Lecture 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 2023

Topics Preview

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

Topics Preview

1. Introduction

2. Decoupling
   - Complete Decoupling
   - Conditional Independent (Tangent) Decoupling

3. Self-Normalization
   - Canonical Assumptions
   - Pseudo-maximization (Method of Mixtures)

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

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

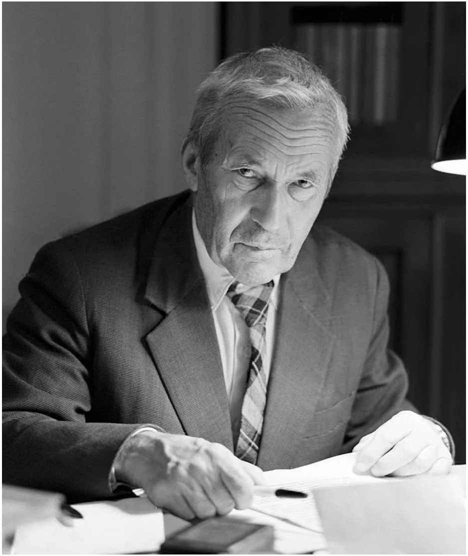
Topics Preview

1. Introduction

2. Decoupling
   - Complete Decoupling
   - Conditional Independent (Tangent) Decoupling

3. Self-Normalization
   - Canonical Assumptions
   - Pseudo-maximization (Method of Mixtures)

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 $d_i \stackrel{\mathscr{L}}{=} y_i$ or $d_i \stackrel{d}{=} y_i$). 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, [10]) developed the following inequality:

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

for every continuous convex function Φ.

De La Peña (1990, [3]), extended the assumption such that for nonnegative dependent r.v.s $(d_1, \ldots, d_n)$, each y_i is an i.i.d. copy of d_i, and $\{y_i\}$ themselves are independent as well. When $p \geq 2$, there exists a constant C_p depending on p only s.t.

$$\|\sum_{i=1}^{n} d_i\|_{L_p} \geq C_p \|\sum_{i=1}^{n} y_i\|_{L_p}. \tag{3}$$

And if Ψ is a nondecreasing concave function on $[0, \infty)$ such that $\Psi(0) = 0$ and $\Psi(x) > 0$ if $x > 0$, then there is a constant $C > 0$, not depending on anything, such that

$$\mathbb{E}\Psi\left(\sum_{i=1}^{n} d_i\right) \leq C\mathbb{E}\Psi\left(\sum_{i=1}^{n} y_i\right). \tag{4}$$

In 2023, Chollette et al [2] showed that this constant $C \leq \frac{e}{e-1} \approx 1.5820$, which holds even for $\Psi(0) \geq 0$.

In 2018, Makarychev and Sviridenko [12] developed the following **sharp** inequality.

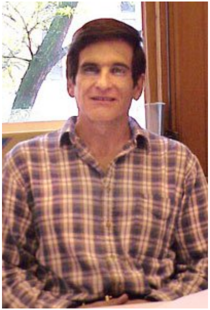
Theorem

Assume the $\{d_i\}_{i=1}^n$ is a sequence of arbitrarily nonnegative random variables. Let $\{y_i\}_{i=1}^n$ be independent random variables with $d_i \stackrel{\mathscr{L}}{=} y_i$ for all i. Let $\pi(1)$ be a Poisson random variable with mean 1, independent of $\{y_i\}$ and $\{d_i\}$. Then, for every convex ϕ and concave ψ,

$$\mathbb{E}\phi\left(\pi(1)\sum_{i=1}^n d_i\right) \geq \mathbb{E}\phi\left(\sum_{i=1}^n y_i\right), \mathbb{E}\psi\left(\pi(1)\sum_{i=1}^n d_i\right) \leq \mathbb{E}\psi\left(\sum_{i=1}^n y_i\right). \quad (5)$$

- This inequality has practical applications, approximating solutions to complex optimization problems. It relaxes integral programming (IP) to linear programming (LP) through randomized procedures. And the expected LP cost can bound the optimal IP cost.
- Notable examples encompass minimum energy-efficient routing and the construction of minimum-degree balanced spanning trees.

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 McConell included in Burkholder [1] that represents a step in extending the theory of martingales to Banach spaces.

Tangent Decoupling

Let $\{d_i\}$ and $\{y_i\}$ be two sequence of r.v.s adapted to σ-fields $\{\mathcal{F}_i'\}$. Then $\{d_i\}$ and $\{y_i\}$ are said to be **tangent** with respect to $\{\mathcal{F}_i'\}$ if, for all i,

$$\mathscr{L}(d_i|\mathcal{F}_{i-1}') = \mathscr{L}(y_i|\mathcal{F}_{i-1}'), \tag{6}$$

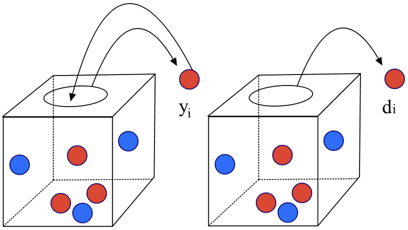
where $\mathscr{L}(d_i|\mathcal{F}_{i-1}')$ denotes the conditional distribution of d_i given $\mathcal{F}_{i-1}'$.

Construction of Tangent Sequence

Specifically, given $\{d_i\}$, we can construct a tangent sequence w.r.t. $\mathcal{F}'_i = \sigma(d_1, ..., d_i; y_1, ..., y_i)$:

- 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 i.i.d. copies of the same random mechanism given $\mathcal{F}'_{i-1}$.

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



Tangent Decoupling: M.G.F.

In 1994, de la Peña [4] 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 ($\mathcal{G} = \sigma(d_1, ..., d_n)$),

$$\mathbb{E}\left[g \exp\left(\lambda \sum_{i=1}^{n} d_i\right)\right] \leq \sqrt{\mathbb{E}\left[g^2 \exp\left(2\lambda \sum_{i=1}^{n} y_i\right)\right]}. \tag{7}$$

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 $\mathcal{G}$.

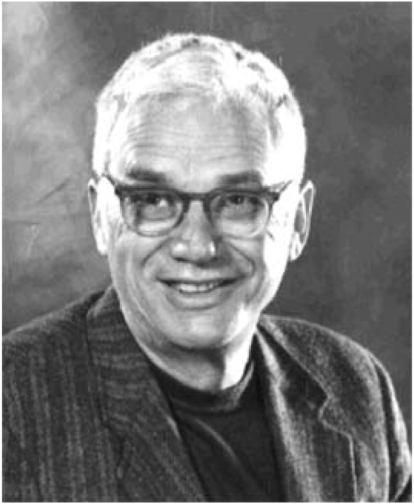
Topics Preview

Self-Normalized statistics

- 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 $X_1, ..., X_n$ (resp., A_t, B_t the function of $(X_s)_{0 \leq s \leq t}$).
- Self-normalization can be traced back to the seminal work of **W.S. Gosset** in 1908 ([9]), 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 $B > 0$,

$$\mathbb{E}\exp(\lambda A - \lambda^2 B^2/2) \le 1, \tag{8}$$

holds for any one of the following schemes:

- for all real λ;
- for all $\lambda \ge 0$;
- for all $0 \le \lambda < \lambda_0$, 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 (1991) [11])

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 (1999) [5])

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 ≤ 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 λ, with distribution function F independent of A and B.
(ii) Now by FUBINI, we have that

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

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, 2004 [6])

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

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

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

$$\mathbb{P}\left(\frac{|A|}{\sqrt{B^2+\mathbb{E}^2 B}} \ge x\right) \le \sqrt{2}\exp\left(-\frac{x^2}{4}\right). \tag{9}$$

Law of Iterated Logarithm Bound

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

Theorem

Assume that

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

is a supermartingale with mean ≤ 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 (1999) [5])

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 ≤ 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.

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] L. Chollete, V.H. de la Peña, and M.J. Klass. "The price of independence in a model with unknown dependence". In: *Mathematical Social Sciences* 123 (2023), pp. 51–58.

[3] 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.

References II

[4] 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.

[5] 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.

[6] V. H. de la Peña, M. J. Klass, and T. L. Lai. "Self-normalized processes: exponential inequalities, moment bounds and iterated logarithm laws". In: (2004).

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

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

References III

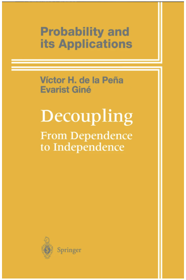
[9] W. S. Gosset (Student). "The probable error of a mean". In: *Biometrika* 6.1 (1908), pp. 1–25.

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

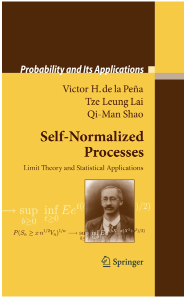
[11] I. Karatzas and S. Shreve. *Brownian motion and stochastic calculus*. Vol. 113. Springer Science & Business Media, 1991.

[12] K. Makarychev and M. Sviridenko. "Solving optimization problems with diseconomies of scale via decoupling". In: *J. ACM* 65.6 (2018), pp. 1–27.

Two Recommended Books



Decoupling: From Dependence to Independence (1999, [8])
Victor H. de la Peña
Evarist Giné



Self-normalized processes: Limit theory and Statistical Applications (2009, [7])
Victor H. de la Peña
Tze Leung Lai
Qi-Man Shao