Topic 3
Conditionally Independent/Tangent 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

1. Review of Tangent Decoupling
2. Examples
3. Inequality for MGF

Topics Preview

1. Review of Tangent Decoupling
2. Examples
3. Inequality for MGF

Background & Definitions

At the beginning of this series, we briefly study the framework of tangent decoupling. In this lecture I provide more details of tangent decoupling.

Definition

Let $\{d_i\}$, $\{y_i\}$ be two sequences of random variables adapted to an increasing sequence of σ-fields $\{\mathcal{F}_i\}$. Then $\{d_i\}$ is said to be **tangent** to $\{y_i\}$ with respect to $\{\mathcal{F}_i\}$ if for all i, $\mathscr{L}(d_i|\mathcal{F}_{i-1}) = \mathscr{L}(y_i|\mathcal{F}_{i-1})$, i.e., the conditional distributions of d_i given $\mathcal{F}_{i-1}$ and y_i given $\mathcal{F}_{i-1}$ are the same.

Definition

A sequence of random variables $\{x_i\}$ is said to be **conditionally symmetric** if x_i is tangent to $-x_i$ w.r.t. $\{\mathcal{F}_i\}$.

Definition

A sequence $\{y_i\}$ of random variables adapted to an increasing sequence of σ-field $\{\mathcal{F}_i\}$ contained in $\mathcal{F}$ is said to be **conditionally independent (CI)** if there exists a σ-algebra $\mathcal{G}$ contained in $\mathcal{F}$ such that $\{y_i\}$ is conditionally independent given $\mathcal{G}$ and $\mathcal{L}(y_i|\mathcal{F}_{i-1}) = \mathcal{L}(y_i|\mathcal{G})$.

Definition

Let $\{d_i\}$ be an arbitrary sequence of random variables, then a conditionally independent sequence $\{y_i\}$ which is also tangent to $\{d_i\}$ will be called a **decoupled** version of $\{d_i\}$.

Construction of Tangent Sequence

Proposition (Kwapień & Woyczyński)

For any sequence of random variables $\{d_i\}$ one can find a decoupled sequence $\{y_i\}$ (on a possibly enlarged probability space) which is tangent to the original sequence and in addition conditionally independent given a master σ-field $\mathcal{G}$. Frequently $\mathcal{G} = \sigma(\{d_i\})$.

More precisely, given $\{d_i\}$, we can construct a tangent sequence w.r.t. $\mathcal{F}_i = \sigma(d_1, ..., d_i)$ (de la Peña [1]):

- First, we take d_1 and y_1 to be two independent copies of the same random mechanism.
- With $(d_1, .., d_{i-1})$, the i-th pair of variables d_i and y_i comes from conditionally independent copies of the same random mechanism given $\mathcal{F}_{i-1}$.
- And y_i's are conditionally independent w.r.t. $= \mathcal{F}_n$.

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

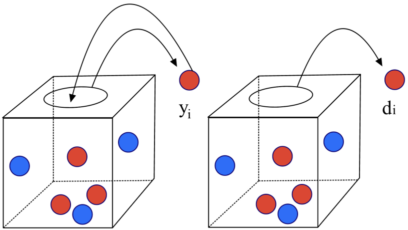
Topics Preview

Example: Simple Sampling

Consider drawing a sample of size n from a box with N balls $\{b_1, ..., b_N\}$, $0 < n \leq N < \infty$. The sequence $\{d_i\}_{i=1}^n$ will represent a sample without replacement. In obtaining a conditionally independent sequence proceed as follows. At the i-th stage of a simple random sample without replacement both d_i and y_i are obtained by sampling uniformly from

$$\{b_1, ..., b_N\} \backslash \{d_1, ..., d_{i-1}\}.$$

It is easy to see that the above procedure will make $\{y_i\}_{i=1}^n$ tangent to $\{d_i\}_{i=1}^n$ with $\mathcal{F}_n = \sigma(d_1, ..., d_n)$. Moreover, $\{y_i\}_{i=1}^n$ is conditionally independent given $\mathcal{G} = \mathcal{F}_n$.



Auto Regressive Model

Let $d_0 = 0$ and for all $i \geq 1$,

$$d_i = \theta d_{i-1} + \epsilon_i \tag{1}$$

where $|\theta| < 1$ and ϵ_i is a sequence of i.i.d., mean zero random variable. Then, a conditionally independent sequence tangent to $\{d_i\}$ is $\{y_i\}$ where for each i,

$$y_i = \theta d_{i-1} + \tilde{\epsilon}_i \tag{2}$$

with $\tilde{\epsilon}_i$ an independent copy of ϵ_i.

Topics Preview

Decoupling Inequality for Products

The results to be introduced in this section are useful in comparing any two tangent sequences when one of them is conditionally independent.

Theorem (de la Peña [1])

Let $\{d_i\}_{i=1}^n$ be a sequence of positive variables. Let $\mathcal{G}$ be a σ-field. Then, for any $\mathcal{G}$-conditionally independent sequence $\{y_i\}_{i=1}^n$, tangent to $\{d_i\}_{i=1}^n$, one has

$$\mathbb{E}\left(\prod_{i=1}^n d_i\right)^{\frac{1}{2}} \leq \left(\mathbb{E}\prod_{i=1}^n y_i\right)^{\frac{1}{2}}. \tag{3}$$

The above result is sharp: Take d_1, y_1 be nonnegative i.i.d. variables. $d_2 = d_1$ and $y_2 = d_1$. Then $\sqrt{d_1 d_2} = d_1$ with mean $\mathbb{E}(d_1)$, and $y_1 y_2 = y_1 d_1$ has the expectation $\mathbb{E}(y_1)\mathbb{E}(d_1) = \mathbb{E}^2(d_1)$.

Proof

Let $\mathcal{F}_i = \sigma(d_1, ..., d_i; y_1, ..., y_i)$. It is easy to see that

$$\mathbb{E}\frac{\prod_{i=1}^{n} d_i}{\prod_{i=1}^{n} \mathbb{E}(d_i|\mathcal{F}_{i-1})} = 1. \tag{4}$$

Since $\{y_i\}$ is tangent to $\{d_i\}$ and conditionally independent given $\mathcal{G}$,

$$\mathbb{E}(x_i|\mathcal{F}_{i-1}) = \mathbb{E}(y_i|\mathcal{F}_{i-1}) = \mathbb{E}(y_i|\mathcal{G}). \tag{5}$$

$$\mathbb{E}\left(\prod_{i=1}^{n} d_i\right)^{\frac{1}{2}} = \mathbb{E}\left[\left(\prod_{i=1}^{n} d_i\right)^{\frac{1}{2}} \times \frac{\left(\prod_{i=1}^{n} \mathbb{E}(d_i|\mathcal{F}_{i-1})\right)^{\frac{1}{2}}}{\left(\prod_{i=1}^{n} \mathbb{E}(d_i|\mathcal{F}_{i-1})\right)^{\frac{1}{2}}}\right]$$

$$= \mathbb{E}\left[\frac{\left(\prod_{i=1}^{n} d_i\right)^{\frac{1}{2}}}{\left(\prod_{i=1}^{n} \mathbb{E}(d_i|\mathcal{F}_{i-1})\right)^{\frac{1}{2}}}\left(\prod_{i=1}^{n} \mathbb{E}(d_i|\mathcal{F}_{i-1})\right)^{\frac{1}{2}}\right]$$

$$\leq \sqrt{\mathbb{E}\frac{\left(\prod_{i=1}^{n} d_i\right)}{\left(\prod_{i=1}^{n} \mathbb{E}(d_i|\mathcal{F}_{i-1})\right)}\mathbb{E}\left(\prod_{i=1}^{n} \mathbb{E}(d_i|\mathcal{F}_{i-1})\right)}$$

(by Hölder's Inequality)

$$= \left(\mathbb{E}\prod_{i=1}^{n} \mathbb{E}(d_i|\mathcal{F}_{i-1})\right)^{\frac{1}{2}}$$

$$\left(\mathbb{E}\prod_{i=1}^{n}\mathbb{E}(d_i|\mathcal{F}_{i-1})\right)^{\frac{1}{2}}$$

$$=\left(\mathbb{E}\left(\prod_{i=1}^{n}\mathbb{E}(y_i|\mathcal{G})\right)\right)^{\frac{1}{2}}$$

$$=\left(\mathbb{E}\left(\mathbb{E}(\prod_{i=1}^{n}y_i|\mathcal{G})\right)\right)^{\frac{1}{2}} \quad \text{(since } \{y_i\} \text{ is } \mathcal{G}\text{-conditionally independent)}$$

$$=\left(\mathbb{E}\prod_{i=1}^{n}y_i\right)^{\frac{1}{2}}.$$

Decoupling Inequality for MGF

A direct consequence of this theorem is the decoupling inequality for the moment generating functions of the sums.

Corollary

Let $\{d_i\}_{i=1}^n$ be a sequence of positive variables. Let $\mathcal{G}$ be a σ-field. Then, for any $\mathcal{G}$-conditionally independent sequence $\{y_i\}_{i=1}^n$, tangent to $\{d_i\}_{i=1}^n$, one has, for all λ finite,

$$\mathbb{E}\exp\left(\lambda\sum_{i=1}^n d_i\right) \leq \sqrt{\mathbb{E}\exp\left(2\lambda\sum_{i=1}^n y_i\right)}. \tag{6}$$

Note that if y_i's are mean zero, the $\sqrt{\cdot}$ symbol may be removed.

We can generalize this corollary to the following extension:

Corollary (de la Peña [2])

If y_i is a decoupled version of d_i, then for all r.v. $g > 0$ adapted to $\sigma\{d_1, ..., d_i\}$

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

This inequality can be used to develop self-normalized inequalities, and we will see an application of this in the establishment of the BERNSTEIN's inequality for self-normalized martingales.

References I

[1] 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.

[2] 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.