1}^{j} X_i - \sum_{i=1}^{k} X_i \Big\| \\
&\le \Big(\sum_{j=1}^{k} \sigma_j \Big) \max_{1 \le j \le k} \Big\| \sum_{i=1}^{j} X_i \Big\| + \Big\| \sum_{i=1}^{k} X_i \Big\| \\
&\le 3 \max_{1 \le j \le k} \Big\| \sum_{i=1}^{j} X_i \Big\|.
\end{aligned}$$

Now the result follows from Theorem 1.1.5 in the i.i.d. case and from Theorem 1.1.1 in the symmetric case. □

REMARK 1.1.7. *Randomization.* It is clear (Fubini's theorem) that the constants α_i in the previous Corollary can be replaced by random variables θ_i uniformly bounded by 1 and independent from the sequence $\{X_j\}$.

REMARK 1.1.8. *Measurability.* The previous propositions can also be proved in the following more general context: The variables X_i are the coordinates of a product probability space $(S^{\mathbb{N}}, \mathcal{S}^{\mathbb{N}}, P^{\mathbb{N}})$, the norm $\|\cdot\|$ is the sup over a not necessarily countable family $\mathcal{F}$ of measurable functions on S and Pr is replaced by outer probability $\Pr^*$. The measurable outer envelope $\|\cdot\|_{\mathcal{F}}^*$ of the sup norm over $\mathcal{F}$ works essentially as a norm and $\Pr^*\{\|\cdot\|_{\mathcal{F}} > t\} = \Pr\{\|\cdot\|_{\mathcal{F}}^* > t\}$ for all $t \ge 0$, and this is essentially all that is needed to show that the above proofs work with only formal changes in the setting just decribed. A good available reference for calculus with outer probabilities is van der Vaart and Wellner (1996). We skip the details, but anticipate that this more general context is the natural one for U-processes, to be discussed in Chapter 4.