Topics in Probability Theory **Fall 2023 (Georgia Institute of Technology)**

Lecture 1: Introduction of Decoupling

Lecturer: ***Víctor H. De La Peña*** *Scribes: Heyuan Yao*

1.1 Decoupling: from Dependence to Independence

Decoupling and self-normalization emerge as pivotal statistical tools, tackling intricate challenges that transcend the realm of conventional statistical methods. Fundamentally, they 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 correlated variables by treating them akin to independent entities**. Notably, it offers a natural framework for **forging precise exponential inequalities tailored for self-normalized martingales**.

Prominent illustrations of self-normalized processes encompass the **t-statistic within correlated random variables**, alongside **(self-normalized) extensions of the Kolmogorov's law of the iterated logarithm**.

Two classical decoupling methods stand out: **complete decoupling** and **conditionally independent (tangent) decoupling**. In the following discussion, I will elucidate the frameworks of these methods, shedding light on their intrinsic mechanics. Moreover, I will delve into the resultant inequalities that arise as a consequence of their application.

1.2 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:

$$\mathbb{E}\sum_{i=1}^n d_i = \mathbb{E}\sum_{i=1}^n y_i \tag{1.1}$$

In complete decoupling, one compares $\mathbb{E}f(\sum d_i)$ to $\mathbb{E}f(\sum y_i)$ for more general functions than the linear mapping $f(x) = cx$.

Let us now place ourselves in the setting of random sampling, where complete decoupling happens. Let the population C consist of N values $c_1, ..., c_N$, $d_1, ..., d_n$ denote a random sample without replacement from C, and $y_1, ..., y_n$ denote a random sample with replacement from C (of course, $n \leq N$). The random variables $y_1, ..., y_n$ are independent and identically distributed (i.i.d). Moreover, for all i, $d_i \overset{\mathscr{L}}{=} y_i$. HOEFFDING developed the following widely used complete decoupling inequality:

Theorem 1.1 *For every continuous convex function* Φ*,*

$$\mathbb{E}f(\sum d_i) \leq \mathbb{E}f(\sum y_i). \tag{1.2}$$

Remark 1.2 *My friend* SHAO *extended this inequality to the case of negatively associated random variables. You may find it in the reference list.*

Applications of complete decoupling include, among others, tools for the optimization of stochastic processes such as the scheduling of dependent computer servers MAKARYCHEV and SVIRIDENKO. In what follows we present their (sharp) complete decoupling inequality for sums of nonnegative dependent random variables that provides a reverse HOEFFDING's inequality for L_p moments. The price one pays is a constant.

Theorem 1.3 *Let $\pi(1)$ be a Poisson random variable with mean 1. Besides the same setting above, we further assume d_i's are nonnegative (of course, so are y_i's). Then, for $p \geq 1$, $\|\sum_{i=1}^n y_i\|_p \leq \|\pi(1)\|_p \|\sum_{i=1}^n d_i\|_p$, i.e.,*

$$\mathbb{E}\left(\sum_{i=1}^n y_i\right)^p \leq \mathbb{E}\pi(1)^p \, \mathbb{E}\left(\sum_{i=1}^n d_i\right)^p \tag{1.3}$$

Remark 1.4 *In particular, when $p = 1$ we obtain the original complete decoupling equality in (1.1).*

1.3 Conditionally Independent (Tangent) Decoupling

The theory of martingale inequalities has been central 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 BURKHODER and McConell that represents a step in extending the theory of martingales to Banach spaces.

Let us first get familiar with the tangent sequences, with the following definition.

Definition 1.5 *Let d_i and e_i be two sequence of r.v.s adapted to $\sigma-$fields $\mathcal{F}_i$. Then d_i and e_i are said to be* ***tangent*** *with respect to $\mathcal{F}_i$ if, for all i,*

$$\mathscr{L}(d_i|\mathcal{F}_{i-1}) = \mathscr{L}(e_i|\mathcal{F}_{i-1}), \tag{1.4}$$

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

Let $x_1, .., x_n$ be an arbitrary sequence of dependent random variables adapted to an increasing sequence of $\sigma-$fields $\mathcal{F}_i$. Then, one can construct a sequence of random variables $y_1, ..., y_n$ that are conditionally independent given $\mathcal{G} = \mathcal{F}$. The construction proceeds as follows:

- First, we take x_1 and y_1 to be two independent copies of the same random mechanism.
- Having constructed $x_1, .., x_{i-1}$; $y_1, .., y_{i-1}$, the i-th pair of variables x_i and y_i comes from i.i.d. copies of the same random mechanism given $\mathcal{F}_{i-1}$.

And you may find Figure 1.3 helpful. It is easy to see that using this construction and taking

$$\mathcal{F}'_i = \mathcal{F}_i \vee \sigma(y_1, ..., y_i),$$

the sequence x_i, y_i satisfy

$$\mathscr{L}(x_i|\mathcal{F}'_{i-1}) = \mathscr{L}(y_i|\mathcal{F}'_{i-1}) = \mathscr{L}(y_i|\mathcal{G})$$

and the sequence $y_1, ..., y_n$ is is conditionally independent given $\mathcal{G} = \mathcal{F}$.

$$\begin{tikzcd}
x_1 \arrow[r] \arrow[dr] & x_2 \arrow[r] \arrow[dr] & x_3 \arrow[r] \arrow[dr] & x_4 \arrow[r] \arrow[dr] & \cdots \arrow[r] \arrow[dr] & x_{j-1} \arrow[r] \arrow[dr] & x_j \\
y_1 & y_2 & y_3 & y_4 & \cdots & y_{j-1} & y_j.
\end{tikzcd}$$

Definition 1.6 *A sequence y_i of random variables satisfying the above conditions is said to be a* ***decoupled*** *$\mathcal{F}'$-tangent version of x_i.*

As in the case of complete decoupling, linearity of expectations provides the canonical example of a decoupling "equality." In conditionally independent decoupling, one replaces dependent random variables with decoupled (conditionally independent) random variables.

Theorem 1.7 *If y_i is a decoupled $\mathcal{F}'$-tangent version of x_i, and for any i, $\mathbb{E}|x_i|< \infty$, then*

$$\mathbb{E}\sum_{i=1}^{n} x_i = \mathbb{E}\sum_{i=1}^{n} y_i. \tag{1.5}$$

Proof: To see this, note that

$$\begin{aligned}\mathbb{E}\sum_{i=1}^{n} x_i = \sum_{i=1}^{n}\mathbb{E}x_i = \sum_{i=1}^{n}\mathbb{E}[\mathbb{E}(x_i|\mathcal{F}'_{i-1})] \\ = \sum_{i=1}^{n}\mathbb{E}[\mathbb{E}(y_i|\mathcal{F}'_{i-1})] = \sum_{i=1}^{n}\mathbb{E}[\mathbb{E}(y_i|\mathcal{G}] \\ = \mathbb{E}[\mathbb{E}(\sum_{i=1}^{n} y_i|\mathcal{G}] = \mathbb{E}\sum_{i=1}^{n} y_i\end{aligned}$$

■

Now let us place ourselves, again, in the setting of random sampling, where, however, tangent decoupling takes place. We show how to decouple a sample without replacement and show how the decoupled sequence relates to sampling without replacement and sampling with replacement. (In survey sampling, we treat draws without replacement as if they were independent, though they are actually weakly coupled.) As before, consider drawing samples of size n from a population C that consists of N values. A conditionally independent sample can be constructed as follows. At the i-th stage of a simple random sampling with- out replacement, both d_i and e_i are obtained sampling uniformly from $c_1, ..., c_N$, excluding $d_1, ..., d_{n-1}1$. This may be attained by at the i-th stage first draw e_i and return it to the population. Then draw e_i and put its value aside.

As we approach the conclusion of our session, let's examine a sharp decoupling inequality bolstered by constraints. This pivotal finding will serve as a cornerstone for our subsequent pursuit: the attainment of a precise expansion of Bernstein's inequality. This extension caters to the scenario of self-normalized martingales, seamlessly bridging the gap from independent random variables to their intricate counterparts.

Theorem 1.8 *If y_i is a decoupled $\mathcal{F}'$-tangent version of x_i, then for all r.v. $g > 0$ adapted to $\mathcal{F}_i$*

$$\mathbb{E}\left[g\exp\left(\lambda\sum_{i=1}^{n} x_i\right)\right] \leq \sqrt{\mathbb{E}\left[g^2\exp\left(2\lambda\sum_{i=1}^{n} x_i\right)\right]} \tag{1.6}$$

A use of the exponential decoupling inequality (with $g = 1$) found in (1.6) gives

$$\mathbb{E}\left[\exp\left(\lambda \sum_{i=1}^{n} x_i\right)\right] \leq \sqrt{\mathbb{E}\left[\exp\left(2\lambda \sum_{i=1}^{n} x_i\right)\right]}$$

Reference

- de la Pena V.H. Bounds on the expectation of functions of martingales and sums of positive RVs in terms of norms of sums of independent random variables, *Proc. Amer. Math. Soc.*, (108, no. 1): 233–239, 1990. MR990432
- de la Pena V.H. A general class of exponential inequalities for martingales and ratios, *Ann. Probab.*, (27, no. 1): 537– 564, 1999. MR1681153
- de la Pena V.H., Gine E. *Decoupling: From Dependence to Independence*, Springer, New York, 1999. MR1666908
- de la Pena V.H., Lai T.L., Shao Q-M. *Self-Normalized Processes: Limit Theory and Statistical Applications*, Springer, 2009.
- de la Pena, V.H. From decoupling and self-normalization to machine learning. *Notices of the American Mathematical Society*, 66(10), 2019.
- Gosset (Student) WS. The probable error of the mean, *Biometrika*, (6): 1–25, 1908.
- Hitczenko P. Comparison of moments for tangent sequences of random variables, *Probab. Theory Related Fields*, (78, no. 2): 223–230, 1988. MR945110
- Hoeffding W. Probability inequalities for sums of bounded random variables, *J. Amer. Statist. Assoc.*, (58, no. 301): 13–30, 1963. MR0144363
- Makarychev K, Sviridenko M. Solving optimization problems with diseconomies of scale via decoupling, *J. ACM*, (65, no. 6): Article 42, 2018. MR3882265
- Shao Q-M. A comparison theorem on moment inequalities between negatively associated and independent random variables, *J. Theoret. Probab.*, (13, no. 2): 343–356, 2000. MR1777538
- Zinn J. Comparison of martingale difference sequences. In: *Probability in Banach Spaces* V, Lecture Notes in Math., (1153). Springer; 1985: 453–457. MR821997