# Topic 1
# Introduction to Decoupling and Self-normalization

Victor H. de la Peña

Professor of Statistics, Columbia University

Artificial Intelligence Institute for Advances in Optimization
Georgia Institute of Technology 2024

# Topics Preview

# Topics Preview

# Motivation and History

- Decoupling and self-normalization constitute domains that have evolved in response to the imperative of expanding martingale techniques to encompass high-dimensional, infinite-dimensional, and intricate nonlinear dependency structures.

# Motivation and History

- Decoupling and self-normalization constitute domains that have evolved in response to the imperative of expanding martingale techniques to encompass high-dimensional, infinite-dimensional, and intricate nonlinear dependency structures.
- **Decoupling** equips us with techniques to proficiently address dependent variables by treating them akin to independent entities. Notably, it offers a natural framework for constructing precise exponential inequalities tailored for self-normalized (super-)martingales.

# Motivation and History

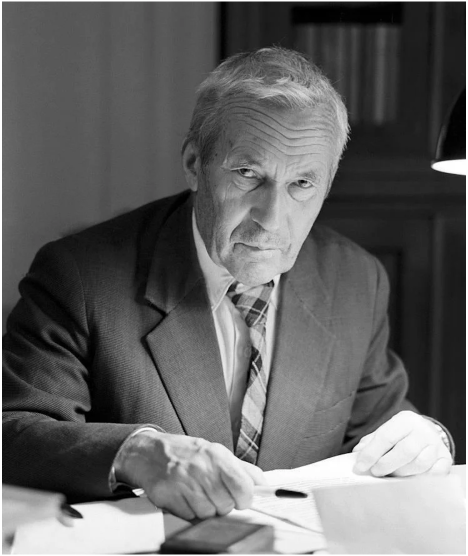
- Decoupling and self-normalization constitute domains that have evolved in response to the imperative of expanding martingale techniques to encompass high-dimensional, infinite-dimensional, and intricate nonlinear dependency structures.
- **Decoupling** equips us with techniques to proficiently address dependent variables by treating them akin to independent entities. Notably, it offers a natural framework for constructing precise exponential inequalities tailored for self-normalized (super-)martingales.
- Prominent illustrations of **self-normalized processes** include the t-statistic with dependent random variables, alongside (self-normalized) extensions of the KOLMOGOROV's law of the iterated logarithm

$$\limsup_{n\to\infty} \frac{|S_n|}{\sqrt{2n\sigma^2 loglogn}} = 1 \ \mathbb{P}\text{-almost everywhere}$$

# Topics Preview

1. Introduction
2. Decoupling
   - Complete Decoupling
   - Conditional Independent (Tangent) Decoupling
3. Self-Normalization
   - Canonical Assumptions
   - Pseudo-maximization (Method of Mixtures)
4. Martingale
5. Probabilistic Inequalities

# Decoupling



"BEHIND EVERY LIMIT THEOREM THERE IS AN INEQUALITY."

This quote has been attributed to **A.N. KOLMOGOROV**.

# Complete Decoupling

Let $\{d_i\}_{i=1}^n$ be a sequence of dependent random variables with $\mathbb{E}|d_i| < \infty$. Let $\{y_i\}_{i=1}^n$ be a sequence of independent variables where for each $i$, $d_i$ and $y_i$ have the same marginal distributions (denoted as $d_i \overset{\mathscr{L}}{=} y_i$ or $d_i \overset{d}{=} y_i$). Since $\mathbb{E}(d_i) = \mathbb{E}(y_i)$, linearity of expectations provides the first "complete decoupling" equality:

$$\mathbb{E} \sum_{i=1}^{n} d_i = \mathbb{E} \sum_{i=1}^{n} y_i \tag{1}$$

In **complete decoupling**, one compares $\mathbb{E}f(\sum d_i)$ to $\mathbb{E}f(\sum y_i)$ for more general functions $f(\cdot)$, taking the linear mapping $f(x) = cx$ as a special case. It remains possible to derive valuable inequalities based on specific assumptions.

Let the population $C$ consist of $N$ values $c_1, ..., c_N$ (e.g. a deck of cards, N=52), and let $d_1, ..., d_n$ ($n \leq N$) denote a random sample **without** replacement drawn from $C$, and $y_1, ..., y_n$ denote a random sample **with** replacement from $C$. The random variables $y_1, ..., y_n$ are i.i.d. and $d_i \stackrel{\mathscr{L}}{=} y_i$ for all $i$. HOEFFDING (1963, [9]) developed the following inequality:

$$\mathbb{E}\Phi(\sum d_i) \leq \mathbb{E}\Phi(\sum y_i). \tag{2}$$

for every continuous convex function $\Phi$.

De la Peña extended the assumption such that for nonnegative dependent r.v.s $(d_1, \ldots, d_n)$, such that for each $i$, $y_i$ and $d_i$ , have the same distribution and $\{y_i\}$ is a sequence of independent random variables.

**Theorem ( De La Peña, [2])**

*Suppose $\Phi$ is a concave nondecreasing function on $[0, \infty)$ such that $\Phi(0) = 0$ and $\Phi(x) > 0$ if $x > 0$, then $\exists\ C > 0$, not depending on anything, such that*

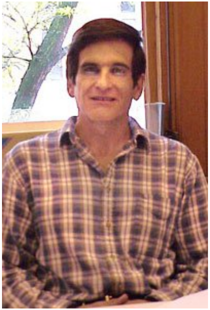
$$\mathbb{E}\Phi\left(\sum d_i\right) \leq C\mathbb{E}\Phi\left(\sum y_i\right). \tag{3}$$

**Theorem ( De La Peña, [2])**

*Suppose $\Phi$ is convex and increasing on $[0, \infty)$. Furthermore, $\exists\alpha > 0$ s.t. $\forall x > 0, c \geq 2$, $\Phi(cx) \leq c^\alpha\Phi(x)$. Then $\exists C_\alpha$ depending only on $\alpha$ s.t.*

$$\mathbb{E}\Phi(\sum d_i) \geq C_\alpha\mathbb{E}\Phi(\sum y_i)$$

# The Origin of Tangent Decoupling



The theory of martingale inequalities has been crucial in the development of modern probability theory. Recently it has been expanded widely through the introduction of the theory of conditionally independent (tangent) decoupling. This approach to decoupling can be traced back to a result of Burkholder and McConnell included in Burkholder [1] that represents a step in extending the theory of martingales to Banach spaces.

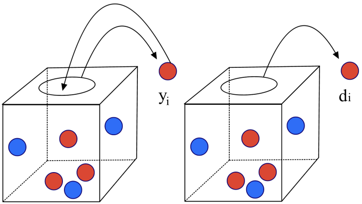
# Construction of Tangent Sequence

Specifically, given $\{d_i\}$, we can construct a tangent sequence w.r.t. $\mathcal{F}_i = \sigma(d_i, \ldots, d_i)$:

- First, we take $d_1$ and $y_1$ to be two independent copies of the same random mechanism.
- With $(d_1, .., d_{i-1}; y_1, .., y_{i-1})$, the i-th pair of variables $d_i$ and $y_i$ comes from conditionally independent copies of the same random mechanism given $\mathcal{F}_{i-1}$,

$$\begin{array}{ccccccccc} d_1 & \rightarrow & d_2 & \rightarrow & d_3 & \rightarrow & \ldots\ldots & \rightarrow & d_n \\ & \searrow & & \searrow & & \searrow & & \searrow & \\ y_1 & & y_2 & & y_3 & & \ldots\ldots & & y_n \end{array}$$

and $y_i$'s are conditionally independent w.r.t. $= \mathcal{F}_n$.



# Tangent Decoupling: M.G.F.

De la Peña [3] introduced the tangent decoupling equality, which compares the sum of dependent variables with the corresponding conditional independent sum.

## Theorem

*If $y_i$ is a decoupled $\mathcal{F}'$-tangent version of $d_i$, then for all r.v. $g > 0$ $\mathcal{G}-$measurable ($\mathcal{G} = \sigma(d_1, ..., d_n)$),*

$$\mathbb{E}\left[g \exp\left(\lambda \sum_{i=1}^{n} d_i\right)\right] \leq \sqrt{\mathbb{E}\left[g^2 \exp\left(2\lambda \sum_{i=1}^{n} y_i\right)\right]}. \tag{4}$$

When setting $g = 1$ almost surely, one gets the decoupling inequality for the moment-generating function.

**Remark**: the $y_i$'s are conditionally independent given $\mathcal{G}$.

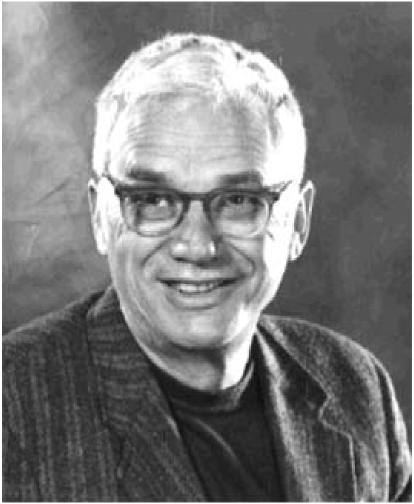
# Topics Preview

# Self-Normalized statistics

- A self-normalized statistic, literally, takes the form $\frac{A_n}{B_n}$ (resp., $\frac{A_t}{B_t}$ for continuous cases), where both $A_n$ and $B_n$ are functions of $X_1, \ldots, X_n$ (resp., $A_t, B_t$ the function of $(X_s)_{0 \le s \le t}$).
- Self-normalization can be traced back to the seminal work of **W.S. Gosset** in 1908 ([8]), which is considered a breakthrough in science. Notably, his Student t-statistic allowed statistical inference about the value of the mean of a (Gaussian) distribution without knowledge of the actual value of the variance, provided one has a random sample from the target population.

Figure: William S. Gosset, June 13, 1876 - October 16, 1937

# Joseph L. Doob, February 27, 1910 - June 7, 2004



Doob suggested the problem of self-normalization, in 1994.

# Canonical Assumptions

Assume that for a pair of random variables $A, B$ with $B > 0$,

$$\mathbb{E} \exp(\lambda A - \lambda^2 B^2/2) \leq 1, \tag{5}$$

holds for any one of the following schemes:

- for all real $\lambda$;
- for all $\lambda \geq 0$;
- for all $0 \leq \lambda < \lambda_0$, where $0 < \lambda_0 < \infty$.

This assumption is valid in a wide array of scenarios involving discrete-time and continuous-time stochastic processes, particularly in the context of (super)martingales.

### Example (Karatzas & Shreve [10])

Let $M_t$ be a continuous, square-integrable martingale, with $M_0 = 0$. Then $\exp\{\lambda M_t - \lambda^2 \langle M \rangle_t / 2\}$ is a supermartingale for all $\lambda \in \mathbb{R}$.

### Example (de la Peña [4])

Let $\{d_i\}$ be a sequence of variables adapted to an increasing sequence of $\sigma-$fields $\{\mathcal{F}_i\}$. Assume that the $d_i$'s are conditionally symmetric (i.e., $\mathcal{L}(d_i|\mathcal{F}_{i-1}) = \mathcal{L}(-d_i|\mathcal{F}_{i-1})$). Then $\exp(\lambda \sum_{i=1}^n d_i - \lambda^2 \sum_{i=1}^n d_i^2/2)$, $n \geq 1$, is a supermartingale with mean $\leq 1$, for all $\lambda \in \mathbb{R}$.

**Remark:** There is no integrability assumption made in this example.

# Pseudo-maximization

Recall the canonical assumption for a pair of random variables $A, B$ with $B > 0$,

$$\mathbb{E}\exp(\lambda A - \lambda^2 B^2/2) \leq 1,$$

If the "global maximizer" of $\lambda A - \lambda^2 B^2/2$, $\hat{\lambda} := \frac{A}{B^2}$ lies in the regime we are interested, and is of course, deterministic, then $\mathbb{E}\exp\left(A^2/2B^2\right) \leq 1$. And by CHEBYSHEV inequality, we have

$$\mathbb{P}\left(\frac{|A|}{|B|} > x\right) = \mathbb{P}\left(\frac{A^2}{2B^2} > \frac{x^2}{2}\right) \leq e^{-\frac{x^2}{2}}\mathbb{E}e^{\frac{A^2}{2B^2}} \leq e^{-\frac{x^2}{2}}.$$

Unfortunately, since $\frac{A}{B^2}$ is random, we need an alternative way, **pseudo-maximization**, an informal framework of which can be stated as:
(i) For $\lambda \in \Lambda$, a measurable set, we construct a probability measure of $\lambda$, with distribution function $F$ independent of $A$ and $B$.
(ii) Now by Fubini, we have that

$$1 \geq \mathbb{E} \exp\left(\lambda A - \frac{\lambda^2 B^2}{2}\right)$$
$$= \mathbb{E}\left[\int_{\mathbb{R}} \exp\left(\lambda A - \frac{\lambda^2 B^2}{2}\right) dF\right]$$
$$= \int_{\mathbb{R}} \mathbb{E}\left[\exp\left(\lambda A - \frac{\lambda^2 B^2}{2}\right)\right] dF.$$

One application of pseudo-maximization is to construct a Gaussian bound for $\frac{A}{\sqrt{B^2+\mathbb{E}^2 B}}$:

**Theorem (de la Peña et al, [5])**

*Let $A, B$ with $B > 0$ be random variables satisfying the canonical assumption*

$$\mathbb{E} \exp(\lambda A - \lambda^2 B^2/2) \leq 1,$$

*for all $\lambda \in \mathbb{R}$. Then*

$$\mathbb{P}\left(\frac{|A|}{\sqrt{B^2 + \mathbb{E}^2 B}} \geq x\right) \leq \sqrt{2} \exp\left(-\frac{x^2}{4}\right). \tag{6}$$

# Law of Iterated Logarithm Bound

Let $\{Y_n\}$ be independent, identically distributed random variables with means zero and variances $\sigma^2$. Let $S_n = Y_1 + ... + Y_n$. Then

$$\limsup_{n\to\infty} \frac{|S_n|}{\sqrt{2n\sigma^2 \log\log n}} = 1 \quad \text{a.s.}.$$

Another application of pseudo-maximization, under the following refinement of the canonical assumption, leads an LIL bound.

**Theorem**

*Assume that*

$$\{\exp\left(\lambda A_t - \lambda \frac{B_t^2}{2}\right), t \geq 0\}$$

*is a supermartingale with mean* $\leq 1$. *Then on the set* $\{\lim_{t\to\infty} B_t^2 = \infty\}$,

$$\limsup_{t\to\infty} \frac{A_t}{\sqrt{2B_t^2 \log\log B_t^2}} \leq 1.$$

As formalized in the previous example:

**Example (de la Peña [4])**

Let $\{d_i\}$ be a sequence of variables adapted to an increasing sequence of $\sigma-$fields $\{\mathcal{F}_i\}$. Assume that the $d_i$'s are conditionally symmetric (i.e., $\mathcal{L}(d_i|\mathcal{F}_{i-1}) = \mathcal{L}(-d_i|\mathcal{F}_{i-1})$). Then $\exp(\lambda \sum_{i=1}^n d_i - \lambda^2 \sum_{i=1}^n d_i^2/2)$, $n \geq 1$, is a supermartingale with mean $\leq 1$, for all $\lambda \in \mathbb{R}$.

We can get, on the set $\{\lim_{n\to\infty} \sum_{i=1}^n d_i^2 = \infty\}$, that

$$\limsup_{n\to\infty} \frac{\sum_{i=1}^n d_i}{\sqrt{2(\sum_{i=1}^n d_i^2)\log\log(\sum_{i=1}^n d_i^2)}} \leq 1,$$

a sharp extension of Kolmogorov's LIL without moment assumptions, which is also valid for i.i.d. centered Cauchy variables.

# Topics Preview

# Background and Definitions

Given a stochastic process $\{X_n\}_{n \in \mathbb{N}_0}$, as the time n increases, so does our knowledge about what has happened in the past, which can be modelled using filtration.

## Definition (Filtration (Discrete))

Given the measurable space $(\Omega, \mathcal{F})$, a sequence of $\sigma$-algebras $\mathcal{F}_1$, $\mathcal{F}_2$, ..., on $\Omega$ such that

$$\mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \ldots \subseteq \mathcal{F}$$

is called a filtration.

Here $\mathcal{F}_n$ represents our knowledge at time n. It contains all events such that at time n it is possible to decide whether A has occurred or not. As n increases, there will be more such events A, i.e. the family $\mathcal{F}_n$ representing our knowledge will become larger.

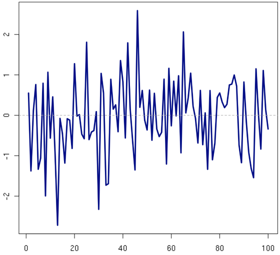
## Definition (Discrete-time Martingales)

Given the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, a sequence $M_1$, $M_2$, ..., of random variables is called a martingale (resp, supermartingale, submartingale) with respect to a filtration $\mathcal{F}_1$, $\mathcal{F}_2$,..., if

1) $M_n$ is integrable for each n = 1,2, ... ;

2) $M_1$, $M_2$, ... is adapted to $\mathcal{F}_1$, $\mathcal{F}_2$,...;

3) $\mathbb{E}(M_{n+1}|\mathcal{F}_n) = (\text{resp}, \ \leq, \ \geq) M_n$ a.s. for each n =1,2, ....




The concept of a martingale has its origin in gambling, namely, it describes a fair game of chance. Similarly, the notions of submartingale and supermartingale defined below are related to favourable and unfavourable games of chance.

## Definition (Filtration)

Given the measurable space $(\Omega, \mathcal{F})$, a family of $\sigma$-algebras $\mathcal{F}_t$ on $\Omega$, parametrized by $t \in T \subset \mathbb{R}$, is called a filtration if

$$\mathcal{F}_s \subseteq \mathcal{F}_t \subseteq \mathcal{F}$$

for any $s, t \in T$ such that $s \leq t$.

## Definition (Continuous-time Martingales)

Given the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, a stochastic process $\xi_t$ $(t \in T)$ is called a martingale (resp, supermartingale, submartingale) with respect to a filtration $\mathcal{F}_t$, if

1) $\xi_t$ is integrable for each $t \in T$;

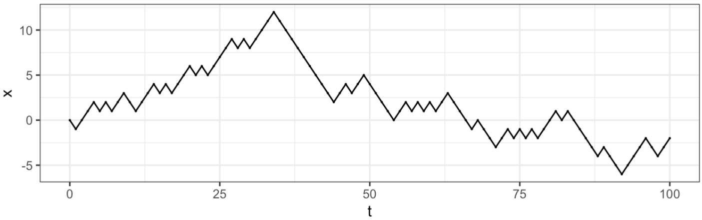
2) $\xi_t$ is adapted to $\mathcal{F}_t$ for each $t \in T$;

3) $\mathbb{E}(\xi_t | \mathcal{F}_s) = (\text{resp}, \ \leq, \ \geq) \xi_s$ a.s. for every $s, t \in T$ and $s \leq t$.

# Examples

## Example

Mean-corrected Sums of i.i.d. r.v.s & Simple Symmetric Random Walk
Let $X_1, X_2, \ldots$ be independent random variables each with mean $\mu$. Let $S_0 = 0$ and for $n > 0$ let $S_n$ be the partial sum $S_n = X_1 + \ldots + X_n$. Then $M_n = S_n - n\mu$ is a martingale with respect to $\mathcal{F}_n$, the information contained in $X_1, \ldots, X_n$.
If we now assume that $X_i$ is the Rademacher variable, then this stochastic process is called a **simple symmetric random walk**.



## Example

Brownian Motion The **Wiener process** (or **Brownian motion**) is a stochastic process $W(t)$ with values in $\mathbb{R}$ defined for $t \in [0, \infty)$ such that
1) $W(0) = 0$ a.s.;
2) the sample paths $t \mapsto W(t)$ are a.s. continuous;
3) for any finite sequence of times $0 < t_1 < ... < t_n$ and Borel sets $A_1, ..., A_n \subset \mathbb{R}$

$$\mathbb{P}(W(t_1) \in A_1, ..., W(t_n) \in A_n)$$
$$= \int_{A_1} ... \int_{A_n} p(t_1, 0, x_1) p(t_2 - t_1, x_1, x_2) ... p(t_n - t_{n-1}, x_{n-1}, x_n) dx_1 ... dx_n,$$

where

$$p(t, x, y) = \frac{1}{\sqrt{2\pi t}} e^{\frac{(x-y)^2}{2t}}$$

defined for any $x, y \in \mathbb{R}$ and $t > 0$, called the transition density.

# Topics Preview

# Some Basic Probabilistic Inequalities

Now let us get familiar with rudimentary probabilistic inequalities including exponential inequalities, decoupling inequalities, and martingale inequalities.

# MARKOV's Inequality

For any integrable random variable X (i.e. $\mathbb{E}X < \infty$), we have that, for any $a > 0$,

$$\mathbb{P}(|X| \geq a) \leq \frac{\mathbb{E}(|X|)}{a}. \tag{7}$$

# HŐLDER's inequality/ CAUCHY-SCHWARZ Inequality

For any square-integrable random variable X, Y (i.e. $\mathbb{E}X^2, \mathbb{E}Y^2 < \infty$), we have that

$$\mathbb{E}|XY| \leq \sqrt{\mathbb{E}X^2 \mathbb{E}Y^2}. \tag{8}$$

And I leave the random vector's version as an exercise.

# Jensen's Inequality

We assume X an integrable real-valued random variable and $\Phi$ a convex function. Then

$$\Phi(\mathbb{E}(X)) \le \mathbb{E}(\Phi(X)). \tag{9}$$

# BERNSTEIN's inequality

Let $\{x_i\}$ be a sequence of independent variables. Assume that $\mathbb{E}(x_i) = 0$ and $\mathbb{E}(x_i^2) = \sigma_i^2 < \infty$ and set $v_n^2 = \sum_{i=1}^{n} \sigma_i^2$. Furthermore, assume that there exists a constant $0 < c < \infty$ such that, $\mathbb{E}(|x_i|^k) \leq (k!\,/2)\sigma_i^2 c^{k-2}$ for all $k > 2$ (satisfied by subexponential random variables). Then for all $x > 0$.

$$\mathbb{P}\left(\sum_{i=1}^{n} x_i > x\right) \leq \exp\left(-\frac{x^2}{2(v_n^2 + cx)}\right).$$

# HOEFFING's Inequality for Sampling Without Replacement

Let the population $C$ consist of $N$ values $c_1, ..., c_N$ (e.g. a deck of cards, N=52), and let $d_1, ..., d_n$ $(n \leq N)$ denote a random sample **without** replacement drawn from $C$, and $y_1, ..., y_n$ denote a random sample **with** replacement from $C$. The random variables $y_1, ..., y_n$ are i.i.d. and $d_i \overset{\mathscr{L}}{=} y_i$ for all $i$. And we have

$$\mathbb{E}\left[\Phi(\sum_{i=1}^{n} d_i)\right] \leq \mathbb{E}\left[\Phi(\sum_{i=1}^{n} y_i)\right]. \tag{10}$$

# Doob's Maximal Inequalities

If $X_1, X_2, \ldots, X_n$ are martingale differences (i.e., $M_n := \sum_{i=1}^{n} X_n$ is a martingale with $M_0 = 0$), then for any $t > 0$,

$$\mathbb{P}\left(\max_{1\le k\le n} |M_k| \ge t\right) \le \frac{1}{t^2}\mathbb{E}[(M_n)^2]. \tag{11}$$

And for all $p > 1$,

$$\mathbb{E}\left(\max_{1\le k\le n} |M_k|^p\right) \le \left(\frac{p}{p-1}\right)^p \mathbb{E}\,|M_n|^p \tag{12}$$

We call it a day with the following important inequalities and one definition, which we will give further discussion on this Wednesday.

**Theorem (Burkholder-Davis-Gundy's Square Function Inequality)**

*If $X_1, X_2, \ldots, X_n$ are martingale differences, then for any $p \geq 1$,*

$$c_p E\left[\left(\sum_{i=1}^{n} X_i^2\right)^{\frac{p}{2}}\right] \leq E\left[\left(\max_{1\leq k\leq n}\left|\sum_{i=1}^{k} X_i\right|\right)^{p}\right] \leq C_p E\left[\left(\sum_{i=1}^{n} X_i^2\right)^{\frac{p}{2}}\right], \tag{13}$$

*where $c_p$ and $C_p$ are positive constants depending only on $p$.*

# KHINTCHINE inequality

Let $\epsilon_i, i = 1, ..., m$ be i.i.d. Rademacher variables with $\mathbb{P}(\epsilon_i = \pm 1) = 1/2$. Let $0 < p < \infty$ and let $x_1, ..., x_n \in \mathbb{C}$. Then

$$A_p \left( \sum_{i=1}^{n} |x_i|^2 \right)^{1/2} \leq \left( \mathbb{E} |\sum_{i=1}^{n} \epsilon_i x_i|^p \right)^{1/p} \leq B_p \left( \sum_{i=1}^{n} |x_i|^2 \right)^{1/2} \tag{14}$$

for some constants $A_p, B_p > 0$ depending only on $p$.

# K-function

## Definition (K-function, Klass [11])

Consider a nontrivial random variable $Y$. Then the K-function, $K_Y(x)$, is implicitly defined by the inverse of

$$g(x) = \frac{x^2}{\int_0^x \mathbb{E}|Y|I_{|Y|>u}du}. \tag{15}$$

**Exercise:** Please verify the equivalent definition of K-function, which is the unique solution of

$$K_Y(x)^2 = x\mathbb{E}[Y^2 \wedge (|Y|K_Y(x))] = x\mathbb{E}Y^2 I_{|Y|\le K_Y(x)} + xK_Y(x)\mathbb{E}|Y|I_{|Y|>K_Y(x)}. \tag{16}$$

You can find the proof in Section 1.4.3 of the book [7].

## Theorem (K-function Bounds)

*Consider $\{y_i\}_{i=1}^n$ a sequence of i.i.d. random centered variables such that $y_1 \sim Y$. Then we have*

$$0.67K_Y(n) \leq \mathbb{E}|\sum_{i=1}^n y_i| \leq 2K_Y(n). \tag{17}$$

# References I


[1] D. L. Burkholder. "A geometric condition that implies the existence of certain singular integrals of Banach-space-valued functions". In: *Conference on harmonic analysis in honor of Antoni Zygmund*. Vol. 1. 1983, pp. 270–286.

[2] V. H. de la Peña. "Bounds on the expectation of functions of martingales and sums of positive RVs in terms of norms of sums of independent random variables". In: *Proc. Amer. Math. Soc.* 108.1 (1990), pp. 233–239.

[3] V. H. de la Peña. "A bound on the moment generating function of a sum of dependent variables with an application to simple random sampling without replacement". In: *Annales de l'IHP Probabilités et statistiques*. Vol. 30. 2. 1994, pp. 197–211.

# References II


[4] V. H. de la Peña. “A general class of exponential inequalities for martingales and ratios”. In: *The Annals of Probability* 27.1 (1999), pp. 537–564.

[5] V. H. de la Peña, M. J. Klass, and T. L. Lai. “Self-normalized processes: exponential inequalities, moment bounds and iterated logarithm laws”. In: *The Annals of Probability* (2004).

[6] V. H. de la Peña, T. L. Lai, and Q. Shao. *Self-normalized processes: Limit theory and Statistical Applications*. Springer, 2009.

[7] V.H. de˙la Peña and E. Giné. *Decoupling: from dependence to independence*. Springer Science & Business Media, 2012.

[8] W. S. Gosset (Student). “The probable error of a mean”. In: *Biometrika* 6.1 (1908), pp. 1–25.

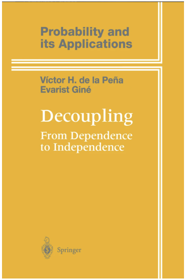
[9] W. Hoeffding. “Probability inequalities for sums of bounded random variables”. In: *J. Amer. Statist. Assoc.* 58.301 (1963), pp. 13–30.

# References III

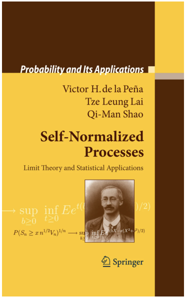
[10] I. Karatzas and S. Shreve. *Brownian motion and stochastic calculus*. Vol. 113. Springer Science & Business Media, 1991.

[11] Michael J Klass. “A method of approximating expectations of functions of sums of independent random variables”. In: *The Annals of Probability* (1981), pp. 413–428.

# Two Recommended Books




*Decoupling: From Dependence to Independence* (1999, [7])
Victor H. de la Peña
Evarist Giné




*Self-normalized processes: Limit theory and Statistical Applications* (2009, [6])
Victor H. de la Peña
Tze Leung Lai
Qi-Man Shao