The Annals of Probability
1999, Vol. 27, No. 1, 432–453

SUB-BERNOULLI FUNCTIONS, MOMENT INEQUALITIES AND STRONG LAWS FOR NONNEGATIVE AND SYMMETRIZED *U*-STATISTICS[1]

BY CUN-HUI ZHANG

Rutgers University

This paper concerns moment and tail probability inequalities and the strong law of large numbers for U-statistics with nonnegative or symmetrized kernels and their multisample and decoupled versions. Sub-Bernoulli functions are used to obtain the moment and tail probability inequalities, which are then used to obtain necessary and sufficient conditions for the almost sure convergence to zero of normalized U-statistics with nonnegative or completely symmetrized kernels, without further regularity conditions on the kernel or the distribution of the population, for normalizing constants satisfying a simple condition. Moments of U-statistics are bounded from above and below by that of maxima of certain kernels, up to scaling constants. The multisample and decoupled versions of these results are also considered.

1. Introduction. This paper concerns moment and tail probability inequalities and the strong law of large numbers (SLLN) for U-statistics with nonnegative or symmetrized kernels and their multisample and decoupled versions.

1.1. *Overview.* Let $\{X, X_n, n \geq 1\}$ be a sequences of iid random variables with a common distribution F. For real Borel functions $h(x_1, \ldots, x_k)$, the U-statistics with kernel h are defined by $S_n^{[k]}/\binom{n}{k}$ with

$$S_n^{[k]} = \sum_{(i_1,\ldots,i_k)\in\Lambda_n^{[k]}} h(X_{i_1}, \ldots, X_{i_k}), \tag{1.1}$$

where $\Lambda_n^{[k]} = \{(i_1, \ldots, i_k): 1 \leq i_1 < i_2 < \cdots < i_k \leq n\}$.

Our investigation is motivated by two problems. The first one is the order of $E\Phi(S_n^{[k]})$: given a nondecreasing nonnegative function Φ satisfying certain regularity conditions [e.g., $\Phi(x) = x^m$], find functionals $\mu_n(F, h, \Phi)$ such that

$$C'_{k,\Phi}\mu_n(F, h, \Phi) \leq E\Phi\big(S_n^{[k]}\big) \leq C''_{k,\Phi}\mu_n(F, h, \Phi) \quad \forall n \geq 1, \tag{1.2}$$

where $C'_{k,\Phi}$ and $C''_{k,\Phi}$ are universal constants. The second problem is the SLLN: given a sequence of normalizing constants $\{b_n\}$ satisfying certain regularity conditions (e.g., $b_n = n^{1/p}$ for some $0 < p < 2$), find necessary and

Received June 1996; revised August 1998.
[1]Supported in part by the National Science Foundation and National Security Agency.
AMS 1991 *subject classifications*. Primary 60F15; secondary 60G50.
Key words and phrases. Sub-Bernoulli function, strong law of large numbers, moment inequality, exponential inequality, tail probability, U-statistics.

sufficient conditions (nasc) on F and h for

$$S_n^{[k]}/b_n^k = b_n^{-k} \sum_{i_1<i_2<\cdots<i_k\le n} h(X_{i_1},\dots,X_{i_k}) \to 0 \quad \text{a.s.} \tag{1.3}$$

We find that the concept of *sub-Bernoulli* function, defined in (1.10) for $k=2$ in a special case and formally defined in Section 2.1, is very useful in our investigation of the preceding two problems and some additional problems. Basically, a nonnegative Borel function $\phi(x_1,\dots,x_k)$ is a sub-Bernoulli function of $(X_1,\dots,X_k)$ with parameter $(\theta_1,\dots,\theta_k)$ if its conditional expectations, given subsets of the X's, are no greater than the corresponding conditional expectations of a product of independent Bernoulli variables with the same parameters. We connect sub-Bernoulli functions to nonnegative kernels $h\ge 0$ through some normalizing kernels $\psi_n=\psi_n(x_1,\dots,x_k)$, positive Borel functions given in Section 3.1 for general k and in (1.9) and (1.11) for $k=2$, such that $\phi_n(x_1,\dots,x_k)=h/\psi_n$ are sub-Bernoulli functions with parameters $(k/n,\dots,k/n)$. It will be shown in Theorem 3.4 in Section 3.2 that, for all $h\ge 0$, nondecreasing nonnegative functions $g(\cdot)$ and integers $m\ge 1$ and $n\ge k$,

$$Eg(\xi_n^{[k]})\left(\frac{S_n^{[k]}}{\xi_n^{[k]}}\right)^m \le Eg(\xi_n^{[k]})E\binom{k+N_k}{k}^m, \tag{1.4}$$

where $\xi_n^{[k]}=\max_{(i_1,\dots,i_k)\in\Lambda_n^{[k]}}\psi_n(X_{i_1},\dots,X_{i_k})$ and N_k is a Poisson variable with $EN_k=k$. It will also be shown in Theorem 3.4 that for $n\ge k$,

$$P\{\xi_n^{[k]}>t\}\le C_k P\{S_n^{[k]}>t/2\}, \tag{1.5}$$

for some universal C_k. These inequalities provide crucial elements in our solutions to (1.2) and (1.3). Moment inequalities for sub-Bernoulli functions also imply an extension of the Bernstein inequality from $k=1$ to general k, Corollary 2.4 in Section 2.3, for decoupled symmetrized bounded kernels. The symmetrized, multisample and/or decoupled versions of the strong law and moment inequalities are also given. Some basic inequalities for sub-Bernoulli functions are provided for general independent (not necessarily identically distributed) variables.

The paper is organized as follows. In the rest of this section, we discuss in detail the case $k=2$, after giving our notation. In Section 2, we describe sub-Bernoulli functions and some basic inequalities. In Section 3, we consider inequalities of type (1.4) and (1.5). In Section 4, we provide the SLLN. Section 5 contains examples about moment conditions for SLLN.

1.2. *Notation.* Let $\{X^{(l)}, X_n^{(l)}, n\ge 1\}$ be independent sequences of iid random variables from possibly different distributions. The multisample version of (1.1) is defined by

$$\tilde S_{\mathbf n} = \sum_{i_1=1}^{n_1}\cdots\sum_{i_k=1}^{n_k} h(X_{i_1}^{(1)},\dots,X_{i_k}^{(k)}), \tag{1.6}$$

where $\mathbf{n} = (n_1, \ldots, n_k)$ gives the sample sizes. The normalized sums $\tilde{S}_{\mathbf{n}}/\prod_{l=1}^{k} n_l$ are the k-sample U-statistics with the kernel h. For $n_1 = \cdots = n_k = n$,

$$\tilde{S}_n^{(k)} = \sum_{(i_1,\ldots,i_k)\in\Lambda_n^k} h\big(X_{i_1}^{(1)}, \ldots, X_{i_k}^{(k)}\big), \tag{1.7}$$

where $\Lambda_n^k = \{(i_1, \ldots, i_k)\colon 1 \le i_l \le n, 1 \le l \le k\}$. When $\{X, X^{(1)}, \ldots, X^{(k)}\}$ are identically distributed, $\tilde{S}_n^{(k)}$ becomes the decoupled version $S_n^{[k]}$ in (1.1). Let $f(x, y)$ be a real Borel function with $f^2(x, y) = h(x, y)$, and $\{\varepsilon, \varepsilon_n, \varepsilon_n^{(l)}, n \ge 1, l \ge 1\}$ be iid Rademacher variables independent of $\{X_n, X_n^{(l)}, n \ge 1, l \ge 1\}$, $P\{\varepsilon = 1\} = P\{\varepsilon = -1\} = 1/2$. Define

$$T_n^{[k]} = \sum_{\mathbf{i}\in\Lambda_n^{[k]}} \varepsilon_{\mathbf{i}} f(X_{\mathbf{i}}), \qquad \tilde{T}_n^{(k)} = \sum_{\mathbf{i}\in\Lambda_n^k} \tilde{\varepsilon}_{\mathbf{i}} f(\tilde{X}_{\mathbf{i}}). \tag{1.8}$$

They are the symmetrized versions of (1.1) and (1.7). Here and in the sequel, $X_{\mathbf{i}} = (X_{i_1}, \ldots, X_{i_k})$, $\tilde{X}_{\mathbf{i}} = (X_{i_1}^{(1)}, \ldots, X_{i_k}^{(k)})$, $\varepsilon_{\mathbf{i}} = \prod_{l=1}^{k} \varepsilon_{i_l}$ and $\tilde{\varepsilon}_{\mathbf{i}} = \prod_{l=1}^{k} \tilde{\varepsilon}_{i_l}^{(l)}$ for $\mathbf{i} = (i_1, \ldots, i_k)$.

In addition to the variables introduced above, we shall use the following notation throughout. Let $\{Y_n, Y_n^{(l)}, n \ge 1, l \ge 1\}$ be independent variables, $\{\theta_n, \theta_n^{(l)}, n \ge 1, l \ge 1\}$ be constants in $[0, 1]$, and $\{\delta_n, \delta_n^{(l)}, n \ge 1, l \ge 1\}$ be independent Bernoulli ones with $P\{\delta_n^{(l)} = 1\} = \theta_n^{(l)}$ and $P\{\delta_n = 1\} = \theta_n$. For $\mathbf{i} = (i_1, \ldots, i_k)$, set $y_{\mathbf{i}} = (y_{i_1}, \ldots, y_{i_k})$, $Y_{\mathbf{i}} = (Y_{i_1}, \ldots, Y_{i_k})$ and $\tilde{Y}_{\mathbf{i}} = (Y_{i_1}^{(1)}, \ldots, Y_{i_k}^{(k)})$. Set $\theta_{\mathbf{i}} = (\theta_{i_1}, \ldots, \theta_{i_k})$ and $\tilde{\theta}_{\mathbf{i}} = (\theta_{i_1}^{(1)}, \ldots, \theta_{i_k}^{(k)})$. Set $\delta_{\mathbf{i}} = \prod_{l=1}^{k} \delta_{i_l}$ and $\tilde{\delta}_{\mathbf{i}} = \prod_{l=1}^{k} \delta_{i_l}^{(l)}$, in the same manner as $\varepsilon_{\mathbf{i}}$ and $\tilde{\varepsilon}_{\mathbf{i}}$ in (1.8). Let $N_{n,\theta}$ be a binomial variable with parameters (n, θ) and N_λ be a Poisson variable with $EN_\lambda = \lambda$. Also, for all real $\{a_i\}$, $\prod_{i\in A} a_i = 1$ if $A = \varnothing$.

The distributions of $X^{(l)}$, $l \ge 1$, are not assumed to be identical, to cover the multi sample case. The variables $\{Y_i, Y_i^{(l)}\}$ are not assumed to have the same distribution (neither between different i nor between different l), to be used to describe basic inequalities for sub-Bernoulli functions. The conditional expectations and moments of sub-Bernoulli functions are dominated by those of products of $\{\delta_i, \delta_i^{(l)}\}$, with nonidentical $\{\theta_i, \theta_i^{(l)}\}$ in general (between different l as well as different i). The Rademacher variables $\{\varepsilon, \varepsilon_n, \varepsilon_n^{(l)}, n \ge 1, l \ge 1\}$ are assumed to be independent of $\{X_n, X_n^{(l)}, Y_n, Y_n^{(l)}, \delta_n, \delta_n^{(l)}, n \ge 1, l \ge 1\}$ throughout. Also, $x_1 \vee \cdots \vee x_m = \max(x_1, \ldots, x_m)$, $x_1 \wedge \cdots \wedge x_m = \min(x_1, \ldots, x_m)$.

1.3. *Case* $k = 2$. Let us discuss $k = 2$ in more detail. Let $h(x, y) = h(y, x) \ge 0$. For $\theta > 0$ and $n \ge 1$, define

$$c_1(y; \theta) = \sup\left\{c > 0\colon E\left(\frac{h(X, y)}{h(X, y) \vee c}\right) \ge \theta\right\}, \qquad \sup\varnothing = 0,$$

$$c_0(\theta) = \sup\left\{c > 0\colon E\left(\frac{h(X_1, X_2)}{h(X_1, X_2) \vee c_1(X_1; \theta) \vee c_1(X_2; \theta) \vee c}\right) \ge \theta^2\right\}$$

and

$$\psi(x, y; \theta) = h(x, y) \vee c_1(x; \theta) \vee c_1(y; \theta) \vee c_0(\theta). \tag{1.9}$$

It can be easily seen (also cf. Lemma 3.1) that for $\phi = h(X_1, X_2)/\psi(X_1, X_2; \theta)$,

$$0 \le \phi \le 1, \ E[\phi \mid X_j] \le E[\delta_1\delta_2 \mid \delta_j], \ j = 1, 2, \qquad E\phi \le E[\delta_1\delta_2], \tag{1.10}$$

if δ_i are iid Bernoulli variables with mean θ. In other words, the conditional expectations of ϕ in (1.10) are dominated by their $\delta_1\delta_2$ versions. In this sense, we call ϕ a sub-Bernoulli function of (X_1, X_2).

The function $c_1(y; 1/n)$ as in (1.9) is the same as $m_n(y) = \sup\{m: nE\{h(X, y) \wedge m) \ge m\}$, a quantity whose essence has been used to approximate the center of the distribution of a sum of iid nonnegative random variables [each distribution in this case is $h(X, y)$]. In fact, Lemma 2.3 of Klass and Zhang (1994) shows that $P\{S_n(y) \ge c(y; 1/n)/3\} \ge 0.2$ and $P\{S_n(y) \le 3c(y; 1/n)\} \ge 0.3$ with $S_n(y) = \sum_{i=1}^n h(X_i, y)$. In this paper,

$$\begin{aligned} \xi_n^{[2]} &= \max_{1 \le i < j \le n} \psi(X_i, X_j; 2/n) \\ &= \max\Big\{\max_{i < j \le n} h(X_i, X_j), \max_{i \le n} c_1(X_i; 2/n), c_0(2/n)\Big\} \end{aligned} \tag{1.11}$$

are used to approximate the center and moments of (1.1) for $k = 2$. The maximum of $h(X_i, X_j)$ represents the extreme term; the maximum of $c_1(X_i; 2/n)$ represents the extreme term of $\sum_{j=1}^n h(X_i, X_j)$ in the sum over i; while $c_0(2/n)$ represents the overall center of the double sum. For more discussions, see Klass and Nowicki (1997).

It will be shown in Theorems 3.2 and 3.4 that (1.4) and its two-sample version

$$E\big(\tilde{S}_n^{(2)}/\tilde{\xi}_n^{(2)}\big)^m g\big(\tilde{\xi}_n^{(2)}\big) \le Eg\big(\tilde{\xi}_n^{(2)}\big)\big\{E(1 + N_1)^m\big\}^2, \tag{1.12}$$

and their extensions to general k, hold for all nondecreasing nonnegative functions g and $m \ge 1$, where $\tilde{\xi}_n^{(2)} = \max_{(i, j) \in \Lambda_n^2} \tilde{\psi}(X_i^{(1)}, X_j^{(2)}; 1/n)$ with $\tilde{\psi}$ being the $(X^{(1)}, X^{(2)})$-version of ψ (cf. Section 3.1). These theorems also assert that (1.5) and its two-sample version

$$P\big\{\tilde{\xi}_n^{(2)} > t\big\} \le 24P\big\{\tilde{S}_n^{(2)} \ge t/2\big\}, \tag{1.13}$$

and their extensions to general k, hold for all positive t. Let Φ be a function satisfying

$$\Phi(x) \uparrow \text{ in } x, \Phi(x) \ge 0, \Phi(cx) \le Mc^\alpha \Phi(x), \qquad c \ge c_* \ge 1, x \ge 0 \tag{1.14}$$

for some $\alpha > 0$. This includes $\Phi(x) = x^\alpha$. It follows from (1.12) and (1.13) that

$$C'_{M,\alpha} E\Phi(\tilde{\xi}_n^{(2)}) \le E\Phi(\tilde{S}_n^{(2)}) \le C''_{M,\alpha} E\Phi(\tilde{\xi}_n^{(2)}), \tag{1.15}$$

with $C'_{M,\alpha} = 1/(24M(2Vc_*)^\alpha)$ and $C''_{M,\alpha} = M[c_*^\alpha + \{E(1+N_1)^m\}^2]$, where $m-1 < \alpha \le m$. The upper bound above follows from (1.12) as $\Phi(\tilde{S}_n^{(2)}) \le M \max\{\{\tilde{S}_n^{(2)}/\tilde{\xi}_n^{(2)}\}^\alpha, 1\}\Phi(\tilde{\xi}_n^{(2)})$ and $c^\alpha \le c^m$ for $c \ge 1$. Inequality (1.15) for general k and its one-sample version [based on (1.4) and (1.5)] are given in Corollaries 3.3 and 3.5. Via different methods, Klass and Nowicki (1997) obtained (1.15) in the independent but non-iid case, using functions $h_{ij}(x, y) \ge 0$ in place of a fixed $h(x, y)$. Their results involved the construction of different constants.

Let $b_n = n^{1/p}$. Sufficient moment conditions for the SLLN (1.3) were given by Hoeffding (1961), Serfling (1980), Sen (1974), Teicher (1992) and Giné and Zinn (1992). By the Kolmogorov and Marcinkiewicz–Zygmund strong laws, (1.3) holds for $k = 1$ if and only if $Eh(X) = 0$ for $p \le 1$ and $E|h(X)|^p < \infty$. However, the case $k \ge 2$ is quite different. Giné and Zinn (1992) gave an example to exhibit that the condition $E|X|^p < \infty$ is not necessary for (1.3) when $h(x, y) = xy$. In Example 5.2 below, (1.3) holds for $b_n = n^{k/p}$ and some symmetric h but $E|h(X_1, \dots, X_k)|^{p_1+\varepsilon} = \infty$ for all $\varepsilon > 0$, where $0 < p < 2$ and $p_1 = p/\{k - p(k-1)/2\} < p$. For the special case $h(x, y) = xy$ and $EX = 0$ whenever $E|X| < \infty$, Cuzick, Giné and Zinn (1995) obtained nasc for the SLLN (1.3) under certain regularity conditions on the distribution of X (e.g., X symmetric, $P\{|X| > x\}$ regularly varying), and Zhang (1996) obtained nasc without regularity conditions on X. Some extensions of these results for $k > 2$ are also available in these papers. The SLLN in this paper give nasc for (1.3) for general nonnegative kernels and its symmetrized and/or multisample versions.

THEOREM 1.1. *Let $S_n^{[2]}$ and $\tilde{S}_n^{(2)}$ be as in* (1.1) *and* (1.7) *and $c_1(\cdot;\theta)$ and $c_0(\theta)$ be as in* (1.9). *Suppose $h(x, y) = h(y, x) \ge 0$, $\{X, X^{(1)}, X^{(2)}\}$ are identically distributed, and $\sup_{n\ge1} n^{-2} b_n^2 \sum_{m=n}^\infty m/b_m^2 < \infty$. Then the SLLN* (1.3) *holds iff $T_n^{[2]}/b_n \to 0$ a.s., iff $\xi_n^{[2]}/b_n^2 \to 0$ a.s., iff $\tilde{S}_n^{(2)}/b_n^2 \to 0$ a.s., iff $\tilde{T}_n^{(2)}/b_n \to 0$ a.s., iff $\tilde{\xi}_n^{(2)}/b_n^2 \to 0$ a.s., iff the following three conditions hold*:

$$c_0(1/n)/b_n^2 \to 0, \tag{1.16}$$

$$\sum_{n=1}^\infty P\left\{c_1(X; 1/n) > b_n^2\right\} < \infty, \tag{1.17}$$

$$\sum_{n=1}^\infty nP\left\{h(X_1, X_2) > b_n^2 \ge c_1(X_1; 1/n) \vee c_1(X_2; 1/n)\right\} < \infty. \tag{1.18}$$

REMARK. Condition (1.18) can be replaced by

$$\sum_{n=1}^\infty nP\left\{h(X_1, X_2) > b_n^2 \vee c_1(X_1; 1/n) \vee c_1(X_2; 1/n)\right\} < \infty. \tag{1.19}$$

The proof of Theorem 1.1 and its extensions for general k and multisample versions are given in Section 4.

2. Sub-Bernoulli functions. In the following three subsections, we shall (1) define sub-Bernoulli functions and describe the motivation, (2) provide upper bounds for conditional expectations of products and moments of sums of sub-Bernoulli functions and (3) provide some exponential inequalities.

2.1. *Sub-Bernoulli functions.* A random variable $\phi(Y_1, \dots, Y_k)$ is called a sub-Bernoulli function of a random vector $(Y_1, \dots, Y_k)$ with parameter $(\theta_1, \dots, \theta_k)$ if $0 \le \phi \le 1$ and for all $A \subseteq \{1, \dots, k\}$,

$$E\big[\phi(Y_1, \dots, Y_k) \big| Y_l, l \in A^c\big] \le \prod_{l \in A} \theta_l. \tag{2.1}$$

From this definition, sub-Bernoulli functions are nonnegative functions of $(Y_1, \dots, Y_k)$ whose conditional expectations given subsets of $\{Y_1, \dots, Y_k\}$ are uniformly bounded from above by the products of the θ's in the complementary subsets of $\{\theta_1, \dots, \theta_k\}$.

Consider the case of $k = 1$. By definition $\phi_i = \phi_i(Y_i)$ are sub-Bernoulli functions of Y_i iff $0 \le \phi_i \le 1$ and $E\phi_i \le \theta_i$. Such variables ϕ_i are dominated in moments by Bernoulli variables δ_i with $E\delta_i = \theta_i$ in the sense that $E\phi_i^m \le E\delta_i^m = \theta_i$ for all $m \ge 1$. Although this does not imply stochastic dominance (i.e., $P\{\phi_i > t\} \le P\{\delta_i > t\}$ may not hold for all t), it is strong enough to assure

$$\begin{aligned} E\bigg(\sum_{i=1}^n \phi_i\bigg)^m &= \sum_{(i_1,\dots,i_m)\in\Lambda_n^m} E\prod_{j=1}^m \phi_{i_j} \\ &\le \sum_{\Lambda_n^m} E\prod_{j=1}^m \delta_{i_j} = E\bigg(\sum_{i=1}^n \delta_i\bigg)^m \end{aligned} \tag{2.2}$$

for all integers $m \ge 1$. Thus, as far as moments of sums are concerned, Bernoulli variables are the *optimal* ones among all sub-Bernoulli variables.

We shall show below that products of independent Bernoulli variables are optimal for general k among all sub-Bernoulli functions.

2.2. *Expectations of products and moments of sums.*

PROPOSITION 2.1. *Suppose $\phi_{\mathbf{i}}(\tilde{Y}_{\mathbf{i}})$ and $\phi_{\mathbf{i},s}(\tilde{Y}_{\mathbf{i}})$, $1 \le s \le m$, are sub-Bernoulli functions of $\tilde{Y}_{\mathbf{i}} = (Y_{i_1}^{(1)}, \dots, Y_{i_k}^{(k)})$ with parameters $\tilde{\theta}_{\mathbf{i}} = (\theta_{i_1}^{(1)}, \dots, \theta_{i_k}^{(k)})$ for $\mathbf{i} = (i_1, \dots, i_k) \in \Lambda_\infty^k$. Then, for all $\mathbf{i}_s = (i_{1,s}, \dots, i_{k,s}) \in \Lambda_\infty^k$, $1 \le s \le m$, and $A \subseteq \Lambda_\infty^2$,*

$$E\bigg[\prod_{s=1}^m \phi_{\mathbf{i}_s}(\tilde{Y}_{\mathbf{i}_s}) \bigg| \tilde{\mathscr{F}}_{A^c}\bigg] \le \prod_{s=1}^m \prod_{l \in A_s} \theta_{i_{l,s}}^{(l)} = E\bigg[\prod_{s=1}^m \prod_{l=1}^k \delta_{i_{l,s}}^{(l)} \bigg| \tilde{\delta}^* = 1\bigg], \tag{2.3}$$

where $\tilde{\mathscr{F}}_A = \sigma(Y_i^{(l)}, (l, i) \in A)$, $A_s = \{l \le k\colon i_{l,s} \ne i_{l,t}\ \forall s < t \le m, (l, i_{l,s}) \in A\}$, $A^c = \Lambda_\infty^2 \setminus A$, and $\tilde{\delta}^* = \prod\{\delta_{i_{l,s}}^{(l)}\colon (l, i_{l,s}) \notin A\}$. Moreover, for all $\Lambda_s \subseteq \Lambda_\infty^k$,

$1 \le s \le m$,

$$E \prod_{s=1}^{m} \biggl(\sum_{\mathbf{i} \in \Lambda_s} \phi_{\mathbf{i},s}(\tilde{Y}_{\mathbf{i}}) \biggr) \le E \prod_{s=1}^{m} \biggl(\sum_{\mathbf{i} \in \Lambda_s} \prod_{l=1}^{k} \delta_{i_l}^{(l)} \biggr). \tag{2.4}$$

Consequently, for $\Lambda_{\mathbf{n}} = \prod_{l=1}^{k} \Lambda_{n_l}^1$ *with* $\mathbf{n} = (n_1, \dots, n_k)$,

$$E\biggl(\sum_{\mathbf{i} \in \Lambda_{\mathbf{n}}} \phi_{\mathbf{i}}(\tilde{Y}_{\mathbf{i}}) \biggr)^m \le \prod_{l=1}^{k} E\biggl(\sum_{i=1}^{n_l} \delta_i^{(l)} \biggr)^m. \tag{2.5}$$

REMARK. Since $\{Y_i^{(l)}\}$ are independent (between different l as well as different i), the indeces $\mathbf{i}$ are allowed to have ties.

PROOF. Set $\phi_{\mathbf{i}} = \phi_{\mathbf{i}}(\tilde{Y}_{\mathbf{i}})$. By (2.1),

$$E\bigl[\phi_{\mathbf{i}_s} \big| \tilde{Y}_{\mathbf{i}_t}, s < t \le m, \tilde{\mathscr{F}}_{A^c}\bigr] = E\bigl[\phi_{\mathbf{i}_s} \big| Y_{i_{l,s}}^{(l)}, l \notin A_s\bigr] \le \prod_{l \in A_s} \theta_{i_{l,s}}^{(l)}.$$

Repeated applications of this inequality for $s = 1, \dots, m$ give the inequality in (2.3). The identity in (2.3) follows from

$$\begin{aligned} &P\bigl\{\tilde{\delta}_{\mathbf{i}_s} = 1 \big| \tilde{\delta}_{\mathbf{i}_t} = 1, s < t \le m, \tilde{\delta}^* = 1\bigr\} \\ &\qquad = P\bigl\{\tilde{\delta}_{\mathbf{i}_s} = 1 \big| \delta_{i_{l,s}}^{(l)} = 1, l \notin A_s\bigr\} = \prod_{l \in A_s} \theta_{i_{l,s}}^{(l)}. \end{aligned}$$

Finally, similarly to (2.2), (2.4) is proved by first writing the product of sums as sum of products and then applying (2.3) with $A = \Lambda_\infty^2$ (trivial $\tilde{\mathscr{F}}_{A^c}$) to each (product) term in the sum to allow substitution of $\phi_{\mathbf{i},s}$ by $\tilde{\delta}_{\mathbf{i}}$. □

For the single sequence $\{Y_n\}$, we have the following analogous result.

PROPOSITION 2.2. *Suppose* $\phi_{\mathbf{i}}(Y_{\mathbf{i}})$ *and* $\phi_{\mathbf{i},s}(Y_{\mathbf{i}})$, $1 \le s \le m$, *are sub-Bernoulli functions of* $Y_{\mathbf{i}}$ *with parameters* $\theta_{\mathbf{i}}$ *for* $\mathbf{i} = (i_1, \dots, i_k) \in \Lambda_\infty^{[k]}$. *Then, for all* $\mathbf{i}_s = (i_{1,s}, \dots, i_{k,s}) \in \Lambda_\infty^{[k]}$, $1 \le s \le m$ *and* $A \subseteq \Lambda_\infty^1$,

$$E\biggl[\prod_{s=1}^{m} \phi_{\mathbf{i}_s}(Y_{\mathbf{i}_s}) \bigg| \mathscr{F}_{A^c} \biggr] \le \prod_{s=1}^{m} \prod_{l \in A_s} \theta_{i_{l,s}} = E\biggl[\prod_{s=1}^{m} \prod_{l=1}^{k} \delta_{i_{l,s}} \bigg| \delta^* = 1 \biggr], \tag{2.6}$$

where $A_s = \{l \le k\colon i_{l,s} \ne i_{j,t}\ \forall s < t \le m \text{ and } 1 \le j \le k, i_{l,s} \in A\}$, $\mathscr{F}_A = \sigma(Y_i, i \in A)$, $A^c = \Lambda_\infty^1 \setminus A$, *and* $\delta^* = \prod\{\delta_{i_{l,s}}\colon i_{l,s} \notin A\}$. *Moreover, for all* $\Lambda_s \subseteq \Lambda_\infty^{[k]}$, $1 \le s \le m$,

$$E \prod_{s=1}^{m} \biggl(\sum_{\mathbf{i} \in \Lambda_s} \phi_{\mathbf{i},s}(Y_{\mathbf{i}}) \biggr) \le E \prod_{s=1}^{m} \biggl(\sum_{\mathbf{i} \in \Lambda_s} \prod_{l=1}^{k} \delta_{i_l} \biggr). \tag{2.7}$$

Consequently, with $T_n = \sum_{i=1}^n \delta_i$,

$$E\biggl(\sum_{\mathbf{i}\in\Lambda_n^{[k]}} \phi_{\mathbf{i}}(Y_{\mathbf{i}})\biggr)^m \le E\binom{T_n}{k}^m. \tag{2.8}$$

REMARK. The symmetrized versions of (2.5) and (2.8) can be easily produced using the Khintchine inequality.

REMARK. For all nondecreasing nonnegative g,

$$EN_{n,\theta}\, g(N_{n,\theta}) \le^m \theta E g(N_{n,\theta}+1)$$

and

$$EN_\lambda g(N_\lambda) = \lambda E g(N_\lambda + 1).$$

These and Corollary 2.1 of Gleser (1975) imply

$$ET_n^m \le EN_{n,\bar\theta_n}^m \le EN_{\lambda_n}^m, \qquad E\binom{T_n}{j}^m \le E\binom{N_{n,\bar\theta_n}}{j}^m \le E\binom{N_{\lambda_n}}{j}^m,$$

where $\bar\theta_n = \lambda_n/n = \sum_{i=1}^n \theta_i/n$ and T_n is as in (2.8). Thus, the T_n in (2.8) and the sums on the right-hand side of (2.5) can be replaced by Poisson variables.

The proof of Proposition 2.2 is omitted as it is nearly identical to the proof of Proposition 2.1

2.3. *Exponential inequalities.* There are several ways of obtaining exponential inequalities for the tail probabilities of U-statistics from moment inequalities. Here we shall only present one for symmetrized and decoupled U-statistics.

PROPOSITION 2.3. *Let* $0 < \theta \le 1$. *Suppose* $\phi_{\mathbf{i}}(\tilde Y_{\mathbf{i}}) = f_{\mathbf{i}}^2(\tilde Y_{\mathbf{i}})$ *are sub-Bernoulli functions of* $\tilde Y_{\mathbf{i}}$ *with a common parameter* $(\theta,\dots,\theta)$ *for all* $\mathbf{i} = (i_1,\dots,i_k)\in\Lambda_n^k$. *Then,*

$$E\exp\biggl(t\biggl|\sum_{\mathbf{i}\in\Lambda_n^k}\frac{\tilde\varepsilon_{\mathbf{i}} f_{\mathbf{i}}(\tilde Y_{\mathbf{i}})}{n^{k/2}}\biggr|^{1/k}\biggr) \le 2E\exp\bigl(t^2 N_{n,\theta}/(2n)\bigr), \tag{2.9}$$

where $N_{n,\theta}$ *is binomial* (n,θ). *Consequently,*

$$P\biggl\{\biggl|\sum_{\mathbf{i}\in\Lambda_n^k}\frac{\tilde\varepsilon_{\mathbf{i}} f_{\mathbf{i}}(\tilde Y_{\mathbf{i}})}{n^{k/2}}\biggr|^{1/k} \ge t\biggr\} \le 2\exp\biggl(\frac{-t^2/2}{\theta + t/(2\sqrt{n})}\biggr). \tag{2.10}$$

COROLLARY 2.4. *Suppose* $\|f_{\mathbf{i}}(\tilde Y_{\mathbf{i}})\|_\infty \le c$ *and* $Ef_{\mathbf{i}}^2(\tilde Y_{\mathbf{i}}) \le \sigma^2$ *for all* $\mathbf{i} = (i_1,\dots,i_k)\in\Lambda_n^k$. *Then*

$$P\biggl\{\biggl|\sum_{\mathbf{i}\in\Lambda_n^k}\frac{\tilde\varepsilon_{\mathbf{i}} f_{\mathbf{i}}(\tilde Y_{\mathbf{i}})}{n^{k/2}}\biggr|^{1/k} \ge t\biggr\} \le 2\exp\biggl(\frac{-t^2/2}{\sigma^{2/k} + tc/(2\sigma^{(k-1)/k}\sqrt{n})}\biggr). \tag{2.11}$$

For $k=1$, (2.11) becomes the Bernstein inequality. For iid $\{X_n\}$ and completely degenerate kernels f with $\|f\|_\infty \le c$ and $Ef^2 \le \sigma^2$, Arcones and Giné [(1993), page 1501] obtained the inequality

$$P\left\{\left|\sum_{\mathbf{i}\in\Lambda_n^k}\frac{f(X_{\mathbf{i}})}{n^{k/2}}\right|^{1/k} \ge t\right\} \le c_k'' \exp\left(\frac{-c_k' t^2}{\sigma^{2/k} + (tc/\sqrt{n})^{2/(k+1)}}\right)$$

and its symmetrized and/or decoupled versions with implicitly specified universal constants c_k' and c_k''. Their inequalities give smaller upper bounds for $\sigma^{(k+1)/k}\sqrt{n}/(ct) = o(1)$ than (2.11) although the breakdown point $t = \sigma^{(k+1)/k}\sqrt{n}/c$ is the same. For related exponential inequalities for the Rademacher chaos, we refer to Ledoux and Talagrand (1991).

PROOF OF PROPOSITION 2.3. Set $\tilde{T}_n = \sum_{\mathbf{i}\in\Lambda_n^k}\tilde{\varepsilon}_{\mathbf{i}} f_{\mathbf{i}}(\tilde{Y}_{\mathbf{i}})$ and $\tilde{S}_n = \sum_{\mathbf{i}\in\Lambda_n^k} f_{\mathbf{i}}^2(\tilde{Y}_{\mathbf{i}})$. Let Z be a $N(0,1)$ variable independent of $N_{n,\theta}$. By the Khintchine inequality and (2.5) of Proposition 2.1,

$$E(\tilde{T}_n)^{2m} \le \left(EZ^{2m}\right)^k E(\tilde{S}_n)^m \le \left(EZ^{2m} EN_{n,\theta}^m\right)^k.$$

By the Jensen inequality $E|\tilde{T}_n|^{2m/k} \le E\left(Z\sqrt{N_{n,\theta}}\right)^{2m}$. Since $e^{|x|} \le e^x + e^{-x}$, the left-hand side of (2.9) is bounded by

$$\begin{aligned} E\left\{2\sum_{m=0}^{\infty}\frac{(\lambda|\tilde{T}_n|^{1/k})^{2m}}{(2m)!}\right\} &\le 2\sum_{m=0}^{\infty}\frac{\lambda^{2m}}{(2m)!}E\left(Z\sqrt{N_{n,\theta}}\right)^{2m} \\ &= 2E\exp\left(\lambda Z\sqrt{N_{n,\theta}}\right) = 2E\exp\left(t^2 N_{n,\theta}/(2n)\right), \end{aligned}$$

with $\lambda = t/\sqrt{n}$. Thus, (2.9) holds.

The proof of (2.10) from (2.9) is nearly identical to the proof of the Bernstein inequality in Chow and Teicher [(1988), page 111]. By (2.9) and the Markov inequality,

$$P\left\{\left|\sum_{\mathbf{i}\in\Lambda_n^k}\frac{\tilde{\varepsilon}_{\mathbf{i}} f_{\mathbf{i}}(\tilde{Y}_{\mathbf{i}})}{n^{k/2}}\right|^{1/k} \ge t\right\} \le 2\exp(-\lambda t)\Bigl(1+\theta\bigl\{\exp(\lambda^2/(2n))-1\bigr\}\Bigr)^n \tag{2.12}$$

for all $\lambda > 0$. Take $\lambda = t/\{\theta + t/(2\sqrt{n})\}$. Since $\lambda^2/(2n) < 2$ and $e^x - 1 \le x/(1-x/2)$ for $0 \le x < 2$,

$$\theta\bigl\{\exp(\lambda^2/(2n))-1\bigr\} \le \frac{\theta\lambda^2/(2n)}{1-\lambda^2/(4n)} \le \frac{\theta\lambda^2/(2n)}{1-\lambda/(2\sqrt{n})} = \frac{t\lambda}{2n}.$$

Thus, the right-hand side of (2.12) is bounded by $2\exp(-t\lambda)\{1+t\lambda/(2n)\}^n \le 2\exp(-t\lambda/2)$ and the proof is complete. □

PROOF OF COROLLARY 2.4. Let $\theta = \sigma^2/c^2 \le 1$ and $\phi_{\mathbf{i}} = \phi_{\mathbf{i}}(\tilde{Y}_{\mathbf{i}}) = \theta^{k-1}c^{-2}\cdot f_{\mathbf{i}}^2(\tilde{Y}_{\mathbf{i}})$. Then $\phi_{\mathbf{i}}$ are sub-Bernoulli functions of $\tilde{Y}_{\mathbf{i}}$ with parameter $(\theta,\dots,\theta)$, as the $\phi_{\mathbf{i}}$ version of (2.1) holds for $j>0$ due to $\|\phi_{\mathbf{i}}\|_\infty \le \theta^{k-1} \le \theta^{k-j}$ and for

$j = 0$ due to $E\phi_{\mathbf{i}} \le \theta^k$. Set $\lambda = (\sqrt{\theta}c)^{1/k}/\sqrt{\theta}$. It follows from (2.10) that the left-hand side of (2.11) is bounded by

$$P\left\{\left|\frac{\theta^{k/2}}{\sqrt{\theta}c}\sum_{\mathbf{i}\in\Lambda_n^k}\frac{\tilde{\varepsilon}_{\mathbf{i}} f_{\mathbf{i}}(\tilde{Y}_{\mathbf{i}})}{n^{k/2}}\right|^{1/k} \ge \frac{t}{\lambda}\right\} \le 2\exp\left(\frac{-t^2/2}{\lambda^2\theta + \lambda t/(2\sqrt{n})}\right).$$

Hence, (2.11) holds as $\lambda^2\theta = (\theta c^2)^{1/k} = \sigma^{2/k}$ and $\lambda = \sigma^{1/k}/\sqrt{\theta} = c/\sigma^{(k-1)/k}$. □

3. Moments of maxima and sums. In this section we provide moment and tail probability inequalities for maxima and sums of products [e.g., (1.4), (1.5), (1.12) and (1.13)]. We shall provide the normalizing kernels in Section 3.1, the inequalities in the iid and multisample cases in Section 3.1 and inequalities for independent not identically distributed variables in Section 3.3. Section 3.4 contains the proofs of Theorems 3.2 and 3.4 in Section 3.2.

We shall use the following notation to shorten expressions: $\Omega_k = \{1, \dots, k\}$, $a_{(j)} = (a_1, \dots, a_j)$, and for all $a_{(k)}$ and $A \subseteq \Omega_k$ with size $|A| = j$, $(a_l\colon l \in A) = (a_{l_1}, \dots, a_{l_j})$ with $l_1 < \dots l_j$ being the ordered labels in A.

3.1. *Construction of normalizing kernels.* Let $h(y_{(k)})$ be a nonnegative Borel function and $Y_{(k)} = (Y_1, \dots, Y_k)$ be a random vector with joint distribution $F_{Y_{(k)}}$. Given $\theta_{(k)} = (\theta_1, \dots, \theta_k)$, we shall find a normalizing kernel $\psi(y_{(k)})$ such that $\phi(y_{(k)}) = h(y_{(k)})/\psi(y_{(k)})$ is a sub-Bernoulli function of $Y_{(k)}$ with parameter $\theta_{(k)}$. In addition, the normalizing kernels should be small enough to be used in the proof of inequalities in both directions such as (1.4) and (1.5).

We shall classify the 2^k inequalities of (2.1) according to $|A|$, the size of A, and consider those with $|A| = k - j$, $j = k, \dots, 0$. The normalizing kernel is defined by

$$\psi(y_{(k)}) = \psi(y_{(k)}; \theta_{(k)}, h, F_{Y_{(k)}}) = \max_{0\le j\le k} h_j(y_{(k)}) = h_0(y_{(k)}), \tag{3.1}$$

$$h_j(y_{(k)}) = h_{j+1}(y_{(k)}) \vee \left\{\max_{|A|=k-j, A\subseteq\Omega_k} c_{j,A}(y_l\colon l \notin A)\right\}, \tag{3.2}$$

$j = k-1, \dots, 0$, $h_k(y_{(k)}) = h(y_{(k)}) = c_{k,\varnothing}(y_{(k)})$, and for $A \subseteq \Omega_k$ and $|A| = k - j$,

$$\begin{aligned} &c_{j,A}(y_l\colon l \in A^c) \\ &\quad= c_{j,A}(y_l\colon l \in A^c; \theta_{(k)}, h, F_{Y_{(k)}}) \\ &\quad= \inf\left\{c \ge 0\colon E\left[\frac{h(Y_{(k)})}{h_{j+1}(Y_{(k)}) \vee c}\middle| Y_l = y_l, l \in A^c\right] \le \prod_{l\in A}\theta_l\right\}. \end{aligned} \tag{3.3}$$

Note that c_0 does not depend on $y_{(k)}$. Define $\phi_{j,A}(y_{(k)}) = \phi_{j,A}(y_{(k)}; \theta_{(k)}, h, F_{Y_{(k)}})$ by

$$\phi_{j,A}(y_{(k)}) = \frac{h(y_{(k)})}{h_{j+1}(y_{(k)}) \vee c_{j,A}(y_l: l \notin A)}. \tag{3.4}$$

LEMMA 3.1. *Let $\theta_{(k)}$, h and $F_{Y_{(k)}}$ be as in (3.1)–(3.4). For all fixed $(y_l: l \notin A)$, the function $\phi_{j,A}(y_{(k)})$ in (3.4) is a sub-Bernoulli function of $(Y_l: l \in A)$ with the parameter $(\theta_l: l \in A)$, and $E[\phi_{j,A}(Y_{(k)}) \,|\, Y_l, l \notin A] = \prod_{l \in A} \theta_l$ for $c_{j,A}(Y_l: l \notin A) > 0$. In particular, $\phi(y_{(k)}) = h/\psi = \phi_{0,\Omega_k}(y_{(k)})$ is a sub-Bernoulli function of $Y_{(k)}$ with the parameter $\theta_{(k)}$.*

PROOF. It immediately follows from (2.1), (3.2) and (3.3) that $\phi_{j,A}(y_{(k)})$ is conditionally sub-Bernoulli. Its mean is given by (3.3) as the conditional expectation on the right-hand side is continuous in c for $c > 0$. □

3.2. *Moment inequalities in the iid and multisample cases.* We shall provide moment inequalities involving maxima and sums, extended (1.4) and (1.5) with iid $\{X_i\}$ and their multisample versions with independent iid sequences $\{X_i^{(l)}\}$, $l \geq 1$.

Let $h(x_1, \ldots, x_k)$ be a fixed kernel. For $\Lambda_n \subseteq \Lambda_n^{[k]}$ and $\tilde{\Lambda}_n \subseteq \otimes_{l=1}^k \Lambda_{n_l}^1$ define

$$S_{\Lambda_n} = \sum_{\mathbf{i} \in \Lambda_n} h(X_{\mathbf{i}}), \qquad \tilde{S}_{\tilde{\Lambda}_n} = \sum_{\mathbf{i} \in \tilde{\Lambda}_n} h(\tilde{X}_{\mathbf{i}}). \tag{3.5}$$

Let $\theta_{(k)} = (\theta_1, \ldots, \theta_k)$ and θ be fixed parameters. Let $F_{\tilde{X}}$ be the joint distribution of $\tilde{X} = (X^{(1)}, \ldots, X^{(k)})$ and $F_{X_{(k)}}$ be the joint distribution of $X_{(k)} = (X_1, \ldots, X_k)$. Define

$$\xi_{\Lambda_n, \theta} = \max_{\mathbf{i} \in \Lambda_n} \psi\big(X_{\mathbf{i}}; (\theta, \ldots, \theta), h, F_{X_{(k)}}\big), \qquad \xi_n^{[k]} = \xi_{\Lambda_n^{[k]}, k/n}, \tag{3.6}$$

$$\tilde{\xi}_{\tilde{\Lambda}_n, \theta_{(k)}} = \max_{\mathbf{i} \in \tilde{\Lambda}_n} \psi\big(\tilde{X}_{\mathbf{i}}; \theta_{(k)}, h, F_{\tilde{X}}\big) \tag{3.7}$$

and

$$\tilde{\xi}_{\mathbf{n}} = \tilde{\xi}_{\otimes_{l=1}^k \Lambda_{n_l}^1, (1/n_1, \ldots, 1/n_k)}, \qquad \tilde{\xi}_n^{(k)} = \tilde{\xi}_{\Lambda_n^k, (1/n, \ldots, 1/n)}, \tag{3.8}$$

where $\psi(\cdot)$ is given by (3.1).

THEOREM 3.2. *Let g be an nondecreasing nonnegative function. Then, for all $\tilde{\Lambda}_n \subseteq \otimes_{l=1}^k \Lambda_{n_l}^1$ and integers $m \geq 1$,*

$$E\big(\tilde{S}_{\tilde{\Lambda}_n} / \tilde{\xi}_{\tilde{\Lambda}_n, \theta_{(k)}}\big)^m g\big(\tilde{\xi}_{\tilde{\Lambda}_n, \theta_{(k)}}\big) \leq E g\big(\tilde{\xi}_{\tilde{\Lambda}_n, \theta_{(k)}}\big) \prod_{l=1}^k E\big(1 + N_{n_l, \theta_l}\big)^m. \tag{3.9}$$

Furthermore, for all $\mathbf{n} = (n_1, \dots, n_k)$ *and real numbers* $t > 0$ *and* $0 < \varepsilon < 1$,

$$P\{\tilde{\xi}_{\mathbf{n}} > t\} \le P\{\tilde{S}_{\mathbf{n}} > \varepsilon t\}\left(\frac{3^k - 2^k}{(1-\varepsilon)^2} + 2^k\right). \tag{3.10}$$

As in (1.15), we have the following corollary.

COROLLARY 3.3. *Let* Φ *be a function satisfying* (1.14) *for some* $\alpha \le m$. *Then*,

$$\frac{(M(2Vc_*)^\alpha)^{-1}E\Phi(\tilde{\xi}_{\mathbf{n}})}{((3^k - 2^k)/4 + 2^k)} \le E\Phi(\tilde{S}_{\mathbf{n}}) \le M\left(c_*^\alpha + \left\{E(1+N_1)^m\right\}^k\right)E\Phi(\tilde{\xi}_{\mathbf{n}}).$$

THEOREM 3.4. *Let* g *be an nondecreasing nonnegative function. Then, for all* $\Lambda_n \subseteq \Lambda_n^{[k]}$ *and integers* $m \ge 1$,

$$E\left(S_{\Lambda_n}/\xi_{\Lambda_n,\theta}\right)^m g\left(\xi_{\Lambda_n,\theta}\right) \le Eg\left(\xi_{\Lambda_n,\theta}\right)E\binom{k+N_{n,\theta}}{k}^m. \tag{3.11}$$

Furthermore, there exists a function $C_{k,\varepsilon}$ *such that for all* $n \ge k$ *and real numbers* $t > 0$ *and* $0 < \varepsilon < 1$,

$$P\left\{\xi_n^{[k]} > t\right\} \le C_{k,\varepsilon}P\left\{S_n^{[k]} > \varepsilon t\right\}. \tag{3.12}$$

For $n \ge k(k+1)$, (1.13) *holds with*

$$C_{k,\varepsilon} = \sum_{j=0}^{k}\binom{k+1}{j}\left\{1 + \frac{E\binom{N_{k+1-j}}{k-j}^2 - 1}{(1-\varepsilon)^2}\right\}.$$

REMARK. If n is a multiplier of $k!$, (3.12) holds for

$$C_{k,\varepsilon} = \sum_{j=0}^{k}\binom{k}{j}\left\{1 + \frac{E\binom{N_{k-j}}{k-j}^2 - 1}{(1-\varepsilon)^2}\right\}.$$

COROLLARY 3.5. *Let* Φ *be a function satisfying* (1.14) *for some* $\alpha \le m$. *Then*,

$$\frac{E\Phi(\xi_n^{[k]})}{M(2Vc_*)^\alpha C_{k,1/2}} \le E\Phi(S_n^{[k]}) \le ME\Phi\left(\xi_n^{[k]}\right)\left\{c_*^\alpha + E\binom{k+N_k}{k}^m\right\}.$$

Theorems 3.2 and 3.4 are proved in Section 3.4.

3.3. *General independent variables.* In this section, we consider the expectations of the product of a maximum and several sums for independent variables $\{Y_i, Y_i^{(l)}, i \ge 1, l \ge 1\}$, which may have different distributions.

PROPOSITION 3.6. *Suppose* $\psi_{\mathbf{i}}(\tilde{Y}_{\mathbf{i}})$ *are measurable functions of* $\tilde{Y}_{\mathbf{i}}$ *and* $\phi_{\mathbf{i},s}(\tilde{Y}_{\mathbf{i}})$ *are sub-Bernoulli ones with parameters* $\tilde{\theta}_{\mathbf{i}}$ *for* $\mathbf{i} \in \Lambda_\infty^k$. *Let* $\Lambda^{k,s} =$

$\otimes_{l=1}^k I_s^{(l)}$ *for some* $I_s^{(l)} \subseteq \Omega_\infty^1$. *Then, for all nondecreasing nonnegative functions* $g(\cdot)$, $\Lambda \subseteq \Lambda_\infty^k$ *and* $m \geq 1$,

$$E g(\tilde{\xi}_\Lambda) \prod_{s=1}^m \biggl(\sum_{\mathbf{i} \in \Lambda^{k,s}} \phi_{\mathbf{i},s}(\tilde{Y}_\mathbf{i}) \biggr) \leq E g(\tilde{\xi}_\Lambda) \prod_{l=1}^k E \prod_{s=1}^m \{1 + T_{l,s}\}, \tag{3.13}$$

where $\tilde{\xi}_\Lambda = \max_{\mathbf{i}\in\Lambda} \psi_\mathbf{i}(\tilde{Y}_\mathbf{i})$ *and* $T_{l,s} = \sum_{i \in I_s^{(l)}}^n \delta_i^{(l)}$.

PROPOSITION 3.7. *Suppose* $\psi_\mathbf{i}(Y_\mathbf{i})$ *are measurable functions of* $Y_\mathbf{i}$ *and* $\phi_{\mathbf{i},s}(Y_\mathbf{i})$ *are sub-Bernoulli ones with parameters* $\theta_\mathbf{i}$ *for* $\mathbf{i} \in \Lambda_\infty^{[k]}$. *Let* $\Lambda^{[k,s]} = \{I_s^{\otimes k}\} \cap \Lambda_\infty^{[k]}$ *for some* $I_s \subseteq \Omega_\infty^1$. *Then, for all nondecreasing nonnegative functions* $g(\cdot)$, $\Lambda \subseteq \Lambda_\infty^{[k]}$ *and* $m \geq 1$,

$$E g(\xi_\Lambda) \prod_{s=1}^m \biggl(\sum_{\mathbf{i} \in \Lambda^{[k,s]}} \phi_{\mathbf{i},s}(Y_\mathbf{i}) \biggr) \leq E g(\xi_\Lambda) E \prod_{s=1}^m \binom{k + T_s}{k}, \tag{3.14}$$

where $\xi_\Lambda = \max_{\mathbf{i}\in\Lambda} \psi_\mathbf{i}(Y_\mathbf{i})$ *and* $T_s = \sum_{i \in I_s} \delta_i$.

REMARK. For general $\Lambda_n \subseteq \{\mathbf{i} \in \Lambda_n^k : i_j \neq i_l \ \forall\ j \neq l\}$, we may apply Proposition 3.7 to $\overline{\phi}_\mathbf{i}(Y_\mathbf{i}) = \sum_{\{\mathbf{i}'\}=\{\mathbf{i}\}} \phi_{\mathbf{i}'}(Y_{\mathbf{i}'})/k!$ and $\overline{\psi}_\mathbf{i}(Y_\mathbf{i}) = \max\{\psi_{\mathbf{i}'}(Y_{\mathbf{i}'}) : \mathbf{i}' \in \Lambda_n, \{\mathbf{i}'\} = \{\mathbf{i}\}\}$, where $\{\mathbf{i}\} = \{i_1, \ldots, i_k\}$ is regarded as a set.

PROOF OF PROPOSITION 3.6. Set $\psi_\mathbf{i} = \psi_\mathbf{i}(\tilde{Y}_\mathbf{i})$ and $\phi_{\mathbf{i},s} = \phi_{\mathbf{i},s}(\tilde{Y}_\mathbf{i})$. Before providing the full proof of (3.13), we shall first take a look at the case where $k = 1$ and $m = 2$. Let (ψ_n, ϕ_n), $n \geq 1$, be independent random vectors with $0 \leq \phi_n \leq 1$ and $E\phi_n = \theta_n$. Let i^* be the index at which $\max_{i \leq n} g(\psi_i)$ is reached. We have

$$\begin{aligned} E\Bigl(\max_{i\leq n} g(\psi_i)\Bigr)\biggl(\sum_{j_1=1}^{n_1} \phi_{j_1}\biggr)\biggl(\sum_{j_1=1}^{n_2} \phi_{j_2}\biggr) &\\ \leq E g(\psi_{i^*})\phi_{i^*}^2 + \sum_{j_1=1}^{n_1} E g(\psi_{i^*})\phi_{i^*}\phi_{j_1} I_{\{j_1 \neq i^*\}} &\\ + \sum_{j_2=1}^{n_2} E g(\psi_{i^*})\phi_{i^*}\phi_{j_2} I_{\{j_2 \neq i^*\}} &\\ + \sum_{j_1=1}^{n_1} \sum_{j_1=1}^{n_2} E g(\psi_{i^*})\phi_{j_1}\phi_{j_2} I_{\{j_1 \neq i^*,\, j_2 \neq i^*\}}. & \end{aligned} \tag{3.15}$$

Since $g(\cdot)$ is nondecreasing and nonnegative,

$$\begin{aligned} E g(\psi_{i^*})\phi_{j_1}\phi_{j_2} I_{\{j_1 \neq i^*,\, j_2 \neq i^*\}} &\leq E\Bigl\{\max_{i\neq j_1,\, i \neq j_2,\, i\leq n} g(\psi_i)\Bigr\} E\phi_{j_1}\phi_{j_2} \\ &\leq E g(\psi_{i^*}) E\delta_{j_1}\delta_{j_2}. \end{aligned} \tag{3.16}$$

Similarly, $Eg(\psi_{i^*})\phi_{i^*}\phi_j I_{\{j\neq i^*\}} \le Eg(\psi_{i^*})E\delta_j$ and $Eg(\psi_{i^*})\phi_{i^*}^2 \le Eg(\psi_{i^*})$ as $0 \le \phi_j \le 1$. Inserting these inequalities into (3.15), we obtain

$$
\begin{aligned}
&Eg(\psi_{i^*})\bigg(\sum_{j_1=1}^{n_1}\phi_{j_1}\bigg)\bigg(\sum_{j_1=1}^{n_2}\phi_{j_2}\bigg)\\
&\quad\le Eg(\psi_{i^*})\big(1+ET_{n_1}+ET_{n_2}+ET_{n_1}T_{n_2}\big)\\
&\quad\le Eg(\psi_{i^*})E\big(1+T_{n_1}\big)\big(1+T_{n_2}\big),
\end{aligned}
\tag{3.17}
$$

where $T_n = \sum_{i=1}^n \delta_i$. This is (3.13) for $k=1$ and $m=2$.

To proof (3.13) with general k and m, we shall compare the indices $\mathbf{i}_s = (i_{1,s},\dots,i_{k,s})$, $1\le s\le m$, with $\mathbf{i}^* = (i_1^*,\dots,i_k^*)$ in products of the form $g(\psi_{\mathbf{i}^*})\prod_{s=1}^m \phi_{\mathbf{i}_s,s}$, where $\mathbf{i}^*\in\Lambda$ is the index at which the maximum $\tilde{\xi}_\Lambda$ is reached. Let $\Omega = \{(l,s)\colon 1\le l\le k, 1\le s\le m\}$. Given $A\subseteq\Omega$ and $(\mathbf{i}_1,\dots,\mathbf{i}_m)$, define $\mathbf{i}_A = (i_{l,s}\colon (l,s)\in A)$ as an $|A|$-dimensional vector of positive integers. Set $\tilde{\Lambda}_A = \otimes_{(l,s)\in A} I_s^{(l)}$, which is the space of combined labels in A. Define

$$
\pi_A = \pi_A(\mathbf{i}_1,\dots\mathbf{i}_m;\mathbf{i}^*) = I\big\{i_{l,s} = i_l^* \;\Leftrightarrow\; (l,s)\in A\big\},
$$

which indicate different patterns of match between $\mathbf{i}_1,\dots\mathbf{i}_m$ and $\mathbf{i}^*$. This facilitates the calculation of sums involving different patterns of match between $\mathbf{i}_\Omega$ and $\mathbf{i}^*$. Since $\sum_{A\in\Omega}\pi_A = 1$, we have, as in (3.15),

$$
\begin{aligned}
Eg(\xi_\Lambda)\prod_{s=1}^m\bigg(\sum_{\mathbf{i}_s\in\Lambda_s}\phi_{\mathbf{i}_s,s}\bigg) &= \sum_{A\subseteq\Omega}\sum_{\mathbf{i}_\Omega\in\tilde{\Lambda}_\Omega} Eg(\psi_{\mathbf{i}^*})\bigg\{\prod_{s=1}^m\phi_{\mathbf{i}_s,s}\bigg\}\pi_{A^c}(\mathbf{i}_\Omega;\mathbf{i}^*)\\
&= \sum_{A\subseteq\Omega}\sum_{\mathbf{i}_A\in\tilde{\Lambda}_A} Eg(\psi_{\mathbf{i}^*})\prod_{s=1}^m\phi_{\mathbf{i}'_s,s},
\end{aligned}
\tag{3.18}
$$

where $(\mathbf{i}_1,\dots\mathbf{i}_m) = \mathbf{i}_\Omega$ and $\mathbf{i}'_s = (i'_{1,s},\dots,i'_{k,s})$ with $i'_{l,s} = i_{l,s}$ for $(l,s)\in A$ and $i'_{l,s} = i_l^*$ for $(l,s)\notin A$, given $\mathbf{i}^*$ and $\mathbf{i}_A$. Given A and $\mathbf{i}_A$, let $\mathbf{i}^o = (i_1^o,\dots,i_k^o)$ be the index at which the maximum of $\psi_{\mathbf{i}}$ is reached over the set $\tilde{\Lambda}(A,\mathbf{i}_A) = \{(i_1,\dots,i_k)\in\Lambda\colon i_l\neq i_{l,k}\,\forall(l,k)\in A\}$, and $\mathbf{i}''_s = (i''_{1,s},\dots,i''_{k,s})$ with $i''_{l,s} = i_{l,s}$ for $(l,s)\in A$ and $i''_{l,s} = i_l^o$ for $(l,s)\notin A$. Note that $\tilde{\Lambda}(A,\mathbf{i}_A)$ is the space of combined indices which do not involve the specified $\mathbf{i}_A$. We have

$$
Eg(\psi_{\mathbf{i}^*})\prod_{s=1}^m\phi_{\mathbf{i}'_s,s}\le Eg(\psi_{\mathbf{i}^o})\prod_{s=1}^m\phi_{\mathbf{i}''_s,s}.
$$

Let $\tilde{\mathscr{F}}(A,\mathbf{i}_A)$ be the σ-field generated by $\{Y_i^{(l)}\colon (l,i)\neq(l,i_{l,s})\;\forall(l,s)\in A\}$. Since $\mathbf{i}^o$ is $\tilde{\mathscr{F}}(A,\mathbf{i}_A)$ measurable, for all $\mathbf{i} = (i_1,\dots,i_k)\in\tilde{\Lambda}(A,\mathbf{i}_A)$,

$$
E\bigg[g(\psi_{\mathbf{i}^o})\prod_{s=1}^m\phi_{\mathbf{i}''_s,s}\bigg|\tilde{\mathscr{F}}(A,\mathbf{i}_A)\bigg] = g(\psi_{\mathbf{i}^o})E\bigg[\prod_{s=1}^m\phi_{\mathbf{i}'''_s,s}\bigg|Y^{(l)}_{i'''_{l,s}},(l,s)\notin A\bigg]
$$

on the event $\{\mathbf{i}^o = \mathbf{i}\}$, where $\mathbf{i}'''_s = (i'''_{1,s}, \ldots, i'''_{k,s})$ with $i'''_{l,s} = i_{l,s}$ for $(l, s) \in A$ and $i'''_{l,s} = i_l$ for $(l, s) \notin A$. Since $\phi_{\mathbf{i},s}$ are sub-Bernoulli functions of $\tilde{Y}_{\mathbf{i}}$, by (2.3),

$$E\left[\prod_{s=1}^{m} \phi_{\mathbf{i}'''_s, s} \,\middle|\, Y^{(l)}_{i'''_{l,s}}, (l, s) \notin A\right] \le E \prod_{(l,s)\in A} \delta^{(l)}_{i_{l,s}}.$$

Thus, for the given A and $\mathbf{i}_A$, we have as in (3.16),

$$Eg(\psi_{\mathbf{i}^*}) \prod_{s=1}^{m} \phi_{\mathbf{i}'_s, s} \le Eg(\psi_{\mathbf{i}^o}) E \prod_{(l,s)\in A} \delta^{(l)}_{i_{l,s}} \le Eg(\psi_{\mathbf{i}^*}) E \prod_{(l,s)\in A} \delta^{(l)}_{i_{l,s}}.$$

Inserting this into (3.18), we find as in (3.17),

$$\begin{aligned} Eg(\xi_\Lambda) \prod_{s=1}^{m} \left(\sum_{\mathbf{i}_s \in \Lambda_s} \phi_{\mathbf{i}_s, s} \right) &\le \sum_{A \subseteq \Omega} \sum_{\mathbf{i}_A \in \Lambda_A} Eg(\psi_{\mathbf{i}^*}) E \prod_{(l,s)\in A} \delta^{(l)}_{i_{l,s}} \\ &\le Eg(\psi_{\mathbf{i}^*}) E \sum_{A\subseteq\Omega} \prod_{(l,s)\in A} T_{l,s} \\ &= Eg(\xi_\Lambda) E \prod_{l=1}^{k} \prod_{s=1}^{m} (1 + T_{l,s}). \end{aligned} \tag{3.19}$$

Hence (3.13) holds for general k and m. □

PROOF OF PROPOSITION 3.7. Let $\mathbf{i}^* \in \Lambda$ be the location of the maximum ξ_Λ. Similarly to the proof of Proposition 3.6, we find via Proposition 2.2 that

$$\begin{aligned} Eg(\xi_\Lambda) \prod_{s=1}^{m} \left(\sum_{\mathbf{i}\in\Lambda^{[k,s]}} \phi_{\mathbf{i},s} \right) &\le \sum_{A\subseteq\Omega} Eg(\xi_\Lambda) E \prod_{s=1}^{m} \left(\sum_{\mathbf{i}\in\Lambda^{[j_s,s]}} \delta_{\mathbf{i}} \right) \\ &= Eg(\xi_\Lambda) E \sum_{A\subseteq\Omega} \prod_{s=1}^{m} \binom{T_s}{j_s}, \end{aligned}$$

where $j_s = \#\{l\colon (l, s) \in A\}$. Since there are $\binom{k}{j}$ subsets of $\{1, \ldots, k\}$ of size $k - j$,

$$\sum_{A\subseteq\Omega} \prod_{s=1}^{m} \binom{T_s}{j_s} = \prod_{s=1}^{m} \sum_{j=0}^{k} \binom{k}{k-j} \binom{T_s}{k-j} = \prod_{s=1}^{m} \binom{k+T_s}{k}.$$

This completes the proof. □

3.4. *Proofs of Theorems* 3.2 *and* 3.4. We shall use Propositions 3.6 and 3.7 to prove (3.9) and (3.11). The Cantelli inequality below is applied to sums of sub-Bernoulli variables in Lemma 3.1 in the proof of (3.10) and (3.12).

LEMMA 3.8 (Cantelli inequality). *Let W be a random variable with $EW = \mu$ and $\mathrm{Var}(W) = \sigma^2$. Then $P\{W \ge \varepsilon\} \ge (\mu - \varepsilon)^2/\{\sigma^2 + (\mu - \varepsilon)^2\}$ for all $\varepsilon < \mu$.*

PROOF OF THEOREM 3.2. Let $\tilde{\phi}_{\mathbf{i}} = h(\tilde{X}_{\mathbf{i}})/\psi(\tilde{X}_{\mathbf{i}}; \theta_{(k)}, h, F_{\tilde{X}})$. By Lemma 3.1, $\tilde{\phi}_{\mathbf{i}}$ are sub-Bernoulli variables with parameter $\theta_{(k)}$. By (3.5) and (3.7), $\tilde{S}_{\tilde{\Lambda}_n}/\tilde{\xi}_{\tilde{\Lambda}_n, \theta_{(k)}} \le \sum_{\mathbf{i}\in\tilde{\Lambda}_n} \tilde{\phi}_{\mathbf{i}}$, so that (3.9) follows directly from Proposition 3.6.

Let us prove (3.10). Let $\mathbf{n} = (n_1, \dots, n_k)$ and $A \subseteq \Omega_k$ be fixed with $|A| = k - j$. Define

$$\tilde{\xi}_{\mathbf{n}, j, A} = \max\{c_{j, A}(X_{i_l}^{(l)}: l \notin A): i_l \in \Lambda_{n_l}^1 \ \forall l \in A^c\}, \tag{3.20}$$

where $c_{j, A}(y_{(j)}) = c_{j, A}(y_{(j)}; (1/n_1, \dots, 1/n_k), h, F_{\tilde{X}})$ as in (3.3). Let $(i_l^*: l \in A^c)$ be the index at which the maximum in (3.20) is reached. Define

$$\Phi_{\mathbf{n}, j, A} = \sum_{\mathbf{i}\in\Lambda_{\mathbf{n}, A}^*} \phi_{j, A}(\tilde{X}_{\mathbf{i}}) \tag{3.21}$$

with $\phi_{j, A}(y_{(k)}) = \phi_{j, A}(y_{(k)}; (1/n_1, \dots, 1/n_k), h, F_{\tilde{X}})$ as in (3.4), where

$$\Lambda_{n, A}^* = \{(i_1, \dots, i_k): i_l \in \Lambda_{n_l}^1 \ \forall l \in A, i_l = i_l^* \ \forall l \in A^c\}.$$

By Lemma 3.1, $E[\Phi_{\mathbf{n}, j, A}|\tilde{\xi}_{\mathbf{n}, j, A}] = 1$ for $\tilde{\xi}_{\mathbf{n}, j, A} > 0$, and by Lemma 3.1 and (2.5),

$$E[\Phi_{\mathbf{n}, j, A}^2|\tilde{\xi}_{\mathbf{n}, j, A}] \le \prod_{l\in A} EN_{n_l, 1/n_l}^2 \le (EN_1^2)^{k-j} = 2^{k-j}.$$

Thus, by Lemma 3.8 with $W = \Phi_{\mathbf{n}, j, A}$,

$$P\{\Phi_{\mathbf{n}, j, A} > \varepsilon|\tilde{\xi}_{\mathbf{n}, j, A}\} \ge \frac{(1-\varepsilon)^2}{(2^{k-j}-1)+(1-\varepsilon)^2}$$

on the set $\{\tilde{\xi}_{\mathbf{n}, j, A} > 0\}$. Since $\tilde{S}_{\mathbf{n}}/\tilde{\xi}_{\mathbf{n}, j, A} \ge \Phi_{\mathbf{n}, j, A}$ by (3.4), this implies

$$P\{\tilde{\xi}_{\mathbf{n}, j, A} > t\} \le \frac{(2^{k-j}-1)+(1-\varepsilon)^2}{(1-\varepsilon)^2} P\{\tilde{S}_{\mathbf{n}} > \varepsilon t\}.$$

Since $\tilde{\xi}_{\mathbf{n}}$ is the maximum of $\tilde{\xi}_{\mathbf{n}, j, A}$ over all $A \subseteq \Omega_k$ and there are $\binom{k}{j}$ of these with $|A| = k - j$,

$$P\{\tilde{\xi}_{\mathbf{n}} > t\} \le P\{\tilde{S}_{\mathbf{n}} > \varepsilon t\} \sum_{j=0}^{k} \binom{k}{j} \frac{(2^{k-j}-1)+(1-\varepsilon)^2}{(1-\varepsilon)^2}.$$

This completes the proof. □

PROOF OF THEOREM 3.4. Let $\phi_{\mathbf{i}} = h(X_{\mathbf{i}})/\psi(X_{\mathbf{i}}; (\theta, \dots, \theta), h, F_{X_{(k)}})$. By Lemma 3.1, $\phi_{\mathbf{i}}$ are sub-Bernoulli variables. By (3.5) and (3.6), $S_{\Lambda_n}/\xi_{\Lambda_n, \theta} \le \sum_{\mathbf{i}\in\Lambda_n} \phi_{\mathbf{i}}$, so that (3.11) follows directly from Proposition 3.7.

We shall only prove (3.12) for $n \ge k(k+1)$ with the explicit $C_{k,\varepsilon}$. The proof of (3.10) can be used to prove (3.12) if $\xi_n^{[k]}$ can be decoupled. Let us divide $\{1, \dots, n\}$ into $k+1$ blocks B_l as evenly as possible. Let $A \subseteq \Omega_k$ with $|A| = k - j$ and j blocks, say $B_1, \dots, B_j$, be fixed. Let $\xi_{j, A}^*$ be the maximum of $c_{j,A}(X_{\mathbf{i}})$ over $\mathbf{i} \in \Lambda_n^{[j]} \cap (\bigcup_{l=1}^{j} B_l)^{\otimes j}$, reached at $(i_1^*, \dots, i_j^*)$, and $\Phi_{j,A}^*$ be the sum as in (3.21), with fixed first j components of $\mathbf{i} = (i_1, \dots, i_k)$, $i_l = i_l^*$, $l \le j$, over $(i_{j+1}, \dots, i_k) \in \Lambda_n^{[k-j]} \cap (\bigcup_{l=j+1}^{k+1} B_l)^{\otimes(k-j)}$. Then, by Lemma 3.8 and (2.8)

$$P\{\Phi_{j,A}^* > \varepsilon \mid \xi_{j,A}^*\} \le \frac{(\mu_j^* - \varepsilon)^2}{v_j^* + (\mu_j^* - \varepsilon)^2},$$

where $\mu_j^* = (k/n)^{k-j}\binom{n_j}{k-j}$ and $v_j^* \le E\binom{N_{\lambda_j^*}}{k-j}^2 - (\mu_j^*)^2$, with n_j being the size of $\bigcup_{l=j+1}^{k+1} B_l$ and $\lambda_j^* = n_j(k/n)$. Consider the smallest possible μ_j^* with $n = s(k+1)+j$ and $n_j = s(k+1-j)$, $|B_l| = s+1$ for $l \le j$, for some s. In this case, $(k/n)\{n_j/(k-j)\} \ge 1$ by algebra for $s \ge k-j$, which holds as $n \ge k(k+1)$. Thus, $\mu_j^* \ge (k/n)^{k-j}\{n_j/(k-j)\}^{k-j} \ge 1$. In the case of largest possible λ_j^*, with $n_j = (s+1)(k+1-j)$ and $n = s(k+1)+(k+1-j)$ and then the smallest possible $s = k$, $n \ge k(k+1)$, we have $\lambda_j^* = n_j(k/n) \le k+1-j$. Therefore, $1+v_j^* \le E\binom{N_{k+1-j}}{k-j}^2 = 1+v_j$, say. As in the proof of Theorem 3.4, by Lemma 3.8,

$$P\{\xi_{j,A}^* > t\} \le \left\{1 + \frac{v_j}{(1-\varepsilon)^2}\right\} P\{S_n^{[k]} > \varepsilon t\}.$$

Now, $\xi_n^{[k]}$ is the maximum of $\xi_{j,A}^*$ over totally $\binom{k+1}{j}$ ways to select these j blocks and then over $j = 0, \dots, k$, so that

$$P\{\xi_n^{[k]} > t\} \le P\{S_n^{[k]} > \varepsilon t\} \sum_{j=0}^{k} \binom{k+1}{j} \left\{1 + \frac{v_j}{(1-\varepsilon)^2}\right\}.$$

The proof is complete. □

4. SLLN. Let $f(x_1, \dots, x_k)$ be a real Borel function. Let $f^2(x_1, \dots, x_k) = h(x_1, \dots, x_k)$. In this section we give nasc for the SLLN (1.3), its multisample version

$$\tilde{S}_n^{(k)}/b_n^k \to 0 \quad \text{a.s.} \tag{4.1}$$

and their symmetrized versions

$$T_n^{[k]}/b_n^{k/2} \to 0 \quad \text{a.s.}, \tag{4.2}$$

$$\tilde{T}_n^{(k)}/b_n^{k/2} \to 0 \quad \text{a.s.}, \tag{4.3}$$

where $\tilde{S}_n^{(k)}$ is given by (1.7), and $T_n^{[k]}$ and $\tilde{T}_n^{(k)}$ are given by (1.8). We shall assume throughout this section that

$$\sup_{n \ge 1} \frac{b_n^k}{n^k} \sum_{m=n}^{\infty} \frac{m^{k-1}}{b_m^k} < \infty. \tag{4.4}$$

We shall also assume that the function f is permutation invariant, $f(x_{i_1}, \dots, x_{i_k}) = f(x_1, \dots, x_k)$ for all $(i_1, \dots, i_k) \in \Lambda_k^{[k]}$.

THEOREM 4.1. *Let* $\tilde{S}_n^{(k)}$ *be given by* (1.7) *and* $\tilde{\xi}_n^{(k)}$ *by* (3.8). *Let* $\varepsilon_i > 0$, $1 \leq i \leq 3$. *Let* n_j *be a sequence of positive integers such that* $1 < \gamma_1 \leq n_{j+1}/n_j \leq \gamma_2 < \infty$, $j \geq 1$. *Then* (4.1) *and* (4.3) *are equivalent to each other and to each and all of the following statements*:

$$\tilde{\xi}_n^{(k)}/b_n^k \to 0 \quad a.s.; \tag{4.5}$$

$$\sum_{j=1}^{\infty} P\big\{\tilde{S}_{n_j}^{(k)} > \varepsilon_1 b_{n_j}^k\big\} < \infty; \tag{4.6}$$

$$\sum_{j=1}^{\infty} P\big\{\tilde{T}_{n_j}^{(k)} > \varepsilon_2 b_{n_j}^{k/2}\big\} < \infty; \tag{4.7}$$

$$\sum_{j=1}^{\infty} P\big\{\tilde{\xi}_{n_j}^{(k)} > \varepsilon_3 b_{n_j}^k\big\} < \infty. \tag{4.8}$$

REMARK. These are the multisample versions of the SLLN, since $X^{(1)}, \dots, X^{(k)}$ are allowed to have different distributions.

THEOREM 4.2. *Let* $\tilde{S}_n^{[k]}$ *be given by* (1.1) *and* $\xi_n^{[k]}$ *by* (3.6). *Let* $\varepsilon_i > 0$ *and* n_j *be as in Theorem* 4.1. *Suppose* $\{X, X^{(1)}, \dots, X^{(k)}\}$ *are iid random variables. Then*, (1.3), (4.2) *and* (4.6) (*and other statements in Theorem* 4.1) *are equivalent to each other and to each and all of the following statements*:

$$\xi_n^{[k]}/b_n^k \to 0 \quad a.s.; \tag{4.9}$$

$$\sum_{j=1}^{\infty} P\big\{S_{n_j}^{[k]} > \varepsilon_1 b_{n_j}^k\big\} < \infty; \tag{4.10}$$

$$\sum_{j=1}^{\infty} P\big\{T_{n_j}^{[k]} > \varepsilon_2 b_{n_j}^{k/2}\big\} < \infty; \tag{4.11}$$

$$\sum_{j=1}^{\infty} P\big\{\xi_{n_j}^{[k]} > \varepsilon_3 b_{n_j}^k\big\} < \infty. \tag{4.12}$$

We state Lemma 3.5 and Proposition 4.2 of Zhang (1996) here as it is applied in some crucial parts in the proofs.

LEMMA 4.3. *Let* η_j *be nonnegative random variables and* A_j *be events. Then*

$$\sum_{j=j_0}^{\infty} \eta_j I_{A_j} \leq I_{A_{j_0}} \sum_{i=j_0}^{\infty} \eta_i + \sum_{j=j_0}^{\infty} I_{A_j^c A_{j+1}} \sum_{i=j+1}^{\infty} \eta_i.$$

For disjoint sets of positive integers $A_1, \dots, A_l$, define the sum of "cross-block" terms

$$S^{[k]}_{A_1\otimes\cdots\otimes A_l} = \sum_{\mathbf{i}\in\Lambda^{[k]}_\infty} h(X_{\mathbf{i}}) I\{\mathbf{i} \in A_1 \otimes \cdots \otimes A_l\},$$

where $A_1 \otimes \cdots \otimes A_l$ is the set of vectors $(i_1, \dots, i_k)$ such that $\{i_1, \dots, i_k\} \subseteq \bigcup_{j=1}^l A_j$ and $\{i_1, \dots, i_k\} \cap A_j \neq \varnothing$ for all $1 \le j \le l$. For example, $S^{[k]}_A / \binom{|A|}{k}$ is the U-statistic based on the set of variables $\{X_i, i \in A\}$, where $|A|$ is the size of the set A.

PROPOSITION 4.4. *Let A_j, $0 \le j \le l$, be disjoint sets of positive integers and $a_{\mathbf{i}}$ be real numbers indexed by vectors $\mathbf{i} = (i_1, \dots, i_k)$. Then*

$$\sum_{\mathbf{i}\in\Lambda} a_{\mathbf{i}} I\{\mathbf{i} \in (A_1 \otimes \cdots \otimes A_l) \cup (A_0 \otimes A_1 \otimes \cdots \otimes A_l)\}$$
$$= \sum_{j=0}^{l} (-1)^{l-j} \sum_{0<m_1<\cdots<m_j\le l} \sum_{\mathbf{i}\in\Lambda} a_{\mathbf{i}} I\{\{i_1, \dots, i_k\} \subseteq A_0 \cup A_{m_1} \cup \cdots \cup A_{m_j}\}$$

for all sets Λ of finitely many vectors. In particular,

$$S^{[k]}_{A_1\otimes\cdots\otimes A_k} = \sum_{j=0}^{k} (-1)^{k-j} \sum_{0<m_1<\cdots<m_j\le k} S^{[k]}_{A_0\cup A_{m_1}\cup\cdots\cup A_{m_j}}.$$

Note that $\left(\prod_{l=1}^k |A_l|\right)^{-1} S^{[k]}_{A_1\otimes\cdots\otimes A_k}$ are multisample U-statistics.

PROOF OF THEOREM 4.1. We shall prove (4.7) $\Rightarrow$ (4.6) $\Rightarrow$ (4.8) $\Rightarrow$ (4.3) $\Rightarrow$ (4.7), (4.8) $\Rightarrow$ (4.1) $\Rightarrow$ (4.6) and (4.8) $\Rightarrow$ (4.5) $\Rightarrow$ (4.8). We use M to denote an arbitrary positive constant. We may choose any value of ε_i (large or small), since (4.1) has nothing to do with the scaling.

(i) (4.7) $\Rightarrow$ (4.6). By Lemma 3.8 with $W = (\tilde{T}^{(k)}_{n_j})^2 / \tilde{S}^{(k)}_{n_j}$ given $\tilde{S}^{(k)}_{n_j}$ and the Khintchine inequality,

$$P\left\{ \frac{(\tilde{T}^{(k)}_{n_j})^2}{\tilde{S}^{(k)}_{n_j}} \ge 1/2 \,\middle|\, \tilde{S}^{(k)}_{n_j} \right\} \ge \frac{1/4}{4 + 1/4}.$$

See the proof of (3.10) and Giné and Zinn [(1994), page 122] for details.

(ii) (4.6) $\Rightarrow$ (4.8). See (3.10) in Theorem 3.2.

(iii) (4.8) $\Rightarrow$ (4.3). This part is very close to the proof of Theorems 2.2 (sufficiency) and 3.1 in Zhang (1996). Let $\varepsilon_3 = 1$ and $\varepsilon > 0$. By the Doob inequality for the reverse martingale $\tilde{T}_n^{(k)}/n^k$, $n \ge n_j$, conditionally on $\tilde{\xi}_{n_j}^{(k)}$,

$$
\begin{aligned}
P\bigg\{&\max_{n_j \le n < n_{j+1}} \frac{(\tilde{T}_n^{(k)})^2}{b_n^k} > \varepsilon^2, \tilde{\xi}_{\Lambda^k_{n_{j+1}},(1/n_j,\dots,1/n_j)} \le b_{n_j}^k\bigg\} \\
&\le \max_{n_j \le n < n_{j+1}} \bigg(\frac{n^{2k}}{\varepsilon^2 b_n^k}\bigg) E \max_{n_j \le n < n_{j+1}} \bigg(\frac{\tilde{T}_n^{(k)}}{n^k}\bigg)^2 I\{\tilde{\xi}_{n_{j+1}}^{(k)} \le b_{n_j}^k\} \\
&\le 4\bigg(\frac{n_{j+1}^{2k}}{\varepsilon^2 b_{n_j}^k}\bigg) E\bigg(\frac{\tilde{T}_{n_j}^{(k)}}{n_j^k}\bigg)^2 I\{\tilde{\xi}_{n_j}^{(k)} \le b_{n_j}^k\} \\
&\le \frac{4\gamma_2^{2k}}{\varepsilon^2 b_{n_j}^k} E(\tilde{T}_{n_j}^{(k)})^2 I\{\tilde{\xi}_{n_j}^{(k)} \le b_{n_j}^k\} \\
&= \frac{4\gamma_2^{2k} n_j^k}{\varepsilon^2 b_{n_j}^k} Eh(\tilde{X}_1) I\{\tilde{\xi}_{n_j}^{(k)} \le b_{n_j}^k\},
\end{aligned}
$$

where $\tilde{X}_1 = (X_1^{(1)}, \dots, X_1^{(k)})$. Thus,

$$
\begin{aligned}
\sum_{j=1}^{\infty} &P\bigg\{\max_{n_j \le n < n_{j+1}} (\tilde{T}_n^{(k)})^2/b_n^k > \varepsilon^2\bigg\} \\
&\le \sum_j P\{\tilde{\xi}_{\Lambda_{n_{j+1}},(1/n_j,\dots,1/n_j)} > b_{n_j}^k\} + M\sum_j \frac{n_j^k}{b_{n_j}^k} Eh(\tilde{X}_1) I\{\tilde{\xi}_{n_j}^{(k)} \le b_{n_j}^k\}.
\end{aligned}
$$

By (4.8), the first sum on the right is finite. By (4.4) and Lemma 4.3,

$$
\begin{aligned}
\sum_j &\frac{n_j^k}{b_{n_j}^k} Eh(\tilde{X}_1) I\{\tilde{\xi}_{n_j}^{(k)} \le b_{n_j}^k\} \\
&\le M + M\sum_j \frac{n_{j+1}^k}{b_{n_{j+1}}^k} Eh(\tilde{X}_1) I\{\tilde{\xi}_{n_j}^{(k)} > b_{n_j}^k, \tilde{\xi}_{n_{j+1}}^{(k)} \le b_{n_{j+1}}^k\}.
\end{aligned}
$$

It follows from (4.8) and (3.9) of Theorem 3.2 (with $\tilde{\xi}_{\Lambda_n,\theta_{(k)}} = \tilde{\xi}_{n_j}^{(k)}$ and $g(t) = I\{t > b_{n_j}^k\}$) that the right-hand side above is bounded by

$$
\begin{aligned}
M + M\sum_j &E\frac{\tilde{S}_{n_j}^{(k)}}{b_{n_{j+1}}^k} I\{\tilde{\xi}_{n_j}^{(k)} > b_{n_j}^k, \tilde{\xi}_{n_{j+1}}^{(k)} \le b_{n_{j+1}}^k\} \\
&\le M + M\sum_j E\frac{\tilde{S}_{n_j}^{(k)}}{\tilde{\xi}_{n_j}^{(k)}} I\{\tilde{\xi}_{n_j}^{(k)} > b_{n_j}^k\} \\
&\le M + M\sum_j P\{\tilde{\xi}_{n_j}^{(k)} > b_{n_j}^k\} < \infty.
\end{aligned}
$$

(iv) (4.3) $\Rightarrow$ (4.7). This is a consequence of Proposition 4.4 and the Borel–Cantelli lemma. For details, see Step 3 of the proof of Theorem 4.1 of Zhang (1996), page 1608.

(v) (4.8) $\Rightarrow$ (4.1). The proof is simpler than (iii).
(vi) (4.1) $\Rightarrow$ (4.6). See (iv).
(vii) (4.8) $\Rightarrow$ (4.5) $\Rightarrow$ (4.8). The Borel–Cantelli lemma. □

PROOF OF THEOREM 4.2. By Proposition 4.4, (1.3) $\Rightarrow$ (4.6), as in (iv) of the proof of Theorem 4.1. Since $h(x_1, \dots, x_k) \ge 0$, (4.6) $\Rightarrow$ (4.10) by de la Peña and Montgomery-Smith (1995). The proofs of (4.10) $\Rightarrow$ (4.12) $\Rightarrow$ (4.2) $\Rightarrow$ (4.11) $\Rightarrow$ (4.10), (4.12) $\Rightarrow$ (1.3), and (4.12) $\Rightarrow$ (4.9) $\Rightarrow$ (4.12), are identical to those of (4.6) $\Rightarrow$ (4.8) $\Rightarrow$ (4.3) $\Rightarrow$ (4.7) $\Rightarrow$ (4.6), (4.8) $\Rightarrow$ (4.1) and (4.8) $\Rightarrow$ (4.5) $\Rightarrow$ (4.8), respectively. □

PROOF OF THEOREM 1.1. (i) (4.12) $\Rightarrow$ (1.16)–(1.18). Let $2^j \le m_j < 2^{j+1}$ be the index at which $P\{\xi_n^{[k]} > b_n^k\}$ reaches its maximum over $2^j \le n < 2^{j+1}$. Taking $n_j = m_{2j}$ in (4.12) with $\varepsilon_3 = 1$ and then $n_j = m_{2j+1}$, we find

$$\sum_{n=1}^{\infty} n^{-1} P\{\xi_n^{[k]} > b_n^k\} < \infty. \tag{4.13}$$

For $k = 2$ and $b_n^2 \ge c_0(2/n)$, $h(X_1, X_2) > b_n^2 \vee c_1(X_1; 2/n) \vee c_1(X_2; 2/n)$ and (1.9) imply $\phi_n(X_1, X_2) = 1$, with $\phi_n(x, y) = h(x, y)/\psi(x, y; 2/n)$, so that the left-hand side of (1.18) is bounded by

$$\sum_{n=2}^{\infty} nE\phi_n(X_1, X_2) I\{\xi_n^{[2]} > b_n^2\}$$
$$= \sum_{n=2}^{\infty} \frac{2}{n-1} E \sum_{\mathbf{i} \in \Lambda_n^{[2]}} \phi_n(X_{\mathbf{i}}) I\{\xi_n^{[2]} > b_n^2\} \le \sum_{n=2}^{\infty} \frac{2E\binom{2+N_2}{2}}{n-1} P\{\xi_n^{[2]} > b_n^2\} < \infty.$$

The inequality above is a consequence of Proposition 3.7, with $g(x) = I\{x > b_n^2\}$. Also, (1.16) follows from $\xi_n^{[2]}/b_n^2 \to 0$ a.s. and (1.17) follows from (4.13), as (4.13) implies $\sum_n n^{-1} P\{\max_{1 \le i \le n} c_1(X_i, 2/n) > b_n^2\} < \infty$.

(ii) (1.16)–(1.18) $\Rightarrow$ (4.12). By the Borel–Cantelli lemma (1.16)–(1.18) $\Rightarrow$ (4.13). We obtain (4.13) $\Rightarrow$ (4.12) by taking m_j to be the index of the minimum in the block $2^j \le n < 2^{j+1}$ in the proof of (4.12) $\Rightarrow$ (4.13) in (i).

5. Examples. For $0 < p < 2$, Giné and Zinn (1992) proved that $E|f(X_1, \dots, X_k)|^p < \infty$ is sufficient for (4.2) with $\sqrt{b_n} = n^{1/p}$. Here we give an example to show that the pth moment condition is in some sense far away from necessary. By the equivalence of (4.1) and (4.3) and the Kolmogorov and Marcinkiewicz–Zygmund strong laws, we have the following example.

EXAMPLE 5.1. Let $\{Y^{(l)}, Y_n^{(l)}, n \ge 1\}$ be independent sequences of iid random variables. Suppose $E|Y^{(l)}|^{p_l} < \infty$, $0 < p_l \le 2$, $1 \le l \le k$, $p_1 + \cdots + p_k < 2k$. Then $n^{-k/p} \prod_{l=1}^{k} (\sum_{i=1}^{n} \tilde{\varepsilon}_i^{(l)} Y_i^{(l)}) \to 0$ a.s., where $k/p = 1/p_1 + \cdots + 1/p_k$.

EXAMPLE 5.2. Take $0 < p < 2$ and set $p_1 = p/(k - p(k-1)/2) < p$. Define $f(x_1, \dots, x_k) = \sum_{\mathbf{i} \in \Lambda_k^{[k]}} x_{i_1}^{2/p_1} x_{i_2} \cdots x_{i_k}$, the permutation symmetrized version of the kernel $x_1^{2/p_1} x_2 \cdots x_k$. Let X be a nonnegative variable with $EX^2 < \infty$ but $EX^{2+\varepsilon} = \infty$ for all $\varepsilon > 0$. Then $n^{-k/p} \sum_{\mathbf{i}\in\Lambda_n^{[k]}} \varepsilon_{i_1}\varepsilon_{i_2}\cdots\varepsilon_{i_k} f(X_{i_1}, \dots, X_{i_k}) \to 0$ a.s., while $E|f(X_1, \dots, X_k)|^{p_1+\varepsilon} = \infty$ for all $\varepsilon > 0$.

PROOF. It is clear that $E|f(X_1, \dots, X_k)|^{p_1+\varepsilon} = \infty$ for all $\varepsilon > 0$. By the equivalence of (4.2) and (4.3), it suffices to show $n^{-k/p} \sum_{\mathbf{i}\in\Lambda_n^{[k]}} \varepsilon_{i_1}\varepsilon_{i_2}\cdots\varepsilon_{i_k} f(X_{i_1}^{(1)}, \dots, X_{i_k}^{(k)}) \to 0$ a.s., which is a consequence of Example 5.1 with $Y^{(1)} = \{X^{(1)}\}^{2/p_1}$, $Y^{(l)} = X^{(l)}$ for $2 \le l \le k$ and $p_2 = \cdots = p_k = 2$.

Acknowledgement. I thank the referee for many valuable suggestions.

REFERENCES

ARCONES, M. and GINÉ, E. (1993). Limit theorems for U-processes. *Ann. Probab.* **21** 1494–1542.

CHOW, Y. S. and TEICHER, H. (1988). *Probability Theory*. Springer, New York.

CUZICK, J., GINÉ, E. and ZINN, J. (1995). Laws of large numbers for quadratic forms, maxima of products and truncated sums of i.i.d. random variables. *Ann. Probab.* **23** 292–333.

DE LA PEÑA, V. and MONTGOMERY-SMITH, S. J. (1995). Decoupling inequalities for the tail probabilities of multivariate U-statistics. *Ann. Probab.* **23** 806–816.

GINÉ, E. and ZINN, J. (1992). Marcinkiewicz type laws of large numbers and convergence of moments for U-statistics. In *Probability in Banach Spaces 8* (R. M. Dudley, M. G. Hahn and J. Kuelbs, eds.) 273–291. Birkhäuser, Boston.

GINÉ, E. and ZINN, J. (1994). A remark on convergence in distribution of U-statistics. *Ann. Probab.* **22** 117–125.

GLESER. L. J. (1975). On the distribution of the number of successes in independent trials. *Ann. Probab.* **3** 182–188.

HOEFFDING, W. (1961). The strong law of large numbers for U-statistics. Institute of Statistics Mimeo Ser. 302, Univ. North Carolina, Chapel Hill.

KLASS, M. J. and NOWICKI (1997). Order of magnitude bounds for expectations of Δ_2-functions of non-negative random bilinear forms and generalized U-statistics. *Ann. Probab.* **25** 1471–1501.

KLASS, M. J. and ZHANG, C.-H. (1994). On the almost sure minimal growth rate of partial sum maxima. *Ann. Probab.* **22** 1857–1878.

LEDOUX, M. and TALAGRAND, M. (1991). *Probability in Banach spaces.* Springer, New York.

SERFLING, R. J. (1980). *Approximation Theorems of Mathematical Statistics.* Wiley, New York.

SEN, P. K. (1974). On L_p-convergence of U-statistics. *Ann. Inst. Statist. Math.* **26** 55–60.

TEICHER, H. (1992). Convergence of self-normalized generalized U-statistics. *J. Theoret. Probab.* **5** 391–405.

ZHANG, C.-H. (1996). Strong laws of large numbers for sums of products. *Ann. Probab.* **24** 1589–1615.

RUTGERS UNIVERSITY
DEPARTMENT OF STATISTICS
HILL CENTER FOR MATHEMATICAL SCIENCES
BUSCH CAMPUS
PISCATAWAY, NEW JERSEY 08854
E-MAIL: czhang@stat.rutgers.edu