Some Old and New Results in Decoupling

by

Victor H. de la Pena
Department of Statistics
Columbia University

Introduction.

In this talk we will discuss two types of decoupling results. The first type is decoupling of tangent sequences, the second being complete decoupling.

Conditionally Independent (Tangent) Decoupling

The theory of martingale inequalities has been central in the development of modern probability theory. This theory 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 (1983) which represents a step in extending the theory of martingales to Banach spaces.

Let $\{d_i\}$ and $\{e_i\}$ be two sequences of random variables adapted to the σ-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,

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

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

Let $d_1, ..., d_n$ be an arbitrary sequence of dependent random variables adapted to an increasing sequence of $\sigma-$fields $\{\mathcal{F}_i\}$. Then, as shown in de la Pena and Gine (1999), one can construct a sequence $e_1, ..., e_n$ of random variables which is conditionally independent given $\mathcal{G} = \sigma(d_1, ..., d_n)$. The construction proceeds as follows: First we take e_1 and d_1 to be two independent copies of the same random mechanism. Having constructed $d_1, ..., d_{i-1}; e_1, ..., e_{i-1}$, the e_i is an i.i.d. copy of variables d_i given $\mathcal{F}_{i-1}$. It is easy to see that using this construction and taking

$$\mathcal{F}'_i = \mathcal{F}_i \vee \sigma(e_1, ..., e_i),$$

the sequences $\{d_i\}$, $\{e_i\}$ satisfy

$$\mathcal{L}(d_i|\mathcal{F}'_{i-1}) = \mathcal{L}(e_i|\mathcal{F}'_{i-1}) = \mathcal{L}(e_i|\mathcal{G}),$$

and the sequence $e_1, ..., e_n$ is conditionally independent given $\mathcal{G} = \mathcal{F}_n$.

A sequence $\{e_i\}$ of random variables satisfying the above conditions is said to be a *decoupled* $\mathcal{F}'_i$-tangent version of $\{d_i\}$.

It is important to remark that linearity of expectations provides the canonical example of a decoupling "equality" as we will show next.

In conditionally independent decoupling one replaces the sequence of dependent random variables $\{d_i\}$ by a decoupled,$\mathcal{G}$ conditionally independent sequence.

If $E|d_i| < \infty$ for all i, then,

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

as we show next.

$$E\sum_{i=1}^{n} d_i = \sum_{i=1}^{n} Ed_i = \sum_{i=1}^{n} E(E(d_i|\mathcal{F}'_{i-1})) = \sum_{i=1}^{n} E(E(e_i|\mathcal{F}'_{i-1}))$$

$$= \sum_{i=1}^{n} E(E(e_i|\mathcal{G})) = E(E(\sum_{i=1}^{n} e_i|\mathcal{G})) = E\sum_{i=1}^{n} e_i.$$

The first general decoupling inequality for tangent sequences was obtained by Zinn (1985) and extended by Hitczenko (1988).

A turning point in the theory of decoupling for tangent sequences has been a 1986 result of Kwapien and Woyczynski (see Kwapien and Woyczinsky (1992) for the exact reference). It is shown in that paper for the first time, and in a precise manner, that one can always obtain a decoupled tangent sequence to any adapted sequence hence, making general decoupling inequalities widely applicable

Developments of the theory are found in hard copy in Kwapien and Woyczynski (1992) and de la Pena and Gine (1999).

Decoupling and Self-Normalization

Next, we present a sharp decoupling inequality with constraints from de la Pena (1999) which naturally lead to the development of self-normalized inequalities for martingales. This result will be used later to obtain a sharp extension of Bernstein's inequality for independent random variables to the case of self-normalized martingales.

Let $\{d_i\}$ and $\{e_i\}$ be two tangent sequence with $\{e_i\}$ decoupled. Then for all $g \geq 0$ adapted to $\sigma(\{d_i\})$

$$(1) \qquad Eg\exp\{\lambda \sum_{i=1}^{n} d_i) \leq \sqrt{Eg^2 \exp(2\lambda \sum_{i=1}^{n} e_i)}.$$

Bernstein's Inequality. Let $\{x_i\}$ be a sequence of independent random variables. Assume that $E(x_j) = 0$ and

$E(x_j^2) = \sigma_j^2 < \infty$ and set $v_n^2 = \sum_{j=1}^n \sigma_j^2$. Furthermore, assume that there exists a constant $0 < c < \infty$ such that, almost surely, $E(|x_j|^k) \leq (k!/2)\sigma_j^2 c^{k-2}$ for all $k > 2$. Then for all $x > 0$,

$$(8) \qquad P(\sum_{i=1}^n x_i > x) \leq \exp(-\frac{x^2}{2(v_n^2 + cx)}).$$

The following is a self-normalized analog for random variables from de la Pena (1999).

Self-normalized Bernstein's Inequality. Let $\{d_i, \mathcal{F}_i\}$ be a martingale difference sequence. Assume the following $E(d_j|\mathcal{F}_{j-1}) = 0$ and $E(d_j^2|\mathcal{F}_{j-1}) = \sigma_j^2 < \infty$ (satisfied by sub-exponential random variables) and set $V_n^2 = \sum_{j=1}^n \sigma_j^2$. Furthermore, assume that there exists a constant $0 < c < \infty$ such that, almost surely, $E(|d_j|^k|\mathcal{F}_{j-1}) \leq (k!/2)\sigma_j^2 c^{k-2}$ for all $k > 2$. Then for all $x, y > 0$,

$$P(\frac{\sum_{i=1}^n d_i}{V_n^2} > x) \leq E[\exp(-\frac{x^2 V_n^2}{2(1+cx)}|\frac{\sum_{i=1}^n d_i}{V_n^2} > x)] \leq$$

$$\sqrt{E[\exp(-\frac{x^2 V_n^2}{2(1+cx)})}.$$

Furthermore,

$$(2) \qquad P(\frac{\sum_{i=1}^n d_i}{V_n^2} > x, \frac{1}{V_n^2} \leq y) \leq \exp(-\frac{x^2}{2y(1+cx)}).$$

Proof: Use (1) to decouple with the constraint $g = 1(\frac{1}{V_n^2} \leq y)$. Continue conditionally on $\mathcal{G}$ using the fact that the e_i are conditionally independent given $\mathcal{G}$.

Example: Conditionally Independent Sampling of Simple Random Sampling Without Replacement.

In this example we show how to decouple a sample without replacement and show how the decoupled sequence relates to sampling with replacement. (eg. In survey sampling we treat draws without replacement as if they were independent, though they are actually weakly coupled). Consider drawing samples of size n from a population C which consists of N values. Let $d_1, ..., d_n$ denote a sample without replacement and $y_1, ..., y_n$ a sample with replacement. A conditionally independent sample can be constructed as follows. At the ith stage of a simple random sampling without replacement both d_i and e_i are obtained sampling uniformly from $\{c_1, ..., c_N\}$ excluding $\{d_1, ..., d_{i-1}\}$. This may be attained by selectively returning items to the C . More precisely, at the i-th stage first draw e_i and return it to the population. Then draw d_i and put its value aside. It is easy to see that the above procedure will make $\{e_i\}$ tangent to $\{d_i\}$ with $\mathcal{F}_n = \sigma(d_1, ..., d_n; e_1, ..., e_n)$. More over, the sequence $\{e_i\}$ is conditionally independent given $\mathcal{G} = \sigma(d_1, ..., d_N)$.

The tools developed have been successfully applied in diverse areas such as extension of martingale results to infinite dimensions including Banach spaces, self-normalized martingales, stochastic integration, empirical processes including U-statistics and U-processes, density estimation, sequential analysis, matrix completion. See Kwapien and Woyczynski (1992), de la Pena and Gine (1999) and de la Pena et. al. (2009). More recent applications of conditionally independent decoupling include Candes and Recht (2009) which deals with exact matrix completion via convex optimization. There is a regent surge of interest in applying and developing the methods presented in this survey. In particular Rahklyn et. al. (2015) uses conditional independent decoupling techniques to study sequential complexities and exponential inequalities for martingales in Banach spaces.

Complete Decoupling

Let $\{d_i\, i = 1, ..., n\}$ be a sequence of dependent random variables with $E|d_i| < \infty$. 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. Since $Ed_i = Ey_i$, linearity of expectations provides the first "complete decoupling" equality:

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

As a concrete example for constructing the sequence $\{y_i\}$, let $\{d_i^{(j)}, i = 1, ..., n; j = 1, ..., n\}$ be independent copies of $\{d_i\}$ and take $y_i = d_i^{(i)}$. Then, it is easy to see that $\{y_i\}$ is a sequence of independent random variables since each row in the array is independent of the others and

$$E\sum_{i=1}^{n} d_i = \sum_{i=1}^{n} Ed_i = \sum_{i=1}^{n} Ed_i^{(i)} = \sum_{i=1}^{n} Ey_i = E\sum_{i=1}^{n} y_i.$$

In complete decoupling, one compares $Ef(\sum d_i)$ to $Ef(\sum y_i)$ for more general functions than $f(x) = x$, including convex and concave functions.

Example: Complete Decoupling of Simple random Sampling without Replacement

Let the population C consist of N values $c_2, c_2, ..., c_N$. Let $d_1, d_2, ..., d_n$ denote a random sample without replacement from C and let $y_1, y_2, ..., y_n$ denote a random sample with replacement from C. The random variables $y_1, ..., y_n$ are independent and identically distributed. Moreover, for all i, d_i and the y_i have the same marginal distributions. Hoeffding (1963) developed the following widely used complete decoupling inequality. For every continuous and convex function Φ

$$E\Phi(\sum_{i=1}^{n} d_i) \leq E\Phi(\sum_{i=1}^{n} y_i).$$

Shao (2000) extended it to the case of negatively associated random variables.

As a special case of the (sharp) complete decoupling inequality for sums of non-negative dependent random variables presented below, provides a reverse Hoeffding's inequality for L_p moments. The price one pays is a constant.

Makarychev and Sviridenko (2018) uses the following complete decoupling inequalities to develop algorithms for the distribution of loads in parallel machines. More precisely to develop stochastic optimization tools for energy efficient routing load balancing in parallel machines.

Theorem Let π be a Poisson random variable with mean 1. Assume that d_i's are a sequence of arbitrarily dependent non-negative random variables. Let $y_1, ..., y_n$ be independent random variables with y_i having the same distributions as d_i for all i. Assume that π is independent of $y_1, ..., y_n$. Then, for all convex functions $\Phi(\cdot)$,

$$E\Phi(\sum_{i=1}^{n} y_i) \leq E\Phi(\pi \sum_{i=1}^{n} d_i).$$

The inequality reverses when the functions Φ are concave. In this case,

$$E\Phi(\pi \sum_{i=1}^{n} d_i) \leq E\Phi(\sum_{i=1}^{n} y_i).$$

If $\Phi, \Phi(0) = 0$ is concave increasing then one can show that,

$$E\Phi(\sum_{i=1}^{n} d_i) \leq \frac{e}{e-1} E\Phi(\sum_{i=1}^{n} y_i).$$

It is interesting to note that when
$\Phi(x) = x$, then

$$E\sum_{i=1}^{n} y_i = E\pi(1)E\sum_{i=1}^{n} d_i = E\sum_{i=1}^{n} y_i.$$

A companion result involves a mean-zero martingale difference sequence $\{d_i\}$

$$E|\sum_{i=1}^{n} d_i|^p \leq c_p E|\sum_{i=1}^{n} y_i|^p,$$

For $1 \leq p \leq 2$. It is easy to see that $c_2 = 1$. A stronger form of the result is

$$E \max_{j=1,\ldots,n} |\sum_{i=1}^{n} d_i|^p \leq C_p E |\sum_{i=1}^{n} y_i|^p.$$

Proof Sketch: First use the Burkholder-Davis-Gundy inequality to change the problem into one of non-negative variables, decouple and use the square function argument again.

Example

If the $\{d_i\}$ are stationary either non-negative, or a martingale difference sequence, then the $\{y_i\}$ are i.i.d.

The above result was introduced by de la Pena (1990) in the case of concave and some convex functions including powers with different constants. The sharp constants were first obtained in de la Pena, Ibragimov Sharakhmetov (2003) and Makarychev and Sviridenko (2018) in its full generality.

More recently in Chollette, Klass, de la Pena (2023) we have developed a companion sharp decoupling inequality for maximums where the d_i are arbitrary dependent random variables (no further assumptions).

$$P(\max_{i=1,\ldots,n} d_i > t) \leq \frac{e}{e-1} P(\max_{i=1,\ldots,n} y_i > t).$$

for all t. The proof follows an optimization argument

Example.

Suppose that you wan to find the value $t_{\alpha,d}$ for which

$$P(\max_{i=1,\ldots,n} d_i > t_{\alpha,d}) \leq \alpha.$$

This can be calculated if you are able to first calculate the value $t_{\frac{e-1}{e}\alpha,y}$ such that

$$P(\max_{i=1,\ldots,n} y_i > t_{\frac{e-1}{e}\alpha,y}) \leq \frac{e-1}{e}\alpha,$$

which might be easier due to independence.

Finally, set $t_{\alpha,d} = t_{\frac{e-1}{e}\alpha,y}$.

Summary The tools developed have been successfully applied in diverse areas such as extension of martingale results to infinite dimensions including Banach spaces, stochastic integration, empirical processes including U-statistics and U-processes, density estimation, sequential analysis, matrix completion and several others. See Kwapien and Woyczynski (1992), de la Pena and Gine (1999) and de la Pena et. al. (2009) for details.

Additional applications of conditionally independent decoupling include Candes and Recht (2009) which deals with exact matrix completion via convex optimization. There is a recent surge of interest in applying and developing the methods presented in this survey. In particular Rahklyn et. al. (2015) uses conditional independent decoupling techniques to study sequential complexities and exponential inequalities for martingales in Banach spaces. Makarychev and Sviridenko (2018) uses complete decoupling inequalities to develop stochastic optimization tools for energy efficient routing load balancing in parallel machines.

Summary

As can be seen from the broad range of results and applications, it is worth looking at problems using the decoupling perspectives.

References:

1. Belloni, A, Chernozhukov, V., Lie, W. (2014). Pivotal estimation via square-root Lasso in nonparametric regression. Ann. Statist. **42** , No. 2, pp 757-788.

2. Burkholder, D. L. (1983). A geometric condition that implies the existence of certain singular integrals of Banach-space-valued functions. *Conference on harmonic analysis in honor of Antoni Zygmund, Vol.***I, II** (Chicago Ill,, 1981) 270-286. Wadsworth. Math. Ser. Wadsworth, Belmont, Calif.

3. Candes, E. J. and Recht, B. (2009). Exact Matrix Completion via Convex Optimization. *Foundations of Computational Mathematics* , Springer.

4. Collette, L., Klass, M. J., de la Pena (2023). The price of independence in a model with unknown dependence. *Mathematical Social Sciences* Elsevier.
Published electronically March 2023.

5. de la Pena, V. H. (1990). Bounds on the expectation of functions of martingales and sums of positive rv's in terms of norms of sums of independent random variables. *Proceedings of the American Mathematical Society*, Vol. **108,** No. 1, pp. 233-239.

6. de la Pena, V. H. (1999). A general class of exponential inequalities for martingales and ratios. *Annals of Probability,* Vol. **27**, No. 1, pp. 537-564.

7. de la Pena, V. H. and Gine, E. (1999). *Decoupling: From Dependence to Independenc.* Springer, New York.

8. de la Pena, V. H., Klass, M. J. and Lai, T. L. (2004). ion in a model with unknown dependence Self-Normalized processes: exponential inequalities, moment bounds and iterated logarithm laws. *Annals of Probability,* Vol. **32** , No. 3A, 2004, pp. 1902-1933.

9. de la Pena, V. H., Ibragimov, R. and Sharakhmetov, Sh. (2003). On extremal distributions and sharp $L_p norms$ 675.

10. Hitczenko, P. (1988). Comparison of moments of tangent sequences of random variables. *Probab. Theory Rel. Fields*, Vol.**78**, pp. 223-230.

11. Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables.*JASA* ,Vol **58**, pp. 13-30.

12. Kwapien, S. and Woyczynski, W. (1992). *Random Series and Stochastic Integrals: Single and Multiple.* Birkhauser, Boston.

13. Makarychev, K. and Sviridenko, M. (2018). Solving Optimization Problems with Diseconomies of Scale via Decoupling. *Journal of the ACM*, Vol **65**, No. 6, Article 42.

14. Rakhlin, A. Sridharan, K, and Tewari, A. (2015). Se-

quential complexities and uniform martingale laws of large numbers. *Probability Theory and Related Fields.* Vol. **161** issue 1-2, pp. 111-153.
15. Shao, Qi-Man, (2000). A comparison theorem on moment inequalities between negatively-associated and independent random variables. *Journal of Theoretical Probability*, Vol **13** (2) pp. 343-356.

Zinn, J. (1985). Comparison of martingale difference sequences. In *Probability in Banach Spaces V, Lecture Notes in Math. 1153* pp 453-457.