Decoupling for Generalized U-statistics and its Application to Nonparametrics

長崎大学　経済学部
永井　圭二

1 Introduction

Decoupling inequalities have had a major role in recent advances on the asymptotic theory of U-statistics. In this section, we briefly overview the decoupling inequalities developed by de la Peña (1992) and de la Peña and Montgomery-Smith (1995).

U-statistics, first considered by Holmos(1964) in connection with unbiased statistics, and formally introduced by Hoeffding(1948), are defined as follows; Given an i.i.d. sequence of random variables $X_1, X_2, \ldots$ with values in a measurable space $(S, \mathbb{S})$, and a measurable function $h : S^m \to \mathbb{R}$, the U-statistics of order m and kernel h based on the sequence $\{X_i\}$ are

$$U_n = \frac{(n-m)!}{n!} \sum_{(i_1,\ldots,i_m)\in I_n^m} h(X_{i_1}, \ldots, X_{i_m}), \quad n \geq m,$$

where

$$I_n^m = \Big\{(i_1, \ldots, i_m) : i_j \in \mathbb{N}, 1 \leq i_j \leq n, \quad i_j \neq i_k \quad \text{if} \quad j \neq k\Big\}.$$

U-statistics appear often in statistics either as unbiased estimators of parameters of interest or, as components of higher-order terms in expansions of smooth statistics (von Mises expansion, delta method).

Hoeffding's U-statistics can be generalized by the following statistics:

$$U_n = \sum_{(i_1,\ldots,i_m)\in I_n^m} h_{i_1,\ldots,i_m}(X_{i_1}, \ldots, X_{i_m}),$$

where $h_{i_1,\ldots,i_m} : S^m \to B$ are mesurable $Banach$-valued functions.

By a tail probability decoupling inequality for U-statistics we mean a inequality between the quantities

$$P\left\{\left\|\sum_{I_n^m} h_{i_1,\ldots,i_m}(X_{i_1}, \ldots, X_{i_m})\right\| > t\right\} \tag{1.1}$$

and

$$P\left\{\left\|\sum_{I_n^m} h_{i_1,\ldots,i_m}(X_{i_1}^{(1)}, \ldots, X_{i_m}^{(m)})\right\| > t\right\} \tag{1.2}$$

possibly multiplied by constants that depends only on m, where the sequences $\{X_i^{(k)}\}$, $k = 1, \ldots, m$, are independent copies of the original sequence $\{X_i\}$.

2 Tail Probability Decoupling Inequality

Recently decoupling inequalities in U-statistics were discovered by de la Peña (1992) and de la Peña and Montgomery-Smith (1995). Their important applications have been appeared in Giné and

Zinn (1994) and Arcones and Giné (1993, 1995), among others. In this section we derive a version of decoupling inequalities and a multiple martingale method for U-statistics which we shall use many times in the following section. See Giné and Zinn (1994) and Arcones and Giné (1993, 1995) for the applications for the decoupling inequalities.

Theorem 2.1 *Suppose $\{X_i\}$ is a sequence of independent random variables and $\{X_i^{(j)}\}, j = 1, \ldots, k$, are k independent copies of $\{X_i\}$. Let f be a function of k real variables. Let $\mathbf{I}_k = \{(i_1, \ldots, i_k); i_j \in \mathbb{N}, \quad i_j \neq i_k \quad \text{if} \quad j \neq k\}$. Then, for any finite region L in R^k consisting of lattice points of positive integers and any $t > 0$,*

$$P\left\{ \max_{\mathbf{n}\in L} \left| \frac{1}{n_1 n_2 \cdots n_k} \sum_{\substack{\mathbf{1}_k \leq \mathbf{i} \leq \mathbf{n} \\ \mathbf{i}\in \mathbf{I}_k}} f(X_{i_1}, X_{i_2}, \ldots, X_{i_k}) \right| > t \right\}$$

$$\leq C_k P\left\{ C_k \max_{\mathbf{n}\in L} \left| \frac{1}{n_1 n_2 \cdots n_k} \sum_{\substack{\mathbf{1}_k \leq \mathbf{i} \leq \mathbf{n} \\ \mathbf{i}\in \mathbf{I}_k}} f(X_{i_1}^{(1)}, X_{i_2}^{(2)}, \ldots, X_{i_k}^{(k)}) \right| > t \right\} \tag{2.1}$$

where C_k depends on k only, $\mathbf{n}= (n_1, n_2, \ldots n_k)$, $\mathbf{i}= (i_1, i_2, \ldots i_k)$, $\mathbf{1}_k$ is the k-tuples of 1's, and $\mathbf{1}_k \leq \mathbf{i} \leq \mathbf{n}$ means $1 \leq i_j \leq n_j$ for all $j = 1, \ldots k$.

参考文献

[1] Victor H. de la Peña and S. J. Montgomery-Smith. Decoupling inequalities for the tail probabilities of multivariate U-statistics. *Ann. Probab.*, 23(2):806–816, 1995.