Topic 2
Complete Decoupling

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

Introduction

Today we start our study of decoupling techniques. We begin by considering complete decoupling.

Let $\{d_i\}_{i=1}^n$ be a sequence of possibly dependent random variables with $\mathbb{E}|d_i| < \infty$. A basic question is: how large can the expectation of a sum of dependent variables be? If we know the expectations of the individual component, linearity of expectations gives the answer.

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

A natural question arises. How large can

$$\mathbb{E}\Phi(\sum_{i=1}^n d_i),$$

be for some nonlinear function Φ?

Complete Decoupling Equality

In **complete decoupling**, one compares $\mathbb{E}f(\sum d_i)$ to $\mathbb{E}f(\sum y_i)$ for more general functions $f(\cdot)$, where $\{y_i\}_{i=1}^n$ is 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 $d_i \stackrel{d}{=} y_i$). We can then rewrite the "complete decoupling" equality (1) as

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

And it remains possible to derive valuable inequalities based on specific assumptions.

Example: Construction of Independent Sequence

Given the dependent d_i's, how can we obtain the independent y_i's? One way of constructing the y_i's is by producing n independent copies of d_1, $d_2, ..., d_n$, in the following manner, where $d_i^{(l)}$ denotes the l-th copy of d_i:

$$
\begin{array}{c}
\mathbf{d_1^{(1)}}, d_2^{(1)}, d_3^{(1)}, \ldots \ldots\ldots, d_n^{(1)} \\
d_1^{(2)}, \mathbf{d_2^{(2)}}, d_3^{(2)}, \ldots \ldots\ldots, d_n^{(2)} \\
d_1^{(3)}, d_2^{(3)}, \mathbf{d_3^{(3)}}, \ldots \ldots\ldots, d_n^{(3)} \\
\vdots \\
d_1^{(n)}, d_2^{(n)}, d_3^{(n)}, \ldots \ldots\ldots, \mathbf{d_n^{(n)}}
\end{array}
$$

Let y_i be $d_i^{(i)}$ and we obtain a sequence of independent copies of $\{d_i\}_{i=1}^n$.

Topics Preview

Simplest Scenario

For $\{y_i\}$ a sequence of independent random variables with the identical distribution characterized by the cumulative distribution function (CDF) F, we are interested in how large can $\mathbb{E}|\sum_{i=1}^{n} y_i|^p$ be. And let us recall KHINTCHINE's inequality for one of the simplest situations, where $Y_i = \pm 1$ with probability $1/2$.

Theorem (Khintchine inequality)

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

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

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

In this case, we note that with $x_i = 1$ for all i (and hence $y_i = \epsilon_i$),

$$\mathbb{E}|\sum_{i=1}^{n} y_i|^p \approx n^{\frac{p}{2}}. \tag{3}$$

Particularly, if we let $p = 1$, then we have that $\mathbb{E}|\sum_{i=1}^{n} y_i| \approx \sqrt{n}$. However, for a general distribution F, we need to calculate the following integral

$$\mathbb{E}|\sum_{i=1}^{n} y_i|^p = \int_{\mathbb{R}^n} |\sum_{i=1}^{n} y_i|^p dF(y_1)...dF(y_n), \tag{4}$$

which is hard to compute. One tool to approximate the expectation is the **K-function**, introduced in the last lecture.

Definition (K-function, Klass [3])

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

$$g(x) = \frac{x^2}{\int_0^x \mathbb{E}|Y| I_{|Y|>u} du}. \tag{5}$$

And you may already verified the equivalent definition of K-function, which is the unique solution of

$$K_Y(x)^2 = x\mathbb{E}[Y^2 \wedge (|Y| K_Y(x))] = x\mathbb{E}Y^2 I_{|Y| \le K_Y(x)} + xK_Y(x)\mathbb{E}|Y| I_{|Y|>K_Y(x)}. \tag{6}$$

$K_Y(x)$ exists and is unique since the function $\mathbb{E}[Y^2/t^2 \wedge |Y|/t)]$ is strictly decreasing and continuous, with values ranging from 0 to ∞.

Now we consider $Y_n = (y_1, ..., y_n)$, where the variables are independent, and analogously let $K(Y_n)$ be the unique positive number k solving

$$\sum_{i=1}^{n} \mathbb{E} y_i^2 I_{|y_i| \leq k} + k \sum_{i=1}^{n} \mathbb{E}|y_i| I_{|y_i| > k} = \sum_{i=1}^{n} \mathbb{E}(y_i^2 \wedge k|y_i|) = k^2, \qquad (7)$$

or equivalently the unique k such that

$$\sum_{i=1}^{n} \mathbb{E}(\frac{y_i^2}{k^2} \wedge \frac{|y_i|}{k}) = 1.$$

We remark that when y_i are i.i.d., this reduces to

$$n\mathbb{E}(y_1^2 \wedge |y_1| K_{y_1}(n)) = K_{y_1}^2(n).$$

K-Function Bound

And we have the following bounds established by K-functions.

Theorem

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

$$0.67K_Y(n) \leq \mathbb{E}|\sum_{i=1}^{n} y_i| \leq 2K_Y(n). \tag{8}$$

Here, I only provide the upper bound when the variables are symmetric. Denote $Y' = YI_{|Y|\leq k}$ and $Y" = YI_{|Y|>k}$, then from (7) we have that $\mathbb{E}n(Y')^2 \leq K_Y(n)^2$ and $\mathbb{E}n|y"| \leq K_Y(n)$.
Thus by triangle inequality, JENSEN inequality, and that y_i's are independent symmetric around 0, we have

$$\begin{aligned}
\mathbb{E}|\sum_{i=1}^{n} y_i| &\leq \mathbb{E}|\sum_{i=1}^{n} y_i' + \sum_{i=1}^{n} y_i"| \\
&\leq \left(\mathbb{E}|\sum_{i=1}^{n} y_i'|^2\right)^{1/2} + \sum_{i=1}^{n} \mathbb{E}|y_i''| \\
&= \left(n\mathbb{E}(Y')^2\right)^{1/2} + n\mathbb{E}|Y"| \\
&\leq 2K_Y(n).
\end{aligned}$$

Besides $\mathbb{L}^1$ norm, the bound can be extended to $\mathbb{E}\Phi(\sum y_i)$ for Φ convex with a generalized K-function associated to Φ. The details can be found in [3] by Klass or sections 2 and 3 in [2].

Topics Preview

1. Introduction
2. The Norm of Sum of Independent Variables
3. Complete Decoupling of Martingales
4. Complete Decoupling of Sums of Nonnegative Dependent Random Variables

Consider now $\{d_i\}$ the sequence of martingale difference, i.e., $M_n := \sum_{i=1}^{n} d_i$ is a martingale. How large are $\mathbb{E}|\sum_{i=1}^{n} d_i|^p$ and $\mathbb{E}\max_{j\leq n}|\sum_{i=1}^{j} d_i|^p$?

Recall BURKHOLDER-DAVIS-GUNDY's Inequality, we know that

$$\mathbb{E}\max_{j\leq n}\left|\sum_{i=1}^{j} d_i\right|^p = \mathbb{E}\max_{j\leq n}|M_j|^p \approx \mathbb{E}\left(\sum_{i=1}^{n} d_i^2\right)^{p/2}.$$

Theorem (BURKHOLDER-DAVIS-GUNDY's Inequality)

If $d_1, d_2, \ldots, d_n$ are martingale differences, then for any $p \geq 1$,

$$c_p E\left[\left(\sum_{i=1}^{n} d_i^2\right)^{\frac{p}{2}}\right] \leq E\left[\left(\max_{1\leq k\leq n}\left|\sum_{i=1}^{k} d_i\right|\right)^p\right] \leq C_p E\left[\left(\sum_{i=1}^{n} d_i^2\right)^{\frac{p}{2}}\right], \quad (9)$$

Decoupling Inequality for Martingales

BUT the answer is not clear unless this martingale has independent increments. And here we still analyze the $\mathbb{L}^1$ scenario for simplicity.

Theorem

Consider now $\{d_i\}$ the sequence of mean-zero martingale differences, and y_i's are independent copies of d_i's. Then

$$\mathbb{E} \max_{j \leq n} \left| \sum_{i=1}^{j} d_i \right| \leq c\mathbb{E} \left| \sum_{i=1}^{n} y_i \right| \tag{10}$$

Remark

Please do not forget we can extend our results from $|\cdot|$ to $\Phi(\cdot)$, with generalized K-functions.

We first denote $D_n = (d_1, ..., d_n)$ and $Y_n = (y_1, ..., y_n)$. Then

$$\sum_{i=1}^{n} \mathbb{E}(\frac{d_i^2}{k^2} \wedge \frac{|d_i|}{k}) = 1 = \sum_{i=1}^{n} \mathbb{E}(\frac{y_i^2}{k^2} \wedge \frac{|y_i|}{k},)$$

and by definition we have that $K(D_n) = K(Y_n)$.

Then according to Burkholder-Davis-Gundy's Inequality,

$$\mathbb{E}\max_{j\le n}|\sum_{i=1}^{j} d_i| \le c\mathbb{E}\left(\sum_{i=1}^{n} d_i^2\right)^{1/2}$$

$$= c\mathbb{E}\left(\sum_{i=1}^{n}((d_i')^2 + (d")^2)\right)^{1/2} \qquad \text{[by definitions of } d_i' \text{ and } d_i"\text{]}$$

$$\le c\left(\mathbb{E}(\sum_{i=1}^{n}(d_i')^2)^{1/2} + \sum_{i=1}^{n}\mathbb{E}|d_i"|\right) \qquad \text{[by concavity of } \sqrt{\cdot}\text{]}$$

$$\le c\left((\mathbb{E}\sum_{i=1}^{n}(d_i')^2)^{1/2} + K(D_n)\right) \text{ [by Jensen and the definition of K function]}$$

$$\le c\left((K(D_n)^2)^{1/2} + K(D_n)\right)$$

$$= 2cK(D_n) = 2cK(Y_n) \le c\mathbb{E}|\sum_{i=1}^{n} y_i|.$$

Topics Preview

Decoupling Inequality for Nonnegative RV

The decoupling inequality for martingale in the last section then serve as an important lemma to establish the complete decoupling inequality for the sums of nonnegative random variables. That is, we want to study $\mathbb{E}\Phi(\sum d_i)$ where d_i are nonnegative.

Decoupling Inequality for Nonnegative RV (Special Case)

First we take $\Phi(\cdot) = \sqrt{\cdot}$ as a special case.

Theorem

Let $d_1, ..., d_n$ be nonnegative arbitrary dependent random variables and y_i's the independent copies of d_i's. Then there exists a constant C such that

$$\mathbb{E}\sqrt{\sum_{i=1}^{n} d_i} \leq C\mathbb{E}\sqrt{\sum_{i=1}^{n} y_i}. \tag{11}$$

Proof:

Recall the decoupling inequality for MGD:

$$\mathbb{E}\max_{j\le n}|\sum_{i=1}^{j} d_i| \le c\mathbb{E}|\sum_{i=1}^{n} y_i| \tag{12}$$

Let $\epsilon_1, ..., \epsilon_n$ be i.i.d. Rademacher variables, independent of $\{d_i\}$ and $\{y_i\}$. Then $M_j := \sum_{i=1}^{j} \sqrt{d_i}\epsilon_i$ is martingale.

Since $\sqrt{d_i}\epsilon_i$ and $\sqrt{y_i}\epsilon_i$ are the martingale differences, by the previous decoupling inequality for martingales (5), we have

$$\mathbb{E}\sqrt{\sum_{i=1}^{n} d_i} = \mathbb{E}\sqrt{\sum_{i=1}^{n} \sqrt{d_i}^2 \epsilon_i^2}$$

$$\leq C_1 \mathbb{E} \max_{j \leq n} \left|\sum_{i=1}^{j} \sqrt{d_i}\epsilon_i\right| \quad \text{[by B-D-G inequality]}$$

$$\leq C_2 \mathbb{E}\left|\sum_{i=1}^{n} \sqrt{y_i}\epsilon_i\right|$$

$$\leq C_3 \mathbb{E}\sqrt{\sum_{i=1}^{n} y_i} \qquad \text{[by B-D-G inequality again]}$$

A Lower Bound for $\Phi(\cdot) = (\cdot)^p$

Generally, according to [2], for nonnegative dependent r.v.s $(d_1, ..., d_n)$, 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.

$$\mathbb{E}|\sum_{i=1}^{n} d_i|^p \geq C_p \mathbb{E}|\sum_{i=1}^{n} y_i|^p. \tag{13}$$

More General Φ

Theorem (Upper Bound, Proposition 1 in [2])

Suppose Φ 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

$$\mathbb{E}\Phi\left(\sum d_i\right) \leq C\mathbb{E}\Phi\left(\sum y_i\right). \tag{14}$$

Theorem (Lower Bound, Proposition 2 in [2])

Suppose Φ 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 α s.t.

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

More Recent Results

1. Chollette et al [1] showed that the constant C in the upper bound satisfy $C \leq \frac{e}{e-1} \approx 1.5820$.
2. Makaryachev et al [4] showed a sharp decoupling inequality for convex Φ using an independent Poisson(1) variable P:

$$\mathbb{E}\Phi(P \cdot \sum d_i) \geq \mathbb{E}\Phi(\sum y_i).$$

Consequently, for $p \geq 1$

$$\mathbb{E}(\sum d_i)^p \geq A_p^{-1}\mathbb{E}(\sum y_i)^p$$

where $A_p := \mathbb{E}P^p$ is the fractional Bell number

References I

[1] 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.

[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] 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.

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