Asymptotic Inference for High Dimensional Data

Jim Kuelbs
Department of Mathematics
University of Wisconsin
Madison, WI 53706-1388

Anand N. Vidyashankar⋆
Department of Statistical Science
Cornell University
Ithaca, NY 14853-4201

Abstract

In this paper we study inference for high dimensional data that are characterized by small sample sizes relative to the dimension of the data. In particular, we provide an infinite-dimensional framework to study statistical models that involve situations in which (i) the number of parameters increase with the sample size (that is allowed to be random) and (ii) there is a possibility of missing data. Under a variety of tail conditions on the components of the data, we provide precise conditions for the joint consistency of the estimators of the mean. In the process, we clarify and improve some of the recent consistency results that appeared in the literature. Furthermore, we provide various results concerning the rate of convergence in the joint consistency results. An important aspect of the work presented is the development of asymptotic normality results for these models. As a consequence, we construct different test statistics for the one-sample problem concerning the mean vector and obtain their asymptotic dist! ributions as a corollary of the infinite dimensional results. Finally, we use these theoretical results to develop an asymptotically justifiable methodology for data analyses. Simulation results presented here bring out the salient features of the proposed methodology and describe situations where the methodology can be successfully applied. The simulations also evaluate the robustness of the proposed methodology under a variety of conditions, some of which are substantially different from the technical conditions required of the theoretical results. Comparisons of the proposed methods to other methods used in the literature are also provided.

⋆ Research Supported in part by a grant from NSF DMS 000-03-07057 and also by grants from the NDCHealth Corporation

Key Words: Covariance matrix estimation, c_0, Functional genomics, High dimensional data, Infinite dimensional central limit theorem, Joint inference, Large p small n, Laws of large numbers, l_ρ, Microarrays, Shrinkage, Structured covariance matrices

AMS 1991 *Subject Classification:* 60B10, 60B12, 60F05, 62A01, 62H15, 62G20, 62F40, 92B15

Short title: Joint Inference for HD Data

1 Introduction

Modern scientific technology is providing a class of statistical problems that typically involve data that are high dimensional, and frequently lead to questions involving simultaneous inference for large sets of parameters. The number of parameters in these datasets is often random, and grows rapidly in comparison to the sample size; furthermore, there can be missing observations. Microarrays epitomize this situation, but similar problems arise in other areas such as polymerase chain reactions, proteomics, functional magnetic resonance imaging, and astronomy. For example, in microarray experiments the number of expressed genes differ between replicates, and certain genes do not express in all replications, which leads to missing data. Statistical analyses of such problems is an area of increasing concern, and various statistical models and methods have been developed to analyze these situations. Some recent references in this area include [11] and [29], which study the large p small n problem. The references [10], [22], [23], [24] and [25] study joint asymptotics in the context of general regression problems when the number of parameters diverge to infinity with the sample size . In particular, [11] investigates the simultaneous estimation of the marginal distributions in the large p small n problem, and it describes how these results can then be used to control the so-called False discovery rates(FDR).

The primary focus of this paper is to develop a general framework for the statistical aspects of these high dimensional problems, incorporating both a random number of parameters and missing data. This is achieved by relating it to various aspects of infinite dimensional problems. Although the methods of our paper apply generally to many high dimensional data problems, we will frequently use the terminology from microarrays to facilitate connections to one of the contemporary scientific disciplines. Now we turn to some specifics of our model.

For each fixed integer $n \geq 1$ we begin with a collection of independent sequences of real valued random variables $\{\xi_{n,i,j} : j \geq 1\}$. All are assumed to be defined on a common probability space, and there is no dependence relationship assumed as n and j vary. In the context of microarrays, for n fixed, each of these sequences represents the expression levels of genes in one replication of the experiment. The index n can be interpreted as either the time frame or as a label for the laboratory where the experiment is being performed. In particular, the random variable $\xi_{n,i,j}$ can then be thought of as the expression level of the j^{th} gene in the i^{th} replicate with index n. The number of replicates, for fixed index n, could be any integer $r(n)$, but for the sake of simplicity we take $r(n) = n$. Nevertheless, the techniques of this paper can be applied to develop results for other choices of $r(n)$.

Since the expressed genes between replicates may not coincide, either due to the random number that appear or for other reasons (which can be viewed as random deletions), we incorporate these two non-mutually exclusive possibilities into our model. We let $N_{n,i}$ denote the random number of variables within the i^{th} replicate having index n. We also assume for each integer $n \geq 1$ that $\{N_{n,i} : i \geq 1\}$ is an i.i.d sequence of integer valued random variables with $P(N_{n,i} \geq 1) = 1$. Of course, in real datasets we also have $P(N_{n,i} = \infty) = 0$, but our results also apply to the situation where $P(N_{n,i} = \infty) > 0$.

To model missing data, we postulate that the missing mechanism is independent of the expression level and the random number of parameters are involved. For this reason, we introduce the Bernoulli random variables $\{R_{n,i,j} : n \geq 1, i \geq 1, j \geq 1\}$ to represent missing data indicators, where

$$P(R_{n,i,j} = 1) = p \text{ for } n \geq 1, i \geq 1, j \geq 1. \tag{1.1}$$

We will assume that $0 < p \leq 1$, and also that the sequences $\{\xi_{n,i,j} : n \geq 1, i \geq 1, j \geq 1\}, \{R_{n,i,j} : n \geq 1, i \geq 1, j \geq 1\}$, and $\{N_{n,i} : n \geq 1, i \geq 1\}$ are independent. The case $p = 1$ corresponds to the case that there is no missing data.

In traditional multivariate analysis, such data is typically represented as random vectors in a fixed dimension d. However, since we are studying the model in which the dimension of the parameter vector diverges to infinity with the sample size, we represent it as a vector in R^∞, the linear space of all real sequences. That is, we set

$$\boldsymbol{X}_{n,i} = \sum_{j \geq 1} \xi_{n,i,j} \theta_{n,i,j} \mathbf{e}_j, \quad i = 1, \cdots, n, \tag{1.2}$$

where

$$\theta_{n,i,j} = I(j \leq N_{n,i}) R_{n,i,j}, \tag{1.3}$$

for $n \geq 1$, $i \geq 1$, $j \geq 1$, and $\{\mathbf{e}_j : j \geq 1\}$ is the canonical basis for R^∞; that is $\mathbf{e}_j = \{\delta_{j,k} : k \geq 1\}$ for $j = 1, 2, \cdots$, where $\delta_{j,k} = 1$ for $j = k$ and 0 if $j \neq k$. In the context of microarrays, the coordinates of the vector $\mathbf{X}_{n,i}$ are thought to be the "normalized expression levels" of genes identified in the i^{th} replicate with index n . In probabilistic terms, the collection $\boldsymbol{X}_{n,1}, \boldsymbol{X}_{n,2}, \cdots, \boldsymbol{X}_{n,n}$ forms a triangular array of n independent R^∞-valued random vectors. Let $N_n^\star = \max_{1 \leq i \leq n} N_{n,i}$ denote the maximum number of components (columns) in the dataset; or in the context of microarrays, the total number of expressed genes present. If $P(N_n^\star < \infty) = 1$, the components of $\boldsymbol{X}_{n,i}$, namely $\xi_{n,i,j}\theta_{n,i,j}$, equal 0 for $j > N_{n,i}$. In other words, $\boldsymbol{X}_{n,i} \in c_0$, where c_0 is the linear space of all real sequences converging to 0. Hence we will be concerned with asymptotic inference for data in c_0. We will also be interested in data in other Banach subspaces of R^∞, as these spaces and their norms arise naturally in various statistical settings. Throughout the paper we allow the possibility that $P(N_n^\star = \infty) > 0$. We also will use the notation $\boldsymbol{x} = \sum_{j \geq 1} x_j \mathbf{e}_j$ to denote a typical vector in R^∞, where $\{\mathbf{e}_j : j \geq 1\}$ denotes the canonical basis vectors defined above.

The space c_0, with the usual sup-norm given by

$$||\boldsymbol{x}||_\infty = \sup_{i \geq 1} |x_i|, \tag{1.4}$$

is naturally appropriate when studying the asymptotic inference for a one-sample problem using the maximum of suitable "averages" of gene expressions. The other Banach subspaces of R^∞ that we use in this paper are $l_\rho, 2 \leq \rho \leq \infty$, where the norm is given by

$$||\boldsymbol{x}||_\rho = (\sum_{j \geq 1} |x_j|^\rho)^{\frac{1}{\rho}} \tag{1.5}$$

when $2 \leq \rho < \infty$, and by (1.4) when $\rho = \infty$. Our main asymptotic results concern the statistics

$$\boldsymbol{S}_{n,n} = \sum_{i=1}^{n} \boldsymbol{X}_{n,i}, \tag{1.6}$$

and

$$\tilde{\boldsymbol{S}}_{n,n} = \sum_{i=1}^{n} \sum_{j \geq 1} \frac{\xi_{n,i,j} \theta_{n,i,j}}{V_{n,j}^{1/2}} \mathbf{e}_j \equiv \sum_{i=1}^{n} \sum_{j=1}^{N_{n,i}} \frac{\xi_{n,i,j} R_{n,i,j}}{V_{n,j}^{1/2}} \mathbf{e}_j, \tag{1.7}$$

where

$$V_{n,j} = \max\{1, \sum_{i=1}^{n} \theta_{n,i,j}\}, \; n \geq 1, j \geq 1. \tag{1.8}$$

Here the coordinate wise random-normalizers $V_{n,j}$ take into account the differences amongst columns due to missing data and random row lengths, and if we replace the $V_{n,j}$ in $\tilde{\boldsymbol{S}}_{n,n}$ by $n^{1/2}$, then we obtain $\boldsymbol{S}_{n,n}/n^{1/2}$. Our results include consistency, rates of convergence, and asymptotic normality for these sums.

The statistic $\tilde{\boldsymbol{S}}_{n,n}$ is important when we consider asymptotic normality in our model, as it essentially normalizes each column by the square root of the number of terms in that column. If $p = 1$, and for positive integer constants $b(n)$ we have $P(N_{n,i} = b(n)) = 1$ for all $i \geq 1$, then (1.7) reduces to

$$\tilde{\boldsymbol{S}}_{n,n} = \sum_{i=1}^{n} \sum_{j=1}^{b(n)} \frac{\xi_{n,i,j}}{n^{1/2}} \mathbf{e}_j. \tag{1.9}$$

If $p = 1$ and we also have $P(N_{n,1} = \infty) = 1$, then

$$\tilde{\boldsymbol{S}}_{n,n} = \sum_{i=1}^{n} \sum_{j \geq 1} \frac{\xi_{n,i,j}}{n^{1/2}} \mathbf{e}_j. \tag{1.10}$$

We also study the statistic

$$\tilde{\boldsymbol{T}}_{n,n} = \sum_{i=1}^{n} \sum_{j \geq 1} \frac{\xi_{n,i,j}\theta_{n,i,j}}{V_{n,j}} \mathbf{e}_j \equiv \sum_{i=1}^{n} \sum_{j=1}^{N_{n,i}} \frac{\xi_{n,i,j} R_{n,i,j}}{V_{n,j}} \mathbf{e}_j. \tag{1.11}$$

For the moment assume that $E(\xi_{n,i,j})(\equiv \mu_{n,j})$ is independent of i, and $\mathcal{H}_{n,j}$ is the sigma-field generated by the random variables $\{N_{n,i} : i \geq 1\}$, $\{R_{n,i,j} : i, j \geq 1\}$ and the event $\cup_{i=1}^{n} \{\theta_{n,i,j} = 1\}$. Also, for each $j \geq 1$, let $\tilde{\mathbf{T}}_{n,n,j}$ denote the j^{th} coordinate of $\tilde{\boldsymbol{T}}_{n,n}$. Then, these coordinate sums are conditionally unbiased estimators of $\mu_{n,j}$, i.e. for each $j \geq 1$ we have

$$E(\tilde{\mathbf{T}}_{n,n,j} | \mathcal{H}_{n,j}) = \mu_{n,j}.$$

Furthermore, for each $j \geq 1$, these components are also conditionally least squares estimators in the following one-way analysis of variance model

$$\xi_{n,i,j} = \mu_{n,j} + \epsilon_{n,i,j}, \quad 1 \leq i \leq V_{n,j}, \tag{1.12}$$

where $E(\epsilon_{n,i,j}) = 0$. Analysis of each of these components can be carried out using traditional techniques involving one dimensional laws of large numbers and the central limit theorem. However, the joint analysis of all the variables (or even finitely many components) brings in correlations between the components of $\xi_{n,i,j}$ for various j, and also the correlations induced by the random sample sizes $N_{n,i}$ and the random normalizations $V_{n,j}$. Our goal is to address the joint behavior of these statistics under a variety of conditions.

Since the statistics $\boldsymbol{S}_{n,n}$, $\tilde{\boldsymbol{S}}_{n,n}$ and $\tilde{\boldsymbol{T}}_{n,n}$ are infinite-dimensional vectors, the choice of the norm will play a critical role in the analysis. As mentioned previously, we will work with different norms on various subspaces of R^{∞}. Expectedly, the asymptotic results that we develop will depend on the assumptions on the random variables $N_{n,i}$ and on the tail behavior of $\xi_{n,i,j}$. When $N_{n,i} = p_n$, where p_n is non-random, exponential in n, and $p = 1$, this is the so-called large p small n problem, and [11] and [29] studied the behavior of $\boldsymbol{S}_{n,n}$ in the sup-norm under various assumptions on the tail behavior of $\xi_{n,i,j}$. For example, the results proved in [29] assume that the random variables have bounded support, while [11] replaces this condition by various exponential decay conditions on the tail behavior of $\xi_{n,i,j}$. The primary technique employed in [11] to obtain consistency results uses Massart's [16], which provides uniform constants for the exponential rate of sup-norm convergence of the empirical distribution function to the true distribution function. This is then used to obtain results for the relevant partial sums of random variables using integration-by-parts techniques. While this approach yields useful results, the integration by parts required seems to obscure the true nature of the matter. From what we do here, we will see that it is more fruitful to study the problem from the point of view of the random variables themselves in that we are able to clarify some of the results described in [11], and also extend them under a broader range of conditions to our more general model.

The rest of the paper is organized as follows: Section 2 develops further notation and contains statements of the main results of the paper. These results concern joint consistency and joint asymptotic normality. Our first result is a global consistency result providing complete convergence for $||\boldsymbol{S}_{n,n}||_\infty$ and $||\tilde{\boldsymbol{S}}_{n,n}||_\infty$, and in order to further understand the magnitude of $||\boldsymbol{S}_{n,n}||_\infty$ our Theorem 1 also establishes the exponential integrability of a suitably scaled version of $||\boldsymbol{S}_{n,n}||_\infty^2$. This result is proved under a sub-Gaussian tail condition on the variables $\{\xi_{n,i,j}\}$. An important question concerning the role of the index n in $\xi_{n,i,j}$ appears in Theorem 3. There the magnitude of $||\boldsymbol{S}_{n,n}||_\infty$ is shown to be even smaller if the variables summed are from an i.i.d. sequence of R^∞ valued independent random vectors, rather than a triangular array. Theorems 2 and 4 both provide additional con! sistency results for $\tilde{\boldsymbol{S}}_{n,n}$ under a variety of conditions, the main ones being on N_n^* and the tail behavior of the random variables $\{\xi_{n,i,j}\}$. In particular, Theorems 2 and 4 cover all the consistency results in [11], and Theorem 4 even provides results when the tail behavior of the $\{\xi_{n,i,j}\}$ is polynomial in nature. Theorem 5 provides both weak and strong consistency for $\tilde{\boldsymbol{T}}_{n,n}$. The proofs of these consistency results are contained in Section 4.

An important aspect of our work is that it investigates the asymptotic limit distribution of $\tilde{\boldsymbol{S}}_{n,n}$, and in Theorems 6 and 7 we provide central limit theorems in c_0 and in ℓ_ρ, $2 \leq \rho < \infty$, under a variety of asumptions. A novel aspect of these results is that the normalizer is different for each component in $\tilde{\boldsymbol{S}}_{n,n}$, which is motivated due to the presence of missing data. In this setting we also obtain the analogous weak laws of large numbers, i.e. we show that the random vectors $\tilde{\boldsymbol{T}}_{n,n}$ converges to zero in probability as n tends to infinity. An immediate consequence of the asymptotic normality results is that $||\tilde{\boldsymbol{S}}_{n,n}||_\infty$ converges in distribution to the sup-norm of the limiting Gaussian random vector. These results are useful in developing tests for sample size determination while designing experiments. Our asymptotic normality results in ℓ_2, and also ℓ_ρ, naturally yield limit distributions of statistics analogous to the Hotelling's T^2 statistic. One can use this convergence to develop tests similar to those based on Hotelling's T^2 statistic. A simulation study undertaken in Section 6 discusses such problems. The proofs of asymptotic normality results are contained in Section 5, and the concluding remarks are in Section 7.

2 Additional Notation and Main Results

In this section we develop the additional notation and assumptions, and present the main results of the paper. As mentioned in the introduction, since the dimension of the vectors grow as n increases, our approach is to view the data as random elements of the space of real sequences R^∞.

The results that we obtain will depend critically on the tail probabilities of the random variable $\{\xi_{n,i,j}\}$. These assumptions are of two types, namely that the tail probabilities decay at an exponential rate, or that they decay polynomially. In the large p small n problem, our results imply that these tail probability conditions are closely tied to the way p must relate to n. For example, in the classic version of this problem where p grows exponentially fast in n, we need tail probabilities that decay exponentially fast, whereas if p grows only as a power of n, then we only need polynomial decay for the tails. The precise nature of this interplay for consistency results is contained in Theorems 2 and 4. In particular, the remarks following these theorems contain precise information on their relationship to the large p small n problem.

First we discuss the exponential decay case. Here we assume that for some $r,\ 0 < r \leq 2$, and all $x \geq 0$ there are constants $c_{n,j}$ and $k_{n,j}$ such that

$$P(|\xi_{n,,i,j}| \geq x) \leq c_{n,j} e^{-k_{n,j} x^r}, \tag{2.1}$$

for all $n \geq 1, j \geq 1$. Random variables satisfying (2.1) with $r = 2$ are usually said to be sub-Gaussian, and if for $1 \leq i \leq n$ we have that each $\xi_{n,i,j}$ takes values in the interval $[a_{n,j}, b_{n,j}]$, then we will see below that (2.1) holds with r=2,

$$c_{n,j} = 2 \text{ and } k_{n,j} = (2(b_{n,j} - a_{n,j})^2)^{-1}, \quad n \geq 1,\ j \geq 1. \tag{2.2}$$

Actually, (2.1) is well known, but it will also emerge from what we prove below. Throughout, when $\xi_{n,i,j} = 0$ with probability one, in (2.1) we take

$$c_{n,j} = 1 \text{ and } k_{n,j} = \infty, \quad n \geq 1,\ j \geq 1. \tag{2.3}$$

In addition, note that $c_{n,j} \geq 1$ is necessary by setting $x = 0$ in (2.1).

It is also useful to notice that if the condition (2.1) holds for some $r^* > 1$, then it holds for all $1 \leq r \leq r^*$ by simply adjusting the constants $c_{n,j}$ and keeping the same $k_{n,j}$. In particular, if (2.1) holds for some $r > 2$, then it holds for $r = 2$, and we are in the sub-Gaussian setting. In [11] this seems to have gone unnoticed, and there one finds results for $r > 2$ which are weaker than the corresponding $r = 2$ results. However, this should not be the case as the previous comment implies the $r = 2$ result applies directly to what is proved there. Of course, in some settings there could be results that distinguish between various r values, even for $r > 2$, but that does not happen here, and is why we restrict r to be in $(0, 2]$. Our methods also yield results when $0 < r < 1$, whereas in [11] r is always greater than equal to one and, as described above, the $r > 2$ consistency results are equivalent to those for ! $r = 2$.

Another situation we will discuss is when the assumption of exponential decay of the tails of $\xi_{n,i,j}$ in (2.1) is replaced by the polynomial decay

$$P(|\xi_{n,i,j}| \geq x) \leq \frac{c_{n,j}}{(1+x)^{k_{n,j}}},\ x \geq 0, \tag{2.4}$$

where $c_{n,j} \geq 1$ and typically for our results, $2 < k_{n,j} < \infty$.

We will assume throughout the paper that $E(\xi_{n,i,j}) = 0$ for all $n, i, j \geq 1$. Should this not be the case, one would simply replace the tail probability conditions in (2.1) and (2.4) by analogous conditions for the variables $\{\xi_{n,i,j} - E(\xi_{n,i,j})\}$, and formulate the results in terms of these variables. One can provide fairly precise formulas for how these tail probability constants for a random variable ξ compare to those for $\xi - E(\xi)$, but the fact remains it is more convenient from a notational point of view to assume at the start that our random variables are mean zero. Hence we do this without further mention.

2.1 Consistency and Rates of Convergence Results

In this subsection we present several consistency and rate of convergence results for $\boldsymbol{S}_{n,n}$, $\tilde{\boldsymbol{S}}_{n,n}$, and $\tilde{\boldsymbol{T}}_{n,n}$ with respect to the sup-norm on c_0. Our first result concerns the statistic $\tilde{\boldsymbol{S}}_{n,n}$.

Theorem 1. *Let $\{\boldsymbol{X}_{n,i} : 1 \leq i \leq n\}$ be as in (1.2), assume (2.1) holds with $r = 2$, and take $\{a_n : n \geq 1\}$ to be a sequence of positive numbers. Furthermore, assume $c_{n,j}, k_{n,j}$ are constants such that $c_{n,j} \geq 1,\ k_{n,j} \leq \infty$ and*

$$\sum_{n\geq 1}\sum_{j\geq 1} \exp\{-(\epsilon a_n)^2 k_{n,j}/(16 c_{n,j})\} < \infty \tag{2.5}$$

for $\epsilon > \epsilon_0$. Then,

$$\sum_{n\geq 1} P(||\tilde{\boldsymbol{S}}_{n,n}||_\infty \geq \epsilon a_n) < \infty \tag{2.6}$$

for all $\epsilon > \epsilon_0$, where $\tilde{\boldsymbol{S}}_{n,n}$ is given as in (1.7). Thus, if the constants $c_{n,j}$ and $k_{n,j}$ are such that uniformly in $n \geq 1$ and for some $\delta > 0$,

$$k_{n,j}/(16 c_{n,j}) \geq \delta L(j+3), \tag{2.7}$$

then for all $\epsilon > 0$ such that $\epsilon^2\delta > 1$ we have

$$\limsup_{n\to\infty} \frac{||\tilde{\boldsymbol{S}}_{n,n}||_\infty}{(L(n+3))^{1/2}} \leq \epsilon. \tag{2.8}$$

In particular, if (2.7) holds, then with probability one

$$M = \sup_{n\geq 1} \frac{||\tilde{\boldsymbol{S}}_{n,n}||_\infty}{(L(n+3))^{1/2}} < \infty, \tag{2.9}$$

and there exists an $\alpha > 0$ such that

$$E(e^{\alpha M^2}) < \infty. \tag{2.10}$$

Moreover, if the $V_{n,j}^{1/2}$ are replaced by $n^{1/2}$ in $\tilde{\boldsymbol{S}}_{n,n}$, then again (2.6), (2.8), and (2.10) continue to hold.

Remark 1. *The above theorem provides an interesting interplay between the constants a_n, $c_{n,j}$, and $k_{n,j}$. For example, in order for $||\tilde{\boldsymbol{S}}_{n,n}||$ to be bounded in probability, the proof shows it is sufficient that the constants $c_{n,j} \geq 1$ and $k_{n,j} \leq \infty$ satisfy*

$$\sup_{n\geq 1}\sum_{j\geq 1} \exp\{-\epsilon k_{n,j}/(16 c_{n,j})\} < \infty \tag{2.11}$$

for some $\epsilon > 0$. In Theorem 6, we will establish a central limit theorem for $\tilde{\boldsymbol{S}}_{n,n}$, and that $\tilde{\boldsymbol{T}}_{n,n}$ converges to zero in probability under related conditions.

In Theorem 1, the impact of the random row sizes $\{N_{n,i} : i \geq 1\}$ is hidden due to our choice of normalizations $\{a_n\}$ as given in (2.5). For example, (2.5) implies the ratio $k_{n,j}/c_{n,j}$ cannot be bounded as j goes to infinity, but in our next result we only require this ratio to be uniformly bounded below in both n and j by a strictly positive constant. Under this different set of conditions the role of $\{N_{n,i} : i \geq 1\}$ appears in the normalizations for $\tilde{\boldsymbol{S}}_{n,n}$ given by $h(n)$ in (2.12). In particular, if $N_{n,i} = p_n$ for $\{i \geq 1,\ n \geq 1\}$, then Theorem 2 and Remark 4 below yield the results in Corollaries 1 and 2 in [11] when $r = 2$. The consistency results in [11] for $1 \leq r < 2$, as well as for many other cases, follow immediately from Theorem 4 below.

Theorem 2. *Let $\{\boldsymbol{X}_{n,i} : 1 \le i \le n\}$ be as in (1.2), and assume (2.1) holds with $r = 2$, and that for $1 \le c < \infty,\ 0 < k < \infty$, we have $c_{n.j} \le c$ and $k_{n,j} \ge k$ for all $n, j \ge 1$. Let,*

$$h(n) = (\theta_1^{-1} L(E(N_n^*)) + \theta_2 Ln)^{1/2}, \tag{2.12}$$

where $\theta_1 = k/(16c)$ and $\theta_2 > 0$. Then

$$\lim_{n\to\infty} P(\frac{||\tilde{\mathbf{S}}_{n,n}||_\infty}{h(n)} \ge 1) = 0, \tag{2.13}$$

and if also $\theta_1\theta_2 > 1$, then

$$\sum_{n\ge1} P(||\tilde{\boldsymbol{S}}_{n,n}||_\infty \ge h(n)) < \infty. \tag{2.14}$$

Finally, if the $V_{n,j}^{1/2}$ are replaced by $n^{1/2}$ in $\tilde{\boldsymbol{S}}_{n,n}$, then again (2.13)-(2.14) hold.

Remark 2. *Note that (2.13) immediately implies*

$$\frac{||\tilde{\boldsymbol{S}}_{n,n}||_\infty}{n^{1/2}} = O_P((\frac{L(E(N_n^*)) + Ln}{n})^{1/2}), \tag{2.15}$$

which relates to the weak consistency of $\tilde{\boldsymbol{S}}_{n,n}$. In particular, if $L(E(N_n^*))/n$ converges to zero, then (2.15) implies that $||\tilde{\boldsymbol{S}}_{n,n}||_\infty/n^{1/2}$ tends to zero in probability. Furthermore, replacing $n^{\frac{1}{2}}$ by n^δ for some $\delta > 0$ yields analogous results. Similar comments can be made concerning strong consistency. For example, (2.14) implies with probability one that

$$\limsup_{n\to\infty} \frac{||\tilde{\boldsymbol{S}}_{n,n}||_\infty}{h(n)} \le 1, \tag{2.16}$$

and hence $\tilde{\boldsymbol{S}}_{n,n}/n^{1/2}$ converges to zero with probability one provided $L(E(N_n^*))/n$ converges to zero as n approaches infinity.

Remark 3. *If $N_{n,i} = p_n$ for $\{i \ge 1,\ n \ge 1\}$, then (2.13)-(2.16) immediately hold under the conditions stated, and they also hold under this assumption if the $V_{n,j}^{1/2}$ are replaced by $n^{1/2}$ in $\tilde{\boldsymbol{S}}_{n,n}$. Furthermore, if $p_n \ge n$, then (2.15) immediately relates to the results of Corollary 2 in [11], as it implies*

$$\frac{||\tilde{\boldsymbol{S}}_{n,n}||_\infty}{n^{1/2}} = O_P((\frac{Lp_n}{n})^{1/2}), \tag{2.17}$$

In particular, (2.15) improves Corollary 2 and its proof considerably whenever $r \ge 2$ there, and the case $0 < r < 2$ will be discussed in what follows. To relate these results to Corollary 1 of [11] follows in a standard way. That is, if $\psi_p(x) = \exp\{x^p\} - 1$ for $x \ge 0$ and $1 \le p < \infty$, then the ψ_p-Orlicz norm of a random variable X is defined to be

$$||X||_{\psi_p} = \inf\{c > 0 : E(\psi_p(|X/c|) \le 1\}, \tag{2.18}$$

where the inf *over an empty set is taken to be infinity. Then, by Lemma 1 below, and Lemma 2.2.1 of [30], we have for each $n \ge 1$ that*

$$||\sum_{i=1}^n \xi_{n,i,j}/n||_{\psi_2} \le 4(3c/(nk))^{1/2}. \tag{2.19}$$

Thus by applying Lemma 2.2.2 of [30], and that

$$\frac{||\boldsymbol{S}_{n,n}||_\infty}{n} = \max_{1\le j\le p_n} |\sum_{i=1}^{n} \xi_{n,i,j}/n|, \tag{2.20}$$

it follows that there is a constant $\beta < \infty$, depending only on ψ_2, such that

$$||\frac{||\boldsymbol{S}_{n,n}||_\infty}{n}||_{\psi_2} \le 4(\frac{3c}{nk})^{1/2}\beta(\log(1+p_n))^{1/2}. \tag{2.21}$$

In view of our Lemma 1, we thus we have an improvement of Corollary 1 in [11], which only provides a comparable result when the $\{\xi_{n,i,j}\}$ are bounded.

To evaluate the sharpness of the exponential moments of M^2 in Theorem 1, we present the following result. In view of the law of the iterated logarithm this is essentially best possible, and occurs because for partial sums of a sequence of independent random vectors one can typically work along geometric subsequences. Of course, for partial sums of triangular arrays this option is generally unavailable.

Theorem 3. *Let $\mathbf{X}_1, \mathbf{X}_2, \cdots$ be independent random vectors with values in R^∞ and such that*

$$\mathbf{X}_i = \sum_{j\ge 1} \xi_{i,j}\mathbf{e}_j,\ i \ge 1, \tag{2.22}$$

where $\{\xi_{i,j} :\ j \ge 1\},\ i \ge 1$, are independent sequences of mean zero random variables satisfying

$$P(|\xi_{i,j}| \ge x) \le c_j e^{-k_j x^2} \tag{2.23}$$

for all $x \ge 0$ and $i \ge 1$. If $k_j/(16c_j) \ge \delta L(j+3)$ for some $\delta > 0$, and all $j \ge 1$, and $\mathbf{S}_n = \mathbf{X_1} + \cdots + \mathbf{X_n}$, then

$$\frac{||\mathbf{S}_n||_\infty}{n} \le M(\frac{LLn}{n})^{1/2}, \tag{2.24}$$

where

$$M = \sup_{n\ge 1} \frac{||\mathbf{S}_n||_\infty}{(nLLn)^{1/2}} \tag{2.25}$$

is such that

$$E(e^{\alpha M^2}) < \infty \tag{2.26}$$

for all $\alpha > 0$ sufficiently small.

We next study the situation when the random variables $\{\xi_{n,i,j}\}$ satisfy the exponential tail condition (2.1) with $0 < r < 2$, or polynomial decay as in (2.4). When $1 \le r < 2$, a special case of these results clarifies Corollary 2 of [11]. This can be seen in Remark 5 below. The $r = 2$ case in this corollary already appeared as a simple consequence of Theorem 1. It should also be observed that this theorem provides sufficient conditions for consistency which involve a precise relationship between the size of p_n in the large p small n problem, and the tail decay of the data. This relationship is shown to exist even when there is only polynomial decay in the data, and as one might expect in this situation the growth of p_n, or $E(N_n^*)$, needs to be further restricted, i.e. in such results p_n and $E(N_n^*)$ grow at a corresponding polynomial rate.

Theorem 4. *Let $\{\boldsymbol{X}_{n,i} : 1 \leq i \leq n\}$ be as in (1.2) and assume that (2.1) holds with $0 < r < 2$. Also assume for all $n \geq 1$ and $j \geq 1$, that $c_{n,j} \leq c$ and $k_{n,j} \geq k$, where $1 \leq c < \infty,\ 0 < k < \infty$. Let $s_n = c_1(L(E(N_n^*)) + 2Ln)^{1/r}$, and*

$$h(n) = (c_2^{-1}L(E(N_n^*)) + c_3 Ln)^{1/2}, \tag{2.27}$$

where $c_1 > 2/k^{1/r}$, $c_2 = k/(128c)$, and $c_3 > 0$. Then,

$$\lim_{n\to\infty} P(\frac{||\boldsymbol{S}_{n,n}||_\infty}{n^{1/2}s_n h(n)} \geq 1) = 0. \tag{2.28}$$

If we also assume $c_2c_3 > 1$, then

$$\sum_{n\geq 1} P(||\boldsymbol{S}_{n,n}||_\infty \geq n^{1/2}s_n h(n)) < \infty. \tag{2.29}$$

Furthermore, if $k > 2$ and the polynomial condition in (2.4) holds, then

$$||\boldsymbol{S}_{n,n}||_\infty = O_P(n^{1/2}s_n(L(E(N_n^*)))^{1/2}), \tag{2.30}$$

where $s_n = (nE(N_n^*))^{\frac{1}{k}+\beta}$ and $\beta > 0$. Additionally, if $E(N_n^*) \geq n$, $b > 8$, and $k\beta > 1/2$, then

$$\sum_{n\geq 1} P(||\boldsymbol{S}_{n,n}||_\infty \geq b s_n n^{1/2}(L(E(N_n^*))^{1/2}) < \infty. \tag{2.31}$$

In particular, if $E(N_n^*)$ is asymptotic to n^γ for $\gamma \geq 1$, then

$$\sum_{n\geq 1} P(||\boldsymbol{S}_{n,n}||_\infty \geq b s_n n^{1/2}(L(E(N_n^*))^{1/2}) < \infty, \tag{2.32}$$

provided $b > 8$ and $(\gamma + 1)k\beta > 1$.

Remark 4. *An immediate consequence of (2.28) is that*

$$\frac{||\boldsymbol{S}_{n,n}||_\infty}{n} = O_P(\frac{(L(E(N_n^*)) + Ln)^{\frac{2+r}{2r}}}{n^{1/2}}), \tag{2.33}$$

and if $(L(E(N_n^\star))^{\frac{(2+r)}{2r}}/n^{1/2} \to 0$, then (2.33) easily implies $\mathbf{S}_{n,n}/n$ converges to zero in probability. In addition, if $N_{n,i} = p_n$ for $n \geq 1, i \geq 1$, where $\{p_n : n \geq 1\}$ is a sequence of integers, and $p_n \geq n$, then it follows from (2.33) that

$$\frac{||\boldsymbol{S}_{n,n}||_\infty}{n} = O_P(\frac{(Lp_n)^{\frac{2+r}{2r}}}{n^{1/2}}). \tag{2.34}$$

Hence using the above for $r \in (0, 2)$, and (2.15) for the case $r = 2$, one obtains an extension and clarification of Corollary 2 and its proof in [11]. It is also interesting to observe that the method of proof for Theorem 4 applied to the $r = 2$ situation only yields

$$\frac{||\boldsymbol{S}_{n,n}||_\infty}{n} = O_P(\frac{Lp_n}{n^{1/2}}). \tag{2.35}$$

Hence we see the methods used for the $r = 2$ case in Theorem 2 are sharper than those we have for other values of r.

Remark 5. *If $c_2 = k/(128c)$ and $c_2c_3 > 1$, then an application of the Borel-Cantelli lemma and (2.29) implies that, with probability one,*

$$\limsup_{n\to\infty} \frac{||\boldsymbol{S}_{n,n}||_\infty}{(L(E(N_n^*)) + 2Ln)^{1/r} n^{1/2} h(n)} \leq 2/k^{1/r}. \tag{2.36}$$

Thus, (2.36) implies that $||\boldsymbol{S}_{n,n}||_\infty/n$ converges to 0 with probability one if $(L(E(N_n^\star))^{\frac{(2+r)}{2r}}/n^{1/2} \to 0$. Furthermore, if $N_{n,i} = p_n$ for $n \geq 1, i \geq 1$, where $\{p_n : n \geq 1\}$ is a sequence of integers and $p_n \geq n$, then, with probability one,

$$\limsup_{n\to\infty} \frac{||\mathbf{S}_{n,n}||_\infty}{n^{1/2}(\log(n^2 p_n))^{1/r}(c_2^{-1}\log p_n + c_3 \log n)^{1/2}} \leq 2/k^{1/r} \tag{2.37}$$

provided $c_2 = k/(128c)$, and $c_2c_3 > 1$. In addition, we easily see from (2.37) that $||\boldsymbol{S}_{n,n}||_\infty/n$ converges to zero with probability one provided $(Lp_n)^{\frac{(2+r)}{2r}}/n^{1/2} \to 0$.

Remark 6. *Under the assumption of polynomial decay given in Theorem 4, and assuming that $E(N_n^*)$ is asymptotic to n^γ for $\gamma \geq 1$, we easily see from (2.32) that $||\mathbf{S}_{n,n}||_\infty/n$ converges to zero almost surely provided k is sufficiently large so that for $\beta > 0$ we have $(\gamma+1)\beta k > 1$ and $(\gamma+1)(1/k+\beta) < 1/2$.*

Now we focus on the analogue of Theorem 2 for the case $0 < r < 2$. This is more delicate, and hence we consider random row lengths, but we do not allow random deletions. More precisely, we study

$$\tilde{\boldsymbol{T}}_{n,n} = \sum_{j\geq 1}\sum_{i=1}^{n} \xi_{n,i,j} I(j \leq N_{n,i})\mathbf{e}_j/V_{n,j}, \tag{2.38}$$

where $V_{n,j} = \sum_{i=1}^n I(j \leq N_{n,i})$.

Theorem 5. *Let $\tilde{\boldsymbol{T}}_{n,n}$ be defined as in (2.38) and assume $\{\xi_{n,i,j} : n \geq 1, i \geq 1, j \geq 1\}$ satisfy (2.1) with $0 < r < 2$, and that for $1 \leq c < \infty$, $0 < k < \infty$ we have $c_{n.j} \leq c$ and $k_{n,j} \geq k$ for all $n, j \geq 1$. Assume further that the support of $\{N_{n,i} : i \geq 1\}$ is a subset of $\{1, 2, \cdots, d(n)\}$ with $P(N_{n,1} = d(n)) = q_n$ where $q_n \geq n^{-\delta}$ for some $\delta \in [0, 1/4)$. Let $s_n = c_1(L(E(N_n^*)) + Ln)^{1/r}$, $c_1 > 4/k^{1/r}$, and*

$$h(n) = s_n\left(\frac{c_2^{-1}L(E(N_n^*)) + c_3Ln}{n^{1-4\delta}}\right)^{1/2}, \tag{2.39}$$

where $c_2 = k/(128c)$ and $c_3 > 0$. Then

$$\lim_{n\to\infty} P\left(\frac{||\tilde{\boldsymbol{T}}_{n,n}||_\infty}{h(n)} \geq 1\right) = 0. \tag{2.40}$$

Moreover,

$$\sum_{n\geq 1} P(||\tilde{\boldsymbol{T}}_{n,n}||_\infty \geq h(n)) < \infty \tag{2.41}$$

provided $c_2c_3 > 1$.

Remark 7. *It follows from (2.40) that*

$$||\tilde{\boldsymbol{T}}_{n,n}||_\infty = O_P\left(\frac{(L(E(N_n^*)) + Ln)^{\frac{2+r}{2r}}}{n^{\frac{1-4\delta}{2}}}\right), \tag{2.42}$$

and (2.41) implies that, with probability one,

$$\limsup_{n\to\infty} \frac{||\tilde{\boldsymbol{T}}_{n,n}||_\infty}{h(n)} \le 1. \tag{2.43}$$

Hence, if $\lim_{n\to\infty}(L(E(N_n^*)))^{\frac{2+r}{2r}}/n^{\frac{1-4\delta}{2}} = 0$*, then we have* $||\tilde{\boldsymbol{T}}_{n,n}||_\infty$ *tending to zero in probability and with probability one.*

2.2 Asymptotic Normality Results

In this section we present results on the asymptotic normality of the quantity $\tilde{\boldsymbol{S}}_{n,n}$. Since the estimator typically lives in c_0, our first result, Theorem 6, is a central limit theorem in that setting. Theorem 7 and Remark 11 contain CLT's in $\ell_\rho,\ 2 \le \rho < \infty$. These results hold when the underlying process is a triangular array with random row lengths and possibly missing data. We also are able to use the coordinate-wise random normalizations $V_{n,j}$ in Theorems 6 and 7, whereas Remark 11 uses the classical normalizations and, as result, holds under far weaker moment conditions. The papers [20] and [21] contain CLT's in c_0, as well as related references, and much is known about the CLT in the spaces $\ell_\rho,\ 2 \le \rho < \infty$. However, none of these results incorporate random row lengths, missing data, or coordinate-wise random normalizations in their formulations. In addition, the results in [20] and [21] require a uniform boundedness assumption on the $\{\xi_{n,i,j}\}$ to obtain results related to what we prove.

A key assumption in any central limit theorem is that there is a limiting covariance function. Since our results include the use of random column-wise normalizers, we have need of a couple different limiting covariances. That is, if

$$\Gamma_n(j_1, j_2) = \sum_{i=1}^{n} E(\xi_{n,i,j_1}\xi_{n,i,j_2})/n$$

is such that

$$\lim_{n\to\infty} \Gamma_n(j_1, j_2) = \Gamma(j_1, j_2) \tag{2.44}$$

for all $j_1, j_2 \ge 1$, then for $k = 1, 2$ we set

$$\Gamma(k, j_1, j_2) = p^k \Gamma(j_1, j_2), \text{ for } j_1 \ne j_2, \tag{2.45}$$

and

$$\Gamma(k, j_1, j_2) = p^{k-1} \Gamma(j_1, j_2), \text{ for } j_1 = j_2. \tag{2.46}$$

Theorem 6. *Let* $\{\boldsymbol{X}_{n,i} : 1 \le i \le n\}$ *be as in (1.2), assume (2.1) holds with* $r = 2$*, and that* $c_{n,j}, k_{n,j}$ *are constants such that* $c_{n,j} \ge 1,\ k_{n,j} < \infty$ *and*

$$\sup_{n,j\ge 1} c_{n,j}/k_{n,j} < \infty. \tag{2.47}$$

Also assume for all $\delta > 0$ *that*

$$\lim_{d\to\infty} \sup_{n\ge 1} \sum_{j\ge d} \exp\{-\delta k_{n,j}/c_{n,j}\} = 0. \tag{2.48}$$

If $\tilde{\mathbf{S}}_{n,n}$ *is given as in (1.7), then*

$$\{\mathcal{L}(\tilde{\mathbf{S}}_{n,n}) :\ n \ge 1\} \text{ is tight in } \mathrm{c}_0. \tag{2.49}$$

In addition, if $\tilde{\boldsymbol{T}}_{n,n}$ is as in (1.11), and for each $j \geq 1$ we have $\lim_{n\to\infty} P(N_{n,1} < j) = 0$, then $\tilde{\boldsymbol{T}}_{n,n}$ converges in probability to zero in c_0. Moreover, if the $V_{n,j}^{1/2}$ are replaced by $n^{1/2}$ in $\tilde{\mathbf{S}}_{n,n}$, then again (2.49) holds, and $\tilde{\boldsymbol{T}}_{n,n}$ converges in probability to zero. Furthermore, if we also assume (2.44), (2.45), and (2.46) hold, and for each $d < \infty$ we have $P(\min_{1\leq i\leq n} N_{n,i} < d) = o(1/n^2)$ as n tends to infinity, then $\Gamma(k,\cdot,\cdot)$ is the covariance of a centered Gaussian measure γ_k on c_0 for $k = 1, 2$, and

$$\mathcal{L}(\tilde{\mathbf{S}}_{n,n}) \ \text{ converges weakly to } \gamma_1 \tag{2.50}$$

on c_0. If the $V_{n,j}^{1/2}$ are replaced by $n^{1/2}$, then (2.50) still holds with limiting measure γ_2.

Remark 8. *The conditions (2.44),(2.47), and (2.48), along with (2.1) when $r = 2$, allows the limiting Gaussian measures γ_k to exist on c_0. Moreover, without such assumptions, with the most important being (2.44) and (2.48), there are examples of triangular arrays of the form indicated when the CLT must fail on c_0, although it may hold on R^∞. Of course, without (2.44), then the CLT will fail even on R^∞.*

Remark 9. *If the constants $c_{n,j}, k_{n,j}$ are such that $k_{n,j}/(16c_{n,j}) \geq \delta_j L(j+3)$ uniformly in $n \geq 1$ for some $\delta_j > 0$ and $\lim_{j\to\infty} \delta_j = \infty$, then it is easy to see that (2.48) holds for all $\delta > 0$.*

Our next result is a central limit theorem in $\ell_\rho,\ 2 \leq \rho < \infty$. As in Theorem 6, this result holds when the underlying process is a triangular array with random row lengths and possibly missing data.

Theorem 7. *Let $\{\boldsymbol{X}_{n,i} : 1 \leq i \leq n\}$ be as in (1.2), and assume (2.1) holds with $r = 2$ and that $c_{n,j}, k_{n,j}$ are constants such that $c_{n,j} \geq 1,\ k_{n,j} < \infty$ and*

$$\sup_{n,j\geq 1} c_{n,j}/k_{n,j} < \infty. \tag{2.51}$$

Also assume for some $\rho,\ 2 \leq \rho < \infty$, we have

$$\lim_{d\to\infty} \sup_{n\geq 1} \sum_{j\geq d} (c_{n,j}/k_{n,j})^{\rho/2} = 0. \tag{2.52}$$

If $\tilde{\mathbf{S}}_{n,n}$ is given as in (1.7), then

$$\{\mathcal{L}(\tilde{\mathbf{S}}_{n,n}) :\ n \geq 1\} \ \text{ is tight in } \ell_\rho. \tag{2.53}$$

In addition, if $\tilde{\boldsymbol{T}}_{n,n}$ is as in (1.11), and for each $j \geq 1$ we have $\lim_{n\to\infty} P(N_{n,1} < j) = 0$, then $\tilde{\boldsymbol{T}}_{n,n}$ converges in probability to zero in ℓ_ρ. Furthermore, if we also assume (2.44), (2.45), and (2.46) hold, and for each $d < \infty$ we have $P(\min_{1\leq i\leq n} N_{n,i} < d) = o(1/n^2)$ as n tends to infinity, then $\Gamma(k,\cdot,\cdot)$ is the covariance of a centered Gaussian measure γ_k on ℓ_ρ for $k = 1, 2$, and

$$\mathcal{L}(\tilde{\mathbf{S}}_{n,n}) \ \text{ converges weakly to } \gamma_1 \tag{2.54}$$

on ℓ_ρ.

Remark 10. *If the $V_{n,j}^{1/2}$ are replaced by $n^{1/2}$, then under the conditions of Theorem 7 both (2.53) and (2.54) hold with limiting measure γ_2. We also have $\tilde{\boldsymbol{T}}_{n,n}$ converges in probability to zero in this situation. This can easily be seen from a simplification of the proof we give for the random normalizers $V_{n,j}^{1/2}$ and, as one might expect, even more is true. That is, one could apply known sufficient* *conditions for triangular arrays of independent random vectors to satisfy the CLT in ℓ_ρ, $2 \le \rho < \infty$. Examples of such results and related references can be found in [2], page 206-207. However, in order to continue in the spirit of this paper, and to phrase our conditions in terms of the random variables $\{\xi_{n,i,j}\}$ as much as possible, we present an alternative result. It is certainly not best possible, but instead emphasizes obtaining basic sufficient conditions in terms of the moments of these random variables. In particular, it replaces (2.1) with r=2, (2.51), and (2.52) by these moment assumptions.*

Remark 11. *Let $\{\boldsymbol{X}_{n,i} : 1 \le i \le n\}$ be as in (1.2) and assume for some $\delta > 0$ that*

$$\sup_{n,i,j\ge 1} E(|\xi_{n,i,j}|^{2+\delta}) < \infty \tag{2.55}$$

and for some $\rho,\ 2 \le \rho < \infty$,

$$\lim_{d\to\infty} \sup_{n,i\ge 1} \sum_{j\ge d} (E(|\xi_{n,i,j}|^{\rho}))^{2/\rho} = 0. \tag{2.56}$$

If $\boldsymbol{S}_{n,n}$ is given as in (1.6), then

$$\{\mathcal{L}(\frac{\boldsymbol{S}_{n,n}}{n^{1/2}}) :\ n \ge 1\} \ \text{ is tight in } \ell_\rho. \tag{2.57}$$

Furthermore, if we also assume (2.44), (2.45), and (2.46) hold, and for each $d < \infty$ we have $P(\min_{1\le i\le n} N_{n,i} < d) = o(1/n^2)$ as n tends to infinity, then $\Gamma(2,\cdot,\cdot)$ is the covariance of a centered Gaussian measure γ_k on ℓ_ρ , and

$$\mathcal{L}(\frac{\boldsymbol{S}_{n,n}}{n^{1/2}}) \ \text{ converges weakly to } \gamma_2 \tag{2.58}$$

on ℓ_ρ.

Remark 12. *The conditions in (2.55) and (2.56) follow from tail conditions of polynomial type. Hence they require far less than (2.1) with $r = 2$, but in contrast, it is also easy to see that (2.52), and also (2.56), are much more restrictive than the analogue in the c_0 setting given in (2.48)*

3 Some Probability Estimates

Here we provide some basic probability estimates used throughout the paper. The first lemma deals with the sub-Gaussian situation, and the inequality we present for bounded random variables is not best possible, as slightly better constants in the basic estimate can be obtained from Theorem 1 in [9]. Nevertheless, we include a proof for this case, as our argument generalizes to the unbounded case. Our approach is to compute the necessary Laplace transforms, and then use Markov's inequality efficiently. This is standard for such problems, but in order to proceed from first principles, and also keep track of relevant constants, we include the details.

Lemma 1. *Let $X_1, \cdots, X_n$ be independent random variables with $E(X_i) = 0$. If $P(X_i \in [a, b]) = 1$ for $1 \le i \le n$, then*

$$P(|\sum_{i=1}^{n} X_i|/n \ge x) \le 2\exp\{-nx^2(2(b-a)^2)^{-1}\} \tag{3.1}$$

for all $x \ge 0$. In particular, when $n = 1$ each X_i is sub-Gaussian with relevant constants $c = 2$ and $k = (2(b-a)^2)^{-1}$. If

$$P(|X_i| \ge x) \le ce^{-kx^2} \tag{3.2}$$

for $1 \le i \le n$ *and all* $x \ge 0$, *then*

$$P(|\sum_{i=1}^{n} X_i|/n \ge x) \le 2\exp\{-nkx^2/(16c)\} \tag{3.3}$$

for all $x \ge 0$.

Proof. First observe that if Y is a mean zero random variable, then Jensen's inequality implies $E(e^{tY}) \ge e^{tE(Y)} = 1$ for all real t. Thus for Y_1, Y_2 independent copies of Y we have $E((Y_1 - Y_2)^l) = 0$ for l odd, and therefore

$$E(e^{tY}) \le E(e^{tY_1})E(e^{-tY_2}) = E(e^{t(Y_1-Y_2)}) = 1 + \sum_{l\ge 1} t^{2l} E((Y_1 - Y_2)^{2l})/(2l)!. \tag{3.4}$$

If $P(X_i \in [a, b]) = 1$ for $1 \le i \le n$ then $E((Y_1 - Y_2)^{2l}) \le (b - a)^{2l}$ and since $(2l)! \ge 2^l (l!)^2$ for $l \ge 1$ we therefore have

$$E(e^{tY}) \le 1 + \sum_{l\ge 1} t^{2l}(b - a)^{2l}/(2^l l!) = e^{t^2(b-a)^2/2}.$$

Applying this estimate to each of the $X_i's$ for $1 \le i \le n$, the independence of the $X_i's$ and Markov's inequality implies that for each $t \ge 0$ we have

$$P(\sum_{i=1}^{n} X_i/n \ge x) \le e^{-ntx} \prod_{i=1}^{n} E(e^{tX_i}) \le e^{-n(tx - t^2(b-a)^2/2)}.$$

Since $x \ge 0$, minimizing the right hand term over $t \ge 0$ we take $t = x/(b - a)^2$, and hence

$$P(\sum_{i=1}^{n} X_i/n \ge x) \le e^{-nx^2/(2(b-a)^2)}$$

Applying the previous argument to $-\sum_{i=1}^{n} X_i$ we thus have (3.1).

To prove (3.3), we first show that if $E(Y) = 0$ and

$$P(|Y| \ge x) \le ce^{-kx^2} \tag{3.5}$$

holds for all $x \ge 0$, then

$$E(e^{tY}) \le e^{4ct^2/k} \tag{3.6}$$

for all $t \ge 0$.

To verify (3.6) let Y_1, Y_2 be independent copies of Y. Then

$$E((Y_1 - Y_2)^{2l}) = \int_0^\infty P((Y_1 - Y_2)^{2l} \ge x)dx = \int_0^\infty P(|Y_1 - Y_2| \ge x^{1/(2l)})dx,$$

and since $P(|Y_1 - Y_2| \ge x) \le 2P(|Y_1| \ge x/2)$, we thus have

$$E((Y_1 - Y_2)^{2l}) \le 2\int_0^\infty P(|Y_1| \ge 2^{-1}x^{1/(2l)})dx \le 2c\int_0^\infty e^{-4^{-1}kx^{(1/l)}}dx.$$

Taking $s = 4^{-1}kx^{1/l}$ in this last integral, and recalling that $c \ge 1$ in (3.5), we have

$$E((Y_1 - Y_2)^{2l}) \le 2c(4/k)^l l \int_0^\infty e^{-s}s^{l-1}ds = 2c(4/k)^l l! \le (8c/k)^l l!.$$

Thus for all real t we have

$$E(e^{tY}) \leq 1 + \sum_{l \geq 1} (t^2)^l l!(8c/k)^l/(2l)! \leq e^{4ct^2/k}.$$

Applying the previous inequality, independence, and Markov's inequality as before, we have for all $x, t \geq 0$ that

$$P(\sum_{i=1}^{n} X_i \geq nx) \leq \exp\{-n(tx - 4ct^2/k)\}.$$

Minimizing the right hand side over $t \geq 0$ we take $t = xk/(8c)$ and hence

$$P(\sum_{i=1}^{n} X_i \geq nx) \leq \exp\{-nkx^2/(16c)\}.$$

Applying the previous argument to $-\sum_{i-1}^{n} X_i$, we thus have (3.3), and the lemma is proven.

The next two lemmas apply Lemma 1 to provide probability estimates in the sup-norm.

Lemma 2. *Let $\{\mathbf{X}_{n,i} : 1 \leq i \leq n\}$ be defined as in (1.2), and assume (2.1) holds for $r = 2$, and the constants $c_{n,j}, k_{n,j}$ are such that $1 \leq c_{n,j}$ and $0 < k_{n,j} \leq \infty$. If $Q_d(\mathbf{x}) = \sum_{j \geq d+1} x_j \mathbf{e}_j$ for $\mathbf{x} \in R^\infty$, and $\tilde{\mathbf{S}}_{n,n}$ is given as in (1.7), then for all $d \geq 0$ and $\delta > 0$*

$$P(||Q_d(\tilde{\mathbf{S}}_{n,n})||_\infty \geq \delta) \leq \sum_{j \geq d+1} 2\exp\{-\delta^2 k_{n,j}/(16c_{n,j})\}. \tag{3.7}$$

In addition, if the $V_{n,j}$ are replaced by $n^{1/2}$ in $\tilde{\mathbf{S}}_{n,n}$, then again (3.7) holds.

Proof. We first establish (3.7) for general $V_{n,j}$. When the $V_{n,j}$ are replaced by $n^{1/2}$, the result then follows by an immediate simplification of this argument.

If $\theta_{n,i,j} = I(j \leq N_{n,i})R_{n,i,j}$ as indicated, then $P(\theta_{n,i,j} = 1) = p_{n,j}p$, where $p_{n,j} = P(j \leq N_{n,i})$ for $n \geq 1, j \geq 1$, and for $k = 0, 1, \cdots, n$ we define the events

$$E_{k,n,j} = \cup_{I \in \mathcal{I}_{k,n,j}} F_I, \tag{3.8}$$

where $\mathcal{I}_{k,n,j}$ denotes all subsets $I = \{i_1, \cdots, i_k\}$ of size k in $\{1, \cdots, n\}$ and

$$F_I = \{\theta_{n,i,j} = 1 \text{ for all } i \in I \text{ and } \theta_{n,i,j} = 0 \text{ for } i \in \{1, \cdots, n\} \cap I^c\}.$$

Note that F_I depends on n and j, but we suppress that in our notation.

Since $V_{n,j} = \max\{1, \sum_{i=1}^{n} \theta_{n,i,j}\}$ and $\sum_{i=1}^{n} \xi_{n,i,j}\theta_{n,i,j} = 0$ on $E_{0,n,j}$, we therefore have for each $\delta > 0,\ n \geq 1$, and $d \geq 0$ that

$$P(||Q_d(\tilde{\mathbf{S}}_{n,n})||_\infty \geq \delta) \leq \sum_{j \geq d+1} \sum_{k=1}^{n} P(\{|\sum_{i=1}^{n} \xi_{n,i,j}\theta_{n,i,j}| \geq \delta V_{n,j}^{1/2}\} \cap E_{k,n,j}).$$

Now

$$P(\{|\sum_{i=1}^{n} \xi_{n,i,j}\theta_{n,i,j}| \geq \delta V_{n,j}^{1/2}\} \cap E_{k,n,j}) = \sum_{I \in \mathcal{I}_{k,n,j}} P(\{|\sum_{l=1}^{k} \xi_{n,i_l,j}| \geq \delta k^{1/2}\} \cap F_I,\ I = \{i_1, \cdots, i_k\}),$$

and letting

$$A_n = \sum_{j \geq d+1} \sum_{k=1}^{n} P(\{|\sum_{i=1}^{n} \xi_{n,i,j}\theta_{n,i,j}| \geq \delta V_{n,j}^{1/2}\} \cap E_{k,n,j}),$$

we have by using the independence of the various sequences of random variables involved that

$$A_n = \sum_{j\geq d+1}\sum_{k=1}^{n}\sum_{I\in\mathcal{I}_{k.n,j}} P(\{|\sum_{l=1}^{k}\xi_{n,i_l,j}| \geq \delta k^{1/2}\})P(F_I,\ I=\{i_1,\dots,i_k\})$$

$$\leq 2\sum_{j\geq d+1}\sum_{k=1}^{n}\exp\{-\delta^2 k_{n,j}/16c_{n,j}\}P(E_{k,n,j}).$$

Of course, in the previous inequality we are applying (3.3) of Lemma 1 to estimate $P(\{|\sum_{l=1}^{k}\xi_{n,i_l,j}| \geq \delta k^{1/2}\})$. Thus we have

$$P(||Q_d(\tilde{\mathbf{S}}_{n,n})||_\infty \geq \delta) \leq \sum_{j\geq d+1} 2\exp\{-\delta^2 k_{n,j}/(16c_{n,j})\},$$

which proves (3.7) for general $V_{n,j}$.

When the $V_{n,j}^{1/2}$ are replaced by $n^{1/2}$, the proof is immediate since the random variables $\{\xi_{n,i,j}\theta_{n,i,j} : n\geq 1, i\geq 1, j\geq 1\}$ also satisfy (2.1), and hence one can apply (3.3) of Lemma 1 immediately to obtain (3.7). Hence the lemma is proven.

In order that the probability estimate in the previous lemma be useful $k_{n,j}/c_{n,j}$ must be unbounded as j tends to infinity. Our next task is to see what happens if we remove this assumption, and only ask that this ratio is uniformly bounded below by a strictly positive constant. This is the content of our next lemma, which is a modification of Lemma 2.

Lemma 3. *Let $\{\mathbf{X}_{n,i} : 1\leq i\leq n\}$ be defined as in (1.2), and assume (2.1) holds for $r=2$, and the constants $c_{n,j}, k_{n,j}$ are such that $1\leq c_{n,j}\leq c<\infty$ and $0<k\leq\ k_{n,j}\leq\infty$. If $\tilde{\mathbf{S}}_{n,n}$ is given as in (1.7), and $N_n^* = \max_{1\leq i\leq n} N_{n,i}$, then*

$$P(||\tilde{\mathbf{S}}_{n,n})||_\infty \geq x) \leq 2E(N_n^*)\exp\{-\frac{kx^2}{16c}\}. \tag{3.9}$$

In addition, if the $V_{n,j}$ are replaced by $n^{1/2}$ in $\tilde{\mathbf{S}}_{n,n}$, then again (3.9) holds.

Remark 13. *If $N_{n,i} = p_n$ for $\{i\geq 1,\ n\geq 1\}$, then (3.9) immediately implies*

$$P(||\tilde{\mathbf{S}}_{n,n}||_\infty \geq x) \leq 2p_n\exp\{-\frac{kx^2}{16c}\}, \tag{3.10}$$

and if the $V_{n,j}^{1/2}$ are replaced by $n^{1/2}$ in $\tilde{\mathbf{S}}_{n,n}$, then again (3.10) holds.

Proof of Lemma 3. Following the proof of Lemma 2 we observe that if $\theta_{n,i,j} = I(j\leq N_{n,i})R_{n,i,j}$, then $P(\theta_{n,i,j}=1) = p_{n,j}p$, where $p_{n,j} = P(j\leq N_{n,i})$ for $n\geq 1, j\geq 1$, and for $m=0,1,\cdots,n$ we define the events

$$E_{m,n,j} = \cup_{I\in\mathcal{I}_{m,n,j}} F_I,$$

where $\mathcal{I}_{m,n,j}$ denotes all subsets $I=\{i_1,\cdots,i_m\}$ of size m in $\{1,\cdots,n\}$ and

$$F_I = \{\theta_{n,i,j}=1 \text{ for all } i\in I \text{ and } \theta_{n,i,j}=0 \text{ for } i\in\{1,\cdots,n\}\cap I^c\}.$$

Recall $V_{n,j} = \max\{1,\sum_{i=1}^{n}\theta_{n,i,j}\}$ and observe that $\sum_{i=1}^{n}\xi_{n,i,j}\theta_{n,i,j} = 0$ on $E_{0,n,j}$. Hence for each $x>0,\ n\geq 1$,

$$P(||\tilde{\mathbf{S}}_{n,n}||_\infty \geq x) \leq \sum_{u\geq 1}\sum_{j=1}^{u}\sum_{m=1}^{n} P(\{|\sum_{i=1}^{n}\xi_{n,i,j}\theta_{n,i,j}| \geq xV_{n,j}^{1/2}\}\cap E_{m,n,j}\cap\{N_n^*=u\}).$$

Now

$$P(\{|\sum_{i=1}^{n}\xi_{n,i,j}\theta_{n,i,j}| \geq xV_{n,j}^{1/2}\}\cap E_{m,n,j}\cap\{N_n^* = u\}) = \sum_{I\in\mathcal{I}_{m,n,j}} P(\{|\sum_{l=1}^{m}\xi_{n,i_l,j}| \geq xm^{1/2}\}\cap F_I\cap\{N_n^* = u\}\},$$

and letting

$$A_n = \sum_{u\geq 1}\sum_{j=1}^{u}\sum_{m=1}^{n} P(\{|\sum_{i=1}^{n}\xi_{n,i,j}\theta_{n,i,j}| \geq xV_{n,j}^{1/2}\} \cap E_{m,n,j} \cap \{N_n^* = u\}),$$

we have by using the independence of the various sequences of random variables involved and (3.3) of Lemma 1 that

$$A_n = \sum_{u\geq 1}\sum_{j=1}^{u}\sum_{m=1}^{n}\sum_{I\in\mathcal{I}_{m.n,j}} P(\{|\sum_{l=1}^{m}\xi_{n,i_l,j}| \geq xm^{1/2}\})P(\{F_I,\ I = \{i_1,\ldots,i_m\}\} \cap \{N_n^* = u\})$$

$$\leq 2\sum_{u\geq 1}\sum_{j=1}^{u}\sum_{m=1}^{n}\exp\{-x^2k_{n,j}/16c_{n,j}\}P(E_{m,n,j} \cap \{N_n^* = u\}).$$

Thus we have

$$P(||\tilde{\mathbf{S}}_{n,n}||_\infty \geq x) \leq 2\sum_{u\geq 1}\sum_{j=1}^{u}\exp\{-x^2k_{n,j}/(16c_{n,j})\}P(N_n^* = u) \qquad (3.11)$$

$$\leq 2\sum_{u\geq 1}\sum_{j=1}^{u}\exp\{-x^2k/(16c)\}P(N_n^* = u),$$

where the last inequality follows since $c_{n,j} \leq c$ and $k_{n,j} \geq k$ for all $n, j \geq 1$. Hence this implies

$$P(||\tilde{\mathbf{S}}_{n,n}||_\infty \geq x \leq 2E(N_n^*)\exp\{-x^2k/(16c)\}. \qquad (3.12)$$

Therefore (3.9) holds, and when the $V_{n,j}^{1/2}$ are replaced by $n^{1/2}$, the proof follows from the ideas used in the general case. That is, by following the previous argument one obtains the analogue of A_n to be

$$A_n = \sum_{u\geq 1}\sum_{j=1}^{u}\sum_{m=1}^{n}\sum_{I\in\mathcal{I}_{m.n,j}} P(\{|\sum_{l=1}^{m}\xi_{n,i_l,j}| \geq xn^{1/2}\})P(\{F_I,\ I = \{i_1,\ldots,i_m\}\} \cap \{N_n^* = u\}).$$

Hence by applying (3.3) of Lemma 1 we have

$$A_n \leq 2\sum_{u\geq 1}\sum_{j=1}^{u}\sum_{m=1}^{n}\exp\{-x^2k_{n,j}n/(16mc_{n,j})\}P(E_{m,n,j} \cap \{N_n^* = u\})$$

$$\leq 2\sum_{u\geq 1}\sum_{j=1}^{u}\sum_{m=1}^{n}\exp\{-x^2k/(16c)\}P(E_{m,n,j} \cap \{N_n^* = u\}),$$

where the last inequality holds since $n \geq m$, $c_{n,j} \leq c$, and $k_{n,j} \geq k$ here. Therefore we again have

$$P(||\tilde{\mathbf{S}}_{n,n}||_\infty \geq x) \leq 2E(N_n^*)\exp\{-x^2k/(16c)\}$$

when the $V_{n,j}^{1/2}$ are replaced by $n^{1/2}$, and the lemma is proven.

Next we turn to a method which will allow us to handle a broader collection of random variables. Here the $\{\xi_{n,i,j}\}$ satisfy (2.1) with $r \in (0,2)$, or the less restrictive conditions of polynomial decay given in (2.4). Of course, the results depend on the rate of decay of the tails of the $\{\xi_{n,i,j}\}$, but under a variety of assumptions we are able to obtain further consistency results in this setting. The relevant probability inequalities are obtained in our next lemma, and can be viewed as a substitute for those in Lemma 1.

Lemma 4. *For each integer $n \geq 1$ let $X_1, \cdots, X_n$ be independent, mean zero random variables, such that for some $r \in (0,2)$ we have*

$$P(|X_i| \geq x) \leq ce^{-kx^r} \tag{3.13}$$

for $1 \leq i \leq n$ and all $x \geq 0$. Then for all $x \geq \sqrt{8}M_{c,k,r}/\sqrt{n}$ and all $s \geq 0$

$$P(|\sum_{i=1}^n X_i| \geq nx) \leq 4\exp\{-\frac{nx^2}{32s^2}\} + 4cn\exp\{-\frac{ks^r}{2^r}\}, \tag{3.14}$$

where $M_{c,k,r}^2 = \int_0^\infty ce^{-kx^{r/2}}dx < \infty$. In addition, for all $x \geq \sqrt{8}M_{c,k,r}/\sqrt{n}$ and all $s \geq 1$, we also have

$$P(|\sum_{i=1}^n X_i| \geq nx) \leq 4\exp\{-\frac{nkx^2}{128cs^2}\} + 4cn\exp\{-\frac{ks^r}{2^r}\}. \tag{3.15}$$

Moreover, if for some $c > 0$ and $k > 2$, (3.13) is replaced by

$$P(|X_i| \geq x) \leq \frac{c}{(1+x)^k} \tag{3.16}$$

for $1 \leq i \leq n$ and all $x \geq 0$, then for all $x \geq \sqrt{8}M_{c,k}/\sqrt{n}$ and all $s \geq 0$

$$P(|\sum_{i=1}^n X_i| \geq nx) \leq 4\exp\{-\frac{nx^2}{32s^2}\} + \frac{2^{2+k}cn}{(2+s)^k}, \tag{3.17}$$

where $M_{c,k}^2 = \int_0^\infty \frac{c}{(1+t^{1/2})^k}dt < \infty$.

Remark 14. *If we take $s = n^{\frac{1}{2+r}}x^{\frac{2}{2+r}}$ then for $x \geq \sqrt{8}M_{c,k,r}/\sqrt{n}$ we have that (3.14) implies*

$$P(|\sum_{i=1}^n X_i| \geq nx) \leq 4\exp\{-\frac{n^{\frac{r}{2+r}}x^{\frac{2r}{2+r}}}{32}\} + 4cn\exp\{-\frac{kn^{\frac{r}{2+r}}x^{\frac{2r}{2+r}}}{2^r}\},$$

which makes the exponents on the right of comparable size. Since the proof of (3.14) and (3.15) also implies (3.14) and (3.15) when $r = 2$, it is interesting to note that the previous inequality is not as sharp as that in (3.3) in Lemma 1 when $r = 2$.

Remark 15. *If the median of each X_i is zero, then (3.14) and (3.15) hold for all $x \geq 0$ and $s \geq 0$. That is, when the medians are zero the key inequality (3.20) below follows directly from (5.8) in [6], page 147, without restrictions on x. A similar remark holds for (3.15) provided $s \geq 1$.*

Proof of Lemma 4. First we observe that for r fixed, uniformly in i, $1 \leq i \leq n$, (3.13) implies

$$E(X_i^2) = \int_0^\infty P(|X_i| > t^{1/2})dt \leq M_{c,k,r}^2 < \infty.$$

Hence if $x \geq \sqrt{8}M_{c,k,r}/\sqrt{n}$ we have by Cheyshev's inequality that

$$P(|\sum_{i=1}^{n} X_i| \geq nx/2) \leq 4\sum_{i=1}^{n} E(X_i^2)/(n^2x^2) \leq 1/2. \tag{3.18}$$

Now let $Y_1, \cdots, Y_n$ be an independent copy of $X_1, \cdots, X_n$ and observe that

$$P(|\sum_{i=1}^{n} X_i| \geq nx)P(|\sum_{i=1}^{n} Y_i| \leq nx/2) \leq P(|\sum_{i=1}^{n}(X_i - Y_i)| \geq nx/2). \tag{3.19}$$

Then for all $x \geq \sqrt{8}M_{c,k,r}/\sqrt{n}$, (3.18) and (3.19) combine to imply

$$P(|\sum_{i=1}^{n} X_i| \geq nx) \leq 2P(|\sum_{i=1}^{n}(X_i - Y_i)| \geq nx/2). \tag{3.20}$$

Taking $s \geq 0$ we define $(X_i - Y_i)^s = (X_i - Y_i)I(|X_i - Y_i| \leq s)$. Then

$$P(|\sum_{i=1}^{n}(X_i - Y_i)| \geq nx/2) \leq I_n(s,x) + II_n(s,x) \tag{3.21}$$

where

$$I_n(s,x) = P(|\sum_{i=1}^{n}(X_i - Y_i)^s| \geq nx/2)$$

and

$$II_n(s,x) = P(\max_{1\leq i\leq n} |(X_i - Y_i) - (X_i - Y_i)^s| > 0).$$

Applying (3.1) of Lemma 1 to $(X_1 - Y_1)^s, \cdots, (X_n - Y_n)^s$ we see that

$$I_n(s,x) \leq 2\exp\{-n(x/2)^2(2(2s)^2)^{-1}\}, \tag{3.22}$$

and (3.13) implies

$$II_n(s,x) \leq \sum_{i=1}^{n} P(|X_i - Y_i| > s) \leq 2\sum_{i=1}^{n} P(|X_i| > s/2) \leq 2cn\exp\{-\frac{ks^r}{2^r}\}. \tag{3.23}$$

Applying (3.20),(3.21),(3.22), and (3.23) we thus have

$$P(|\sum_{i=1}^{n} X_i| \geq nx) \leq 4\exp\{-\frac{nx^2}{32s^2}\} + 4cn\exp\{-\frac{ks^r}{2^r}\}. \tag{3.24}$$

Thus (3.14) of Lemma 4 is proved.

The proof of (3.15) follows that for (3.14) up to the point we apply (3.1) of Lemma 1 to $I_n(s,x)$ in (3.22). At this point we now apply (3.3) of Lemma 1 to the random variables $(X_1 - Y_1)^s, \cdots, (X_n - Y_n)^s$. That is, (3.13) implies that for all $x \geq 0$ and $1 \leq i \leq n$ that

$$P(|(X_i - Y_i)^s| \geq x) \leq P(|X_i^s| \geq x/2) + P(|Y_i^s| \geq x/2) \leq 2ce^{-\frac{kx^2}{4s^2}},$$

where the last inequality follows since $x^r/2^r \geq x^2/(4s^2)$ when $0 \leq x \leq 2s$, $0 < r < 2$, and $s \geq 1$. Hence by (3.3) of Lemma 1, with k replaced by $k/(4s^2)$ and c by $2c$, we obtain

$$I_n(s,x) \leq 2\exp\{-nkx^2/(128cs^2)\}. \tag{3.25}$$

Now combining (3.25) and the estimate for $II_n(s,x)$ in (3.23) to (3.20) and (3.21), we obtain (3.15).

Next we observe that uniformly in i, $1 \le i \le n$, (3.16) and $k > 2$ implies

$$E(X_i^2) = \int_0^\infty P(|X_i| > t^{1/2})dt \le \int_0^\infty \frac{c}{(1+t^{1/2})^k}dt < \infty.$$

Hence if we choose $x \ge \sqrt{8}M_{c,k,}/\sqrt{n}$ we have by Cheyshev's inequality that

$$P(|\sum_{i=1}^n X_i| \ge nx/2) \le 4\sum_{i=1}^n E(X_i^2)/(n^2x^2) \le 1/2. \tag{3.26}$$

Now let $Y_1, \cdots, Y_n$ be an independent copy of $X_1, \cdots, X_n$ and observe that

$$P(|\sum_{i=1}^n X_i| \ge nx)P(|\sum_{i=1}^n Y_i| \le nx/2) \le P(|\sum_{i=1}^n (X_i - Y_i)| \ge nx/2). \tag{3.27}$$

Then for all $x \ge \sqrt{8}M_{c,k,}/\sqrt{n}$, (3.26) and (3.27) combine to imply

$$P(|\sum_{i=1}^n X_i| \ge nx) \le 2P(|\sum_{i=1}^n (X_i - Y_i)| \ge nx/2), \tag{3.28}$$

Taking $s \ge 0$ we define $(X_i - Y_i)^s = (X_i - Y_i)I(|X_i - Y_i| \le s)$. Then

$$P(|\sum_{i=1}^n (X_i - Y_i)| \ge nx/2) \le I_n(s,x) + II_n(s,x) \tag{3.29}$$

where

$$I_n(s,x) = P(|\sum_{i=1}^n (X_i - Y_i)^s| \ge nx/2)$$

and

$$II_n(s,x) = P(\max_{1\le i\le n} |(X_i - Y_i) - (X_i - Y_i)^s| > 0).$$

Applying (3.1) of Lemma 1 to $(X_1 - Y_1)^s, \cdots, (X_n - Y_n)^s$ we see that

$$I_n(s,x) \le 2\exp\{-n(x/2)^2(2(2s)^2)^{-1}\}, \tag{3.30}$$

and (3.15) implies

$$II_n(s,x) \le \sum_{i=1}^n P(|X_i - Y_i| > s) \le 2\sum_{i=1}^n P(|X_i| > s/2) \le \frac{2cn}{(1+s/2)^k}. \tag{3.31}$$

Applying (3.28),(3.29), (3.30), and (3.31) we thus have

$$P(|\sum_{i=1}^n X_i| \ge nx) \le 4\exp\{-\frac{nx^2}{32s^2}\} + \frac{2^{2+k}cn}{(2+s)^k}. \tag{3.32}$$

Thus Lemma 4 is proven.

4 Proofs of Consistency Results

4.1 Proof of Theorem 1

Applying Lemma 2 with $d = 0$, we have for all $x > 0$ and each integer $n \geq 1$ that

$$P(||\tilde{\mathbf{S}}_{n,n}||_\infty \geq x) \leq \sum_{j\geq 1} 2\exp\{-x^2 k_{n,j}/(16c_{n,j})\}. \tag{4.1}$$

Taking $x = \epsilon a_n$ in (4.1) and applying (2.5), we thus have (2.6) for general $V_{n,j}^{1/2}$, and also when the $V_{n,j}^{1/2}$ are replaced by $n^{1/2}$.

If the constants $c_{n,j}$ and $k_{n,j}$ satisfy (2.7) as indicated, then with $a_n = (L(n+3))^{1/2}$ and $x = \epsilon a_n$ in (4.1) we have

$$P(||\tilde{\mathbf{S}}_{n,n}||_\infty \geq \epsilon(L(n+3))^{1/2}) \leq \sum_{j\geq 1} 2\exp\{-\epsilon^2\delta L(n+3)L(j+3)\} = 2\sum_{j\geq 1}(j+3)^{-\epsilon^2\delta L(n+3)}. \tag{4.2}$$

Thus for $\epsilon^2\delta > 1$

$$\sum_{n\geq 1} P(||\tilde{\mathbf{S}}_{n,n}||_\infty \geq \epsilon(L(n+3))^{1/2}) \leq 2\sum_{n\geq 1}\int_3^\infty x^{-\epsilon^2\delta L(n+3)}dx \leq 2\sum_{n\geq 1}\frac{3^{-L(n+3)-1}}{(L(n+3)-1)} < \infty, \tag{4.3}$$

and hence (2.8) holds for general $V_{n,j}$. In particular, we then have from (4.3) that (2.9) is immediate, and it remains to show $E(e^{\alpha M^2}) < \infty$ for all $\alpha > 0$ sufficiently small. Now

$$E(e^{\alpha M^2}) = \int_0^\infty P(e^{\alpha M^2} > t)dt \leq 3 + \int_3^\infty P(M > (\frac{\log t}{\alpha})^{1/2})dt,$$

and

$$\int_3^\infty P(M > (\frac{\log t}{\alpha})^{1/2})dt \leq \sum_{n\geq 1}\int_3^\infty P(||\tilde{\mathbf{S}}_{n,n}||_\infty \geq (L(n+3))^{1/2}(\frac{\log t}{\alpha})^{1/2})dt$$

$$\leq 2\sum_{n\geq 1}\sum_{j\geq 1}\int_3^\infty \exp\{-\delta L(j+3)L(n+3)\frac{\log t}{\alpha}\}dt,$$

where the last inequality follows from (4.1) and that (2.7) holds. Therefore, for $\alpha < \delta/2$ we have

$$E(e^{\alpha M^2}) \leq 3 + 2\sum_{n\geq 1}\sum_{j\geq 1}\int_3^\infty \exp\{-2L(j+3)L(n+3)\log t\}dt$$

$$= 3 + 6\sum_{n\geq 1}\sum_{j\geq 1}\frac{\exp\{-2\log 3L(j+3)L(n+3)\}}{2L(j+3)L(n+3)-1}.$$

Now $x, y \geq 1+\eta$ for some $\eta > 0$ implies $xy \geq (x+y)(1+\eta)/2$ and hence since $j, n \geq 1$ implies $L(j+3), L(n+3) \geq L4 \geq 1+\eta$ for $\eta = L4 - 1 > 0$ we have

$$E(e^{\alpha M^2}) \leq 3 + 6\sum_{n\geq 1}\sum_{j\geq 1}\frac{\exp\{-(1+\eta)\log 3(L(j+3)+L(n+3))\}}{2L(j+3)L(n+3)-1} < \infty$$

since $(1+\eta)\log 3 > 1$. Since (4.1) holds when the $V_{n,j}^{1/2}$ are replaced by $n^{1/2}$, the proof also holds in this situation. Thus Theorem 1 is proven.

4.2 Proof of Theorem 2

Under our assumptions, Lemma 3 with $x = h(n)$ implies that

$$P(||\tilde{\mathbf{S}}_{n,n}||_\infty \geq h(n)) \leq 2E(N_n^*)\exp\{-\frac{kh(n)^2}{16c}\}.$$

Since $x = h(n) = (\theta_1^{-1}L(E(N_n^*)) + \theta_2 Ln)^{1/2}$, $\theta_1 = k/(16c)$ and $\theta_2 > 0$ we have

$$P(||\tilde{\mathbf{S}}_{n,n}||_\infty \geq h(n)) \leq 2L(E(N_n^*))\exp\{-L(E(N_n^*)) - \theta_1\theta_2 Ln\}.$$

Since $\theta_1 > 0$ we thus have (2.13) if $\theta_2 > 0$, and (2.14) follows immediately provided $\theta_1\theta_2 > 1$. Since the above holds for general $V_{n,j}^{1/2}$, and also the $n^{1/2}$ normalizations, Theorem 2 is proven

4.3 Proof of Theorem 3

Since $E(\xi_{i,j}) = 0$ for all $i, j \geq 1$, Theorem 3 will follow from Theorem 3.1 of [12] with $B = \ell^\infty$, with the usual sup-norm, provided we show $(a) P(\mathbf{X}_i \in \ell^\infty) = 1$ for all $i \geq 1$, (b) $\sup_{i\geq 1} E(e^{\alpha||\mathbf{X}_i||_\infty^2}) < \infty$ for some $\alpha > 0$, and (c) $\mathbf{S}_n/n^{1/2}$ is bounded in probability with respect to the sup-norm.

Now

$$||\mathbf{X}_i||_\infty^2 = \sup_{j\geq 1}|\xi_{i,j}|^2,$$

where (2.26) holds uniformly in i, j. Thus

$$P(||\mathbf{X}_i||_\infty^2 > t) \leq \sum_{j\geq 1} P(|\xi_{i,j}|^2 > t) \leq \sum_{j\geq 1} c_j e^{-k_j t} \leq \sum_{j\geq 1}\frac{k_j}{16\delta L(j+3)}e^{-k_j t},$$

since $k_j/(16c_j) \geq \delta L(j+3)$. Hence

$$E(e^{\alpha||\mathbf{X}_i||_\infty^2}) \leq 3 + \int_3^\infty P(e^{\alpha||\mathbf{X}_i||_\infty^2} > t)dt \leq 3 + \sum_{j\geq 1}\frac{k_j}{16\delta L(j+3)}\int_3^\infty e^{-k_j \log t/\alpha}dt,$$

where $k_j \geq 16\delta L(j+3)$ since $c_j \geq 1$. Taking $\alpha > 0$ sufficiently small so that $k_j/\alpha - 1 > k_j/(2\alpha) > 1$, we then have

$$\begin{aligned} E(e^{\alpha||\mathbf{X}_i||_\infty^2}) &\leq 3 + \sum_{j\geq 1}\frac{k_j}{16\delta L(j+3)}\int_3^\infty t^{-k_j/\alpha}dt \\ &= 3 + \sum_{j\geq 1}\frac{k_j}{16\delta L(j+3)}\frac{3^{-k_j/\alpha+1}}{(k_j/\alpha - 1)} \\ &\leq 3 + 3\sum_{j\geq 1}\frac{2\alpha}{16\delta L(j+3)}3^{-k_j/\alpha} < \infty. \end{aligned}$$

Thus (a) and (b) above hold, and it remains to verify (c).

Now

$$P(||\mathbf{S}_n||_\infty/n^{1/2} \geq x) \leq \sum_{j\geq 1} P(|\sum_{i=1}^n \xi_{i,j}| \geq xn^{1/2}) \leq 2\sum_{j\geq 1}\exp\{-k_j x^2/(16c_j)\}$$

for all $x \geq 0$, where the last inequality follows from Lemma 1. Since we have $k_j/(16c_j) \geq \delta L(j+3)$ for some $\delta > 0$, it follows that

$$P(||\mathbf{S}_n||_\infty/n^{1/2} \geq x) \leq \sum_{j\geq 1}\exp\{-\delta L(j+3)x^2\} = \sum_{j\geq 1}(j+3)^{-\delta x^2},$$

and hence

$$P(||\mathbf{S}_n||_\infty/n^{1/2} \ge x) \le \int_0^\infty (t+3)^{-\delta x^2} dt = 2\frac{3^{-\delta x^2+1}}{\delta x^2 - 1} \le \epsilon,$$

provided $x \ge \beta_{\epsilon,\delta}$. Thus (c) above holds, and Theorem 3.1 of [12] therefore implies (2.29). Hence Theorem 3 holds.

4.4 Proof of Theorem 4

First observe that for all $x \ge 0$ that

$$P(||\mathbf{S}_{n,n}||_\infty \ge nx) = P(\max_{1\le j\le N_n^*} |\sum_{i=1}^n \xi_{n,i,j}\theta_{n,i,j}| \ge nx)$$

$$= P(\bigcup_{j=1}^{N_n^*}\{|\sum_{i=1}^n \xi_{n,i,j}\theta_{n,i,j})| \ge nx\}). \tag{4.4}$$

Letting $\mathbf{b} = (b_1, \cdots, b_n)$, where b_i is a positive integer for $1 \le i \le n$, and setting $E_{n,\mathbf{b}} = \{N_{n,1} = b_1, \cdots, N_{n,n} = b_n\}$ we thus have

$$P(||\mathbf{S}_{n,n}||_\infty \ge nx) = \sum_{(b_1,\cdots,b_n)} P(\bigcup_{j=1}^{\max(b_1,\cdots,b_n)} \{|\sum_{i=1}^n \xi_{n,i,j}\theta_{n,i,j}| \ge nx\}|E_{n,\mathbf{b}})P(E_{n,\mathbf{b}})$$

$$\le \sum_{(b_1,\cdots,b_n)} \sum_{j=1}^{\max(b_1,\cdots,b_n)} P(|\sum_{i=1}^n \xi_{n,i,j}\theta_{n,i,j}| \ge nx|E_{n,\mathbf{b}})P(E_{n,\mathbf{b}}). \tag{4.5}$$

Fixing n and j, and defining $X_i = \xi_{n,i,j}\theta_{n,i,j}I(j \le b_i)$ for $1 \le i \le n$, we see $X_1, \cdots, X_n$ are independent random variables, and it is easy to check from our assumptions on $c_{n,j}$ and $k_{n,j}$, and (2.1), that for all $x \ge 0$

$$P(|X_i| \ge x) \le ce^{-kx^r}.$$

Therefore $X_1, \cdots, X_n$ satisfy the conditions in Lemma 4 and using the independence of the sequences $\{\xi_{n,i,j}\}$, $\{R_{n,i,j}\}$, and $\{N_{n,i}\}$ we have for $x \ge \sqrt{8}M_{c,k,r}/\sqrt{n}$ and $s \ge 1$ that (3.15) implies

$$P(|\sum_{i=1}^n \xi_{n,i,j}\theta_{n,i,j}| \ge nx|E_{n,\mathbf{b}}) = P(|\sum_{i=1}^n \xi_{n,i,j}\theta_{n,i,j}| \ge nx)$$

$$\le 4\exp\{-nkx^2/(128cs^2)\} + 4cn\exp\{-\frac{ks^r}{2^r}\}, \tag{4.6}$$

which combined with (4.5) implies

$$P(||\mathbf{S}_{n,n}||_\infty \ge nx) \le 4\sum_{(b_1,\cdots,b_n)} \sum_{j=1}^{\max(b_1,\cdots,b_n)} [\exp\{-nkx^2/(128cs^2)\} + cn\exp\{-\frac{ks^r}{2^r}\}]P(E_{n,\mathbf{b}}) \tag{4.7}$$

$$= 4E(N_n^*)[\exp\{-nkx^2/(128cs^2)\} + cn\exp\{-\frac{ks^r}{2^r}\}].$$

Recalling $h(n) = (c_2^{-1}L(E(N_n^*)) + c_3Ln)^{1/2}$ and taking $s = s_n = c_1(L(E(N_n^*)) + 2Ln)^{1/r}$ and $x = x_n = s_n\{c_2^{-1}L(E(N_n^*)) + c_3Ln\}^{1/2}/n^{1/2}$ in (4.7), then for all sufficiently large n we have

$x \geq \sqrt{8}M_{c,k,r}/\sqrt{n}$, $s \geq 1$, and (4.7) holds. Thus (2.28) holds if $c_1 > 2/k^{1/r}$ and $c_3 > 0$, and (2.29) follows if we also have $c_2c_3 > 1$. Thus Theorem 4 is proven when (2.1) holds and $0 < r < 2$.

We now turn to the situation where there is only polynomial decay in the tails of the data $\xi_{n,i,j}$ as in (2.4), where $c_{n,j} \leq c$ and $k_{n,j} \geq k$ for all $n \geq 1, j \geq 1$ and $1 \leq c < \infty$, $2 < k < \infty$. Then the random variables $\xi_{n,i,j}\theta_{n,i,j}$ are also easily seen to satisfy (2.4), and arguing as in (4.4-4.7), and applying (3.17)), we have for $s \geq 0$ and $x \geq \sqrt{8}M_{c,k}/n^{1/2}$ that

$$P(||\mathbf{S}_{n,n}||_\infty \geq nx) \leq 4E(N_n^*)[\exp\{-\frac{nx^2}{32s^2}\} + \frac{2^k cn}{(2+s)^k}]. \tag{4.8}$$

Taking $s = s_n = (nE(N_n^*))^{\frac{1}{k}+\beta}$, $\beta > 0$, and $x = x_n = bs_n(LE(N_n^*))^{1/2}/n^{1/2}$, then for all n sufficiently large

$$P(||\mathbf{S}_{n,n}||_\infty \geq bn^{1/2}s_n(LE(N_n^*))^{1/2}) \leq 4E(N_n^*)[\exp\{-\frac{b^2}{32}L(E(N_n^*))\} + \frac{2^k cn}{(2+(nE(N_n^*))^{\frac{1}{k}+\beta})^k}].$$

Thus (2.30) holds when $\beta > 0$ by taking b large. Moreover, (2.31) holds if $b \geq 8$ and $k\beta > 1/2$, and (2.32) holds when if $b > 8$ and $\beta k(\gamma+1) > 1$. Thus Theorem 4 is proven.

4.5 Proof of Theorem 5

If $\delta = 0$, then $N_{n,i} = d(n)$ for $i \geq 1$, and since $p = 1$ it is easy to check that $\tilde{\mathbf{T}}_{n,n} = \mathbf{S}_{n,n}/n$. Hence the result follows from what has been proven previously in Theorem 4, and we need only prove the result when $0 < \delta < 1/4$. However, when comparing Theorems 4 and 5, be careful to notice that $h(n)$ is used differently in these results.

Let $\tilde{\mathbf{T}}_{n,n}$ be defined by (2.38) and let λ_n equal the number of $N_{n,i}$, $1 \leq i \leq n$, that are equal to $N_n^* = \max_{1\leq i\leq n} N_{n,i}$. Then

$$P(||\tilde{\mathbf{T}}_{n,n}||_\infty \geq x) = \sum_{\rho=1}^{n} P(||\tilde{\mathbf{T}}_{n,n}||_\infty \geq x, \lambda_n = \rho) \leq I_n + II_n, \tag{4.9}$$

where

$$I_n = \sum_{\rho=1}^{m(n)} P(\lambda_n = \rho), \tag{4.10}$$

and

$$II_n = \sum_{\rho=m(n)+1}^{n} P(||\tilde{\mathbf{T}}_{n,n}||_\infty \geq x, \lambda_n = \rho). \tag{4.11}$$

In order to estimate I_n we observe that

$$P(\lambda_n = \rho) = C_{n,\rho}P(N_{n,n-\rho+1} = \cdots = N_{n,n} > \max_{1\leq i\leq n-\rho} N_{n,i}),$$

where $C_{n,\rho} = n!/((n-\rho)!\rho!)$. Therefore, since the $N_{n,i}$, $1 \leq i \leq n$, are i.i.d and their range is in $\{1, 2, \cdots, d(n)\}$, with $P(N_{n,1} = d(n)) = q_n > 0$, we have

$$P(\lambda_n = \rho) = \sum_{j=2}^{d(n)} C_{n,\rho}P(\max_{1\leq i\leq n-\rho} N_{n,i} < j)P(N_{n,n-\rho+1} = \cdots = N_{n,n} = j)$$

$$\leq C_{n,\rho}(1-q_n)^{n-\rho}\sum_{j=2}^{d(n)} P(N_{n,1}=j)^\rho \leq C_{n,\rho}(1-q_n)^{n-\rho}.$$

Thus

$$P(\lambda_n=\rho)\leq n^\rho \exp\{-(n-\rho)q_n\},$$

and hence

$$I_n \leq m(n)n^{m(n)}\exp\{-(n-m(n))q_n\}. \tag{4.12}$$

To estimate II_n we observe by (2.38) that

$$P(||\tilde{\mathbf{T}}_{n,n}||_\infty \geq x) = P(\max_{1\leq j\leq N_n^*}|\sum_{i=1}^n \xi_{n,i,j}I(j\leq N_{n,i})|\geq V_{n,j}x)$$

$$= P(\bigcup_{j=1}^{N_n^*}\{|\sum_{i=1}^n \xi_{n,i,j}I(j\leq N_{n,i})|\geq V_{n,j}x\}).$$

Letting $\mathbf{b}=(b_1,\cdots,b_n)$, where b_i is a positive integer for $1\leq i\leq n$, and setting $E_{n,\mathbf{b},\rho}=\{N_{n,1}=b_1,\cdots,N_{n,n}=b_n,\lambda_n=\rho\}$ we thus have

$$P(||\tilde{\mathbf{T}}_{n,n}||_\infty \geq x,\lambda_n=\rho) = \sum_{(b_1,\cdots,b_n)\in B_{n,\rho}} P(\bigcup_{j=1}^{\max(b_1,\cdots,b_n)}\{|\sum_{i=1}^n \xi_{n,i,j}I(j\leq N_{n,i})|\geq V_{n,j}x\}|E_{n,\mathbf{b},\rho})P(E_{n,\mathbf{b},\rho})$$

$$\leq \sum_{(b_1,\cdots,b_n)\in B_{n,\rho}}\sum_{j=1}^{\max(b_1,\cdots,b_n)} P(|\sum_{i=1}^n \xi_{n,i,j}I(j\leq N_{n,i})|\geq V_{n,j}x|E_{n,\mathbf{b},\rho})P(E_{n,\mathbf{b},\rho}), \tag{4.13}$$

where $B_{n,\rho}=\{(b_1,\cdots,b_n):$ there are exactly ρ largest $b_i's\}$. Now if $\lambda_n=\rho$, then $V_{n,j}\geq\rho$ for $1\leq j\leq N_n^*$, and for n,j fixed, the random variables $\{\xi_{n,i,j}I(j\leq N_{n,i}):1\leq i\leq n\}$ satisfy the conditions in Lemma 4 with $0<r<2$. Hence with x in (3.15) replaced by $xV_{n,j}/n$, we have

$$P(|\sum_{i=1}^n \xi_{n,i,j}I(j\leq N_{n,i})|\geq xV_{n,j}|E_{n,\mathbf{b},\rho}) = P(|\sum_{i=1}^n \xi_{n,i,j}I(j\leq N_{n,i})|\geq n(xV_{n,j}/n)|E_{n,\mathbf{b},\rho}) \tag{4.14}$$

$$\leq 4[\exp\{-\frac{kx^2V_{n,j}^2}{128cns^2}\}+cn\exp\{-\frac{ks^r}{2^r}\}]$$

$$\leq 4[\exp\{-\frac{kx^2\rho^2}{128cns^2}\}+cn\exp\{-\frac{ks^r}{2^r}\}],$$

provided $V_{n,j}\geq\rho$, $s\geq 1$, and $xV_{n,j}/n\geq\sqrt{8}M_{c,k,r}/\sqrt{n}$.

Combining (4.14) with (4.13) we have

$$P(||\tilde{\mathbf{T}}_{n,n}||_\infty \geq x,\lambda_n=\rho) \leq 4\sum_{(b_1,\cdots,b_n)\in B_{n,\rho}}\sum_{j=1}^{\max(b_1,\cdots,b_n)}[\exp\{-\frac{kx^2\rho^2}{128cns^2}\}+cn\exp\{-\frac{ks^r}{2^r}\}]P(E_{n,\mathbf{b},\rho})$$

$$\leq 4\sum_{(b_1,\cdots,b_n)}\sum_{j=1}^{\max(b_1,\cdots,b_n)}[\exp\{-\frac{kx^2\rho^2}{128cns^2}\}+cn\exp\{-\frac{ks^r}{2^r}\}]P(E_{n,\mathbf{b},\rho}) \tag{4.15}$$

$$= 4E(N_n^*|\lambda_n=\rho)[\exp\{-\frac{kx^2\rho^2}{128cns^2}\}+cn\exp\{-\frac{ks^r}{2^r}\}]P(\lambda_n=\rho).$$

Hence

$$II_n \leq 4 \sum_{\rho=m(n)+1}^{n} E(N_n^*|\lambda_n = \rho)[\exp\{-\frac{kx^2\rho^2}{128cns^2}\} + cn\exp\{-\frac{ks^r}{2^r}\}]P(\lambda_n = \rho) \tag{4.16}$$

$$\leq 4E(N_n^*)[\exp\{-\frac{kx^2m^2(n)}{128cns^2}\} + cn\exp\{-\frac{ks^r}{2^r}\}$$

Combining (4.9), (4.12), and (4.16) we see

$$P(||\tilde{\mathbf{T}}_{n,n}||_\infty \geq x) \leq m(n)n^{m(n)}\exp\{-(n-m(n))q_n\} + 4E(N_n^*)[\exp\{-\frac{kx^2m^2(n)}{128cns^2}\} + cn\exp\{-\frac{ks^r}{2^r}\}]$$

provided $s \geq 1$, and $xm(n)/n \geq \sqrt{8}M_{c,k,r}/\sqrt{n}$.

Setting $s = s_n = c_1(L(E(N_n^*)) + Ln)^{1/r}$,

$$x = h(n) = s_n(\frac{n[c_2^{-1}L(E(N_n^*)) + c_3Ln]}{m^2(n)})^{1/2},$$

where $c_1 > 4/k^{1/r}$, $c_2 = k/(128c)$, $c_3 > 0$, $q_n \geq n^{-\delta}$, and $m(n) = n^{1-2\delta}$, we have $xm(n)/n \geq \sqrt{8}M_{c,k,r}/\sqrt{n}$ and $s \geq 1$ for all n sufficiently large. Furthermore, from the previous inequality and that $0 < \delta < 1/4$, we have for all n sufficiently large that

$$P(||\tilde{\mathbf{T}}_{n,n}||_\infty \geq s_n[\frac{(128c/k)L(E(N_n^*)) + c_3Ln}{n^{1-4\delta}}]^{1/2}) \leq \exp\{(1-2\delta+n^{1-2\delta})Ln - (n^{1-\delta} - n^{1-3\delta})\}$$

$$+ 4E(N_n^*)[\exp\{-\frac{kx^2m^2(n)}{128cns^2}\} + cn\exp\{-\frac{ks^r}{2^r}\}]$$

$$\leq \exp\{-\frac{n^{1-\delta}}{2}\} + 4[\exp\{-c_2c_3Ln\} + c\exp\{-(1+\eta)Ln\}],$$

where $\eta > 0$ is sufficiently small. Hence once $\delta \in (0, 1/4)$ is specified, then for each such δ we have (2.41) provided $c_1 > 4/k^{1/r}$ and $c_3 > 0$, and (2.42) provided $c_1 > 4/k^{1/r}$ and $c_2c_3 > 1$. Thus Theorem 5 is proven.

5 Proofs of Asymptotic Normality Results

The proofs proceed with a sequence of lemmas.

5.1 Proof of Theorem 6

Our first lemma provides tightness in Theorem 6, and shows $\tilde{\boldsymbol{T}}_{n,n}$ converges in probability to zero in c_0.

Lemma 5. *Under the conditions (2.47) and (2.48) of Theorem 6,*

$$\{\mathcal{L}(\tilde{\mathbf{S}}_{n,n}) : \ n \geq 1\} \text{ is tight in } c_0. \tag{5.1}$$

In addition, if for each $j \geq 1$ we have $\lim_{n\to\infty} P(N_{n,1} < j) = 0$, then $\tilde{\boldsymbol{T}}_{n,n}$ converges in probability to zero in c_0.

Proof. For general $V_{n,j}$, or if the $V_{n,j}^{1/2}$ are replaced by $n^{1/2}$, (3.7) of Lemma 2 implies that

$$P(||Q_d(\tilde{\mathbf{S}}_{n,n})||_\infty \geq \delta) \leq \sum_{j\geq d+1} 2\exp\{-\delta^2 k_{n,j}/(16c_{n,j})\}. \tag{5.2}$$

Hence (2.48) implies for $\delta > 0$ arbitrary that

$$\lim_{d\to\infty}\sup_{n\geq 1} P(||Q_d(\tilde{\mathbf{S}}_{n,n})||_\infty > \delta) = 0. \tag{5.3}$$

Now (2.1) easily implies

$$E(\xi_{n,i,j}^2) \leq c_{n,j}/k_{n,j},$$

and the independence of the sequences of random variables involved implies for each $j \geq 1$ that

$$P(|\sum_{i=1}^n \xi_{n,i,j}\theta_{n,i,j}| \geq bV_{n,j}^{1/2}) \leq b^{-2}E(E((\sum_{i=1}^n \xi_{n,i,j}\theta_{n,i,j})^2/V_{n,j}|\theta_{n,1,j},\cdots,\theta_{n,n,j})) \leq b^{-2}c_{n,j}/k_{n,j}.$$

Thus (2.47), (5.3), and an application of the remark on page 49 of [19] easily combine to prove the tightness in (5.1) for general $V_{n,j}$ and also when the $V_{n,j}^{1/2}$ are replaced by $n^{1/2}$.

If $\tilde{\boldsymbol{T}}_{n,n}$ is defined as in (1.11), then for each $\epsilon > 0$ and $d \geq 1$ we have

$$P(||\tilde{\boldsymbol{T}}_{n,n}||_\infty > 2\epsilon) \leq P(||\tilde{\boldsymbol{T}}_{n,n} - Q_d(\tilde{\boldsymbol{T}}_{n,n})||_\infty > \epsilon) + P(||Q_d(\tilde{\boldsymbol{T}}_{n,n})||_\infty > \epsilon).$$

Now (5.3) immediately implies there exists $d \geq 1$ such that

$$\sup_{n\geq 1} P(||Q_d(\tilde{\boldsymbol{T}}_{n,n})||_\infty > \epsilon) < \epsilon/2,$$

and the independence of the sequences of random variables involved implies for each $j \geq 1$ and $b > 0$ that

$$\begin{aligned} P(|\sum_{i=1}^n \xi_{n,i,j}\theta_{n,i,j}| \geq bV_{n,j}) &\leq b^{-2}E(E((\sum_{i=1}^n \xi_{n,i,j}\theta_{n,i,j})^2/V_{n,j}^2|\theta_{n,1,j},\cdots,\theta_{n,n,j})) \\ &\leq b^{-2}E(V_{n,j}^{-1})c_{n,j}/k_{n,j}. \end{aligned}$$

Since $\lim_{n\to\infty} P(N_{n,1} < j) = 0$, the law of large numbers applied to the i.i.d. sequence of random variables $\{R_{n,i,j} : i \geq 1\}$ implies for each fixed $j \geq 1$ and $M > 0$ that

$$\limsup_{n\to\infty} P(V_{n,j} \leq M) = 0.$$

Thus

$$\limsup_{n\to\infty} E(V_{n,j}^{-1}) = 0,$$

so for each fixed $j \geq 1$ and $b > 0$ we have

$$\lim_{n\to\infty} P(|\sum_{i=1}^n \xi_{n,i,j}\theta_{n,i,j}| \geq bV_{n,j}) = 0.$$

Now

$$P(||\tilde{\boldsymbol{T}}_{n,n} - Q_d(\tilde{\boldsymbol{T}}_{n,n})||_\infty > \epsilon) \leq \sum_{j=1}^d P(|\sum_{i=1}^n \xi_{n,i,j}\theta_{n,i,j}| \geq \epsilon V_{n,j}),$$

and hence from the above we have for each $\epsilon > 0$ that

$$\lim_{n\to\infty} P(||\tilde{\boldsymbol{T}}_{n,n}||_\infty > 2\epsilon) \le \epsilon.$$

Thus the lemma is proven.

Now that we have tightness of $\{\mathcal{L}(\tilde{\mathbf{S}}_{n,n}) : \ n \ge 1\}$ in c_0, the next step of the proof is to show that the finite dimensional distributions induced by $\cup_{d\ge 1} c_{0,d}^*$ are the same for every limiting measure of $\{\mathcal{L}(\tilde{\mathbf{S}}_{n,n}) : \ n \ge 1\}$. Here c_0^* denotes the continuous linear functionals on c_0 and

$$c_{0,d}^* = \{f \in c_0^* : f(Q_d(\mathbf{x})) = 0 \ \text{ for all } \mathbf{x} \in c_0\}. \tag{5.4}$$

We start by showing that the limiting covariance functions $\Gamma(k,\cdot,\cdot)$ given in (2.44)-(2.46) determine the limiting variance of $f(\tilde{\mathbf{S}}_{n,n})$ for each $d \ge 1$ and $f \in c_{0,d}^*$. This follows from our next lemma.

Lemma 6. *If (2.44)-(2.46) hold and $P(\min_{1\le i\le n} N_{n,i} < d) = o(1/n^2)$ as n tends to infinity, then for all $d \ge 1$ and $f \in c_{0,d}^*$ we have*

$$\lim_{n\to\infty} E(f^2(\tilde{\mathbf{S}}_{n,n})) = \sum_{u=1}^{d}\sum_{v=1}^{d} \Gamma(1,u,v) f(\mathbf{e}_u) f(\mathbf{e}_v). \tag{5.5}$$

If $V_{n,j}^{1/2}$ is replaced by $n^{1/2}$ in $\tilde{\boldsymbol{S}}_{n,n}$, then (5.5) holds with $\Gamma(1,\cdot,\cdot)$ replaced by $\Gamma(2,\cdot,\cdot)$ in the right hand term.

Proof. Since $f \in c_{0,d}^*$, we have

$$E(f^2(\tilde{\mathbf{S}}_{n,n})) = E((\sum_{i=1}^{n}\sum_{j=1}^{d} \xi_{n,i,j}\theta_{n,i,j} f(\mathbf{e}_j)/V_{n,j}^{1/2})^2) = \sum_{i=1}^{n} E((\sum_{j=1}^{d} \xi_{n,i,j}\theta_{n,i,j} f(\mathbf{e}_j)/V_{n,j}^{1/2})^2),$$

where the last equality follows immediately since the sequences $\{\theta_{n,i,j}\}$ and $\{V_{n,j}\}$ are independent of the sequence $\{\xi_{n,i,j}\}$, and $E(\xi_{n,i,j}) = 0$ with the random variables $\xi_{n,i,j}$ independent in i. Hence again using the independence cited above we have

$$\begin{aligned} E(f^2(\tilde{\mathbf{S}}_{n,n})) &= \sum_{i=1}^{n}\sum_{u=1}^{d}\sum_{v=1}^{d} E(\xi_{n,i,u}\xi_{n,iv}) E(\frac{\theta_{n,i,u}}{V_{n,u}^{1/2}}\frac{\theta_{n,i,v}}{V_{n,v}^{1/2}}) f(\mathbf{e}_u) f(\mathbf{e}_v) \\ &= \sum_{u=1}^{d}\sum_{v=1}^{d} [\sum_{i=1}^{n} \Gamma_{n,i}(u,v) E(\frac{\theta_{n,i,u}}{V_{n,u}^{1/2}}\frac{\theta_{n,i,v}}{V_{n,v}^{1/2}})] f(\mathbf{e}_u) f(\mathbf{e}_v), \end{aligned} \tag{5.6}$$

where $\Gamma_{n,i}(u,v) = E(\xi_{n,i,u}\xi_{n,i,v})$. Hence (5.5) and Lemma 5 follows from (5.6) once we prove the following lemma. The situation when $V_{n,j}^{1/2}$ is replaced by $n^{1/2}$ in $\tilde{\boldsymbol{S}}_{n,n}$ is simpler, so for the time being we assume the $V_{n,j}$ are random. The nonrandom case will be taken up later.

Lemma 7. *Under the assumptions of Theorem 6 we have*

$$\lim_{n\to\infty} \sum_{i=1}^{n} \Gamma_{n,i}(u,v) E(\frac{\theta_{n,i,u}}{V_{n,u}^{1/2}}\frac{\theta_{n,i,v}}{V_{n,v}^{1/2}}) = \Gamma(1,u,v). \tag{5.7}$$

Proof. Since $\lim_{n\to\infty}\sum_{i=1}^{n}\Gamma_{n,i}(u,v)/n = \Gamma(u,v)$ by assumption, (5.7) will follow if we first show for $u \neq v$ that

$$E(\frac{\theta_{n,i,u}}{V_{n,u}^{1/2}}\frac{\theta_{n,i,v}}{V_{n,v}^{1/2}}) = \frac{p}{n}a_{n,i}, \tag{5.8}$$

where $\lim_{n\to\infty}\sup_{1\leq i\leq n}|a_{n,i}-1| = 0$, and that

$$\sup_{n\geq 1, i\geq 1}\Gamma_{n,i}(u,v) < \infty. \tag{5.9}$$

Now

$$\sup_{n\geq 1, i\geq 1}\Gamma_{n,i}(u,v) = \sup_{n\geq 1, i\geq 1}E(\xi_{n,i,u}\xi_{n,i,v}), \tag{5.10}$$

and, as mentioned earlier, (2.1) with $r = 2$ implies

$$E(\xi_{n,i,u}^2)) \leq c_{n,j}/k_{n,j}.$$

Therefore the Cauchy-Schwartz inequality and (2.47) easily combine to imply (5.9). Hence when $u \neq v$ it remains to prove (5.8).

To verify (5.8) for random $V_{n,j}$, let

$$\Lambda_{n,i} = \frac{\theta_{n,i,u}}{V_{n,u}^{1/2}}\frac{\theta_{n,i,v}}{V_{n,v}^{1/2}},$$

and set

$$W_{n,u} = \max\{1.\sum_{i=1}^{n}R_{n,i,u}\}$$

for $n \geq 1, u \geq 1$. Thus we have

$$\begin{aligned}E(\Lambda_{n,i}) &= E(\Lambda_{n,i}I(\min_{1\leq i\leq n}N_{n,i}\geq d)) + E(\Lambda_{n,i}I(\min_{1\leq i\leq n}N_{n,i} < d))\\ &= E(\frac{R_{n,i,u}}{W_{n,u}^{1/2}}\frac{R_{n,i,v}}{W_{n,v}^{1/2}}) - E(\frac{R_{n,i,u}}{W_{n,u}^{1/2}}\frac{R_{n,i,v}}{W_{n,v}^{1/2}}I(\min_{1\leq i\leq n}N_{n,i} < d))\\ &\qquad + E(\Lambda_{n,i}I(\min_{1\leq i\leq n}N_{n,i} < d)),\end{aligned} \tag{5.11}$$

where

$$E(\Lambda_{n,i}I(\min_{1\leq i\leq n}N_{n,i} < d)) \leq P(\min_{1\leq i\leq n}N_{n,i} < d) = o(1/n) \tag{5.12}$$

and

$$E(\frac{R_{n,i,u}}{W_{n,u}^{1/2}}\frac{R_{n,i,v}}{W_{n,v}^{1/2}}I(\min_{1\leq i\leq n}N_{n,i} < d) \leq P(\min_{1\leq i\leq n}N_{n,i} < d) = o(1/n). \tag{5.13}$$

Hence (5.8) will follow if we show for all $i,\ 1 \leq i \leq n$, and $u \neq v$ that

$$E(\frac{R_{n,i,u}}{W_{n,u}^{1/2}}\frac{R_{n,i,v}}{W_{n,v}^{1/2}}) = \frac{p}{n}a_n, \tag{5.14}$$

where $\lim_{\to\infty}a_n = 1$. To verify (5.14), we establish the following lemma, which immediately implies (5.14) when $u \neq v$. If $u = v$, then from the above we see that the analogue of (5.14) required is that $E(R_{n,i,u}/W_{n,u}) = c_n/n$ where $\lim_{n\to\infty}c_n = 1$. This follows since from the proof of Lemma 8 below we actually have $c_n = 1-(1-p)^n$. Hence the proof of Lemma 7, and also Lemma 6 for random $V_{n,j}$, will follow once Lemma 8 is established.

Lemma 8. *Let* $\{X_i : 1 \leq i \leq n\}$ *and* $\{Y_i : 1 \leq i \leq n\}$ *be independent collections of i.i.d Bernoulii random variables with* $P(X_i = 1) = P(Y_i = 1) = p$. *Let* $A_n = \max\{1, \sum_{i=1}^n X_i\}$ *and* $B_n = \max\{1, \sum_{i=1}^n Y_i\}$. *Then for all* $i,\ 1 \leq i \leq n$, *we have*

$$E(\frac{X_i}{A_n^{1/2}}\frac{Y_i}{B_n^{1/2}}) = \frac{p}{n}b_n, \tag{5.15}$$

where $\lim_{\to\infty} b_n = 1$.

Proof. By the independence assumed we have

$$E(\frac{X_i}{A_n^{1/2}}\frac{Y_i}{B_n^{1/2}}) = E(\frac{X_i}{A_n^{1/2}})E(\frac{Y_i}{B_n^{1/2}}),$$

and hence since $\{X_i : 1 \leq i \leq n\}$ and $\{Y_i : 1 \leq i \leq n\}$ i.i.d Bernoulii random variables with $P(X_i = 1) = P(Y_i = 1) = p$, it suffices to verify that

$$E(\frac{X_1}{A_n^{1/2}}) = (\frac{p}{n})^{1/2}c_n, \tag{5.16}$$

where $\lim_{\to\infty} c_n = 1$. Now

$$E(\frac{X_1}{A_n^{1/2}}) = pE((\frac{1}{1+\sum_{i=2}^n X_i})^{1/2}) \leq p(E(\frac{1}{1+\sum_{i=2}^n X_i}))^{1/2} = p(\frac{1-(1-p)^n}{np})^{1/2},$$

where the second equality follows from a formula in [27], page 198.

Hence since we assume $0 < p \leq 1$ we have

$$E(\frac{X_i}{A_n^{1/2}}) \leq (\frac{p}{n})^{1/2}. \tag{5.17}$$

Thus Lemma 8 will follow provided we establish a comparable lower bound.

Now for each $\epsilon > 0,\ 0 < \epsilon < p$, we have

$$E(\frac{X_1}{A_n^{1/2}}) = pE(\frac{1}{(1+\sum_{i=2}^n X_i)^{1/2}}) \geq p \sum_{\{k:|k/(n-1)-p|<\epsilon/Ln\}} \frac{1}{(1+k)^{1/2}} P(\sum_{i=2}^n X_i = k),$$

and hence

$$E(\frac{X_1}{A_n^{1/2}}) \geq \frac{p}{(1+(n-1)p-(n-1)\epsilon/Ln)^{1/2}} \sum_{\{k:|k/(n-1)-p|<\epsilon/Ln\}} P(\sum_{i=2}^n X_i = k).$$

Since $\epsilon > 0$, Theorem 1 of [18] implies

$$\sum_{\{k:|k/(n-1)-p|<\epsilon/Ln\}} P(\sum_{i=2}^n X_i = k) \geq 1 - 2\exp\{-2(n-1)(\frac{\epsilon}{Ln})^2\},$$

and hence

$$E(\frac{X_1}{A_n^{1/2}}) \geq \frac{p}{(1+(n-1)p-(n-1)\epsilon/Ln)^{1/2}}[1-2\exp\{-2(n-1)(\frac{\epsilon}{Ln})^2\}] = (\frac{p}{n})^{1/2}d_n,$$

where $\lim_{\to\infty} d_n = 1$. This implies the comparable lower bound to (5.17) and therefore Lemma 8 holds.

As mentioned earlier, Lemma 8 completes the proof of Lemma 7, and hence Lemma 6 is established with (5.5) providing a limiting variance function when the $V_{n,j}$ are random. If the $V_{n,j}^{1/2}$ are replaced by $n^{1/2}$ in $\tilde{\boldsymbol{S}}_{n,n}$, then the proof of Lemma 7 with the right hand of (5.7) being $\Gamma(2,u,v)$ is much simpler, and the details are left for the reader. Hence Lemma 6 is proven.

Now that Lemma 6 is verified, the next step is to show for all $d \geq 1,\ f \in c_{0,d}^*$, and random $V_{n,j}$ that all limit laws of $\{\mathcal{L}(f(\tilde{\mathbf{S}}_{n,n})) : n \geq 1\}$ are centered Gaussian random variables with variance given by

$$\sigma^2(f) = \sum_{u=1}^{d}\sum_{v=1}^{d}\Gamma(1,u,v)f(\mathbf{e}_u)f(\mathbf{e}_v). \tag{5.18}$$

Of course, if the $V_{n,j}^{1/2}$ are replaced by $n^{1/2}$ in $\tilde{\boldsymbol{S}}_{n,n}$, then (5.18) holds with $\Gamma(1,u,v)$ replaced by $\Gamma(2,u,v)$.

To verify this step of the proof, we first prove a lemma which will put us in position to allow an application of Lyapunov's central limit theorem.

Lemma 9. *For each integer $d \geq 1$ and $\mathbf{x} \in c_0$, let*

$$\Pi_d(\mathbf{x}) = \sum_{j=1}^{d} x_j \mathbf{e}_j.$$

Under the conditions of the theorem we have for each $d \geq 1$ that

$$\lim_{n\to\infty}\sum_{i=1}^{n} E(||\Pi_d(\tilde{\mathbf{X}}_{n,i})||_\infty^4) = 0, \tag{5.19}$$

where

$$\tilde{\mathbf{X}}_{n,i} = \sum_{j\geq 1}\frac{\xi_{n,i,j}\theta_{n,i,j}}{V_{n,j}^{1/2}}\mathbf{e}_j.$$

Proof. Since Jensen's inequality implies $(\frac{a+b}{2})^4 \leq \frac{a^4+b^4}{2}$ for all $a, b \geq 0$, we easily see that

$$||\Pi_d(\tilde{\mathbf{X}}_{n,i})||_\infty^4 \leq |\sum_{j=1}^{d}\frac{|\xi_{n,i,j}|\theta_{n,i,j}}{V_{n,j}^{1/2}}|^4 \leq 2^{3(d-1)}\sum_{j=1}^{d}\frac{\xi_{n,i,j}^4\theta_{n,i,j}}{V_{n,j}^2}.$$

Hence

$$E(||\Pi_d(\tilde{\mathbf{X}}_{n,i})||_\infty^4) \leq 2^{3(d-1)}\sum_{j=1}^{d}E(\frac{\xi_{n,i,j}^4\theta_{n,i,j}}{V_{n,j}^2}),$$

and the lemma will follow if we show

$$\lim_{n\to\infty}\sum_{i=1}^{n}E(\frac{\xi_{n,i,j}^4\theta_{n,i,j}}{V_{n,j}^2}) = 0 \tag{5.20}$$

for $j = 1, \cdots, d$ and all $d \geq 1$. Now

$$E(\frac{\xi_{n,i,j}^4\theta_{n,i,j}}{V_{n,j}^2}) \leq (E(\xi_{n,i,j}^8))^{1/2}(E(\frac{\theta_{n,i,j}}{V_{n,j}^4}))^{1/2},$$

and using (2.1) with $r = 2$ we have

$$E(\xi_{n,i,j}^8) = \int_0^\infty P(|\xi_{n,i,j}| > t^{1/8})dt \leq c_{n,j}\int_0^\infty \exp\{-k_{n,j}t^{1/4}\}dt = 24c_{n,j}/k_{n,j}^4.$$

Applying (2.47) and that $c_{n,j} \geq 1$ for all $n \geq 1, j \geq 1$ we therefore have

$$\sup_{n\geq 1, j\geq 1} E(\xi_{n,i,j}^8) < \infty.$$

Moreover,

$$E(\frac{\theta_{n,i,j}}{V_{n,j}^4}) = E(\frac{I(j \leq N_{n,i})R_{n,i,j}}{V_{n,j}^4}) = E(\frac{R_{n,i,j} I(\min_{1\leq i\leq n} N_{n,i} \geq d)}{\max\{1, (\sum_{i=1}^n R_{n,i,j})^4\}}) + O(P(\min_{1\leq i\leq n} N_{n,i} < d)).$$

Hence

$$E(\frac{\theta_{n,i,j}}{V_{n,j}^4}) = E(\frac{R_{n,i,j}}{\max\{1, (\sum_{i=1}^n R_{n,i,j})^4\}}) + 2O(P(\min_{1\leq i\leq n} N_{n,i} < d)) = pE(\frac{1}{(1+\sum_{i=2}^n X_i)^4}) + o(1/n^2),$$

where $X_1, X_2, \cdots, X_n$ are i.i.d. Bernoulli random variables with $P(X_i = 1) = p$ and our assumption that $P(\min_{1\leq i\leq n} N_{n,i} < d) = o(1/n^2)$.

Therefore let

$$A_n = \sum_{k=0}^n \frac{1}{(1+k)^4} P(\sum_{i=1}^n X_i = k),$$

and we want an appropriate upper bound on A_{n-1}. Now

$$A_n \leq B_n + 2\exp\{-2n/(Ln)^2\}$$

where

$$B_n = \sum_{\{k:|k/n-p|\leq 1/Ln\}} \frac{1}{(1+k)^4} P(\sum_{i=1}^n X_i = k)$$

and the exponential term follows from an immediate application of Theorem 1 in [18]. Now

$$B_n \leq \frac{1}{(1+n(p-1/Ln))^4} \sum_{\{k:|k/n-p|\leq 1/Ln\}} P(\sum_{i=1}^n X_i = k) \leq \frac{1}{(np)^4[1+\frac{1}{np} - \frac{1}{pLn}]^4} \leq \frac{2}{(np)^4}$$

for all n sufficiently large. Therefore for all $n \geq n_0$ we have

$$A_{n-1} \leq \frac{2}{((n-1)p)^4} + 2\exp\{-\frac{2(n-1)}{(L(n-1))^2}\},$$

which implies

$$E(\frac{\theta_{n,i,j}}{V_{n,j}^4}) = E(\frac{I(j \leq N_{n,i})R_{n,i,j}}{V_{n,j}^4}) \leq \frac{2}{((n-1)p)^4} + o(1/n^2)$$

uniformly in $i \geq 1, j \geq 1$. Thus uniformly in $i, j \geq 1$ we have

$$E(\frac{\xi_{n,i,j}^4 \theta_{n,i,j}}{V_{n,j}^2}) \leq (E(\xi_{n,i,j}^8))^{1/2}(E(\frac{\theta_{n,i,j}}{V_{n,j}^4}))^{1/2} = o(1/n),$$

which implies (5.20). Thus (5.19) holds by the inequality prior to (5.20), and Lemma 9 is proven for random $V_{n,j}$. If the $V_{n,j}^{1/2}$ are replaced by $n^{1/2}$ in $\tilde{\boldsymbol{S}}_{n,n}$, then the proof is even easier and details are left to the reader. Hence Lemma 9 holds for both normalizations.

The next lemma completes the proof of Theorem 6.

Lemma 10. *The functions $\Gamma(1,\cdot,\cdot)$ and $\Gamma(2,\cdot,\cdot)$ defined by (2.44)-(2.46), are covariances of centered Gaussian measures γ_1 and γ_2, respectively, on c_0. Furthermore, if the $V_{n,j}$ are random, then $\tilde{\mathbf{S}}_{n,n}$ converges weakly to γ_1 on c_0, and if the $V_{n,j}^{1/2}$ are replaced by $n^{1/2}$, then $\tilde{\mathbf{S}}_{n,n}$ converges weakly to γ_2 on c_0. In addition, for each $f \in c_0^*$ and $k = 1, 2$ we have*

$$\int_{c_0} f^2(\mathbf{x}) d\gamma_k(\mathbf{x}) = \sum_{u=1}^{\infty} \sum_{v=1}^{\infty} \Gamma(k, u, v) f(\mathbf{e}_u) f(\mathbf{e}_v).$$

Proof. First assume the $V_{n,j}$ are random. Then since (5.19) is verified, we also see for all $d \geq 1$ and $f \in c_{0,d}^*$ that

$$\lim_{n \to \infty} \sum_{i=1}^{n} E(f^4(\tilde{\mathbf{X}}_{n,i})) = 0.$$

Hence by (5.5) and Lyapunov's Central Limit Theorem, see [4], page 209, we have that $f(\tilde{\mathbf{S}}_{n,n})$ converges in distribution to a mean zero Gaussian random variable with variance given by (5.18) for all $d \geq 1$ and $f \in c_{0,d}^*$. If μ is a probability measure on the Borel subsets of c_0, and for all $k \geq 1, d \geq 1$, $f_1, \cdots, f_k \in c_{0,d}^*$, and A is an arbitrary Borel set of R^k, then the probability distributions

$$F^{f_1,\ldots,f_k}(A) = \mu(\{\mathbf{x} \in c_0 : (f_1(\mathbf{x}), \cdots, f_k(\mathbf{x})) \in A\})$$

are the finite dimensional distributions of μ on c_0 induced by $\cup_{d \geq 1} c_{0,d}^*$, and they uniquely determine μ on the Borel subsets of c_0. In view of the tightness obtained in Lemma 5 we thus have that $\tilde{\mathbf{S}}_{n,n}$ converges weakly to a unique probability on the Borel subsets of c_0, which for the moment we call μ. What remains is to show that for every $f \in c_0^*$ this limiting measure makes f a centered Gaussian random variable with variance determined by $\Gamma(1,\cdot,\cdot)$. Recalling that pointwise limits of centered Gaussian random variables are again centered Gaussian variables with limiting variances the limits of the variances, and that $\cup_{d \geq 1} c_{0,d}^*$ is weak star dense in c_0^*, it follows that μ is a centered Gaussian measure on c_0. Furthermore, if $f \in c_0^*$ and $f_d(\mathbf{x}) = f(\Pi_d(\mathbf{x}))$, $\mathbf{x} \in c_0$, then for random $V_{n,j}$ we have

$$\int_{c_0} f^2(\mathbf{x}) d\mu(\mathbf{x}) = \lim_{d \to \infty} \int_{c_0} f_d^2(\mathbf{x}) d\mu(\mathbf{x}) = \lim_{d \to \infty} \sum_{u=1}^{d} \sum_{v=1}^{d} \Gamma(1, u, v) f(\mathbf{e}_u) f(\mathbf{e}_v). \tag{5.21}$$

Since $\sup_{i \geq 1} E(\xi_{n,i,j}^2) \leq c_{n,j}/k_{n,j}$, we have from (2.47), (2.44)-(2.46), and Cauchy-Schwarz that

$$\sup_{j_1, j_2 \geq 1} |\Gamma(1, j_1, j_2)| < \infty.$$

Now $c_0^* = \ell_1$, and hence the dominated convergence theorem easily implies μ is a centered Gaussian measure on c_0 with covariance given by $\Gamma(1,\cdot,\cdot)$. Moreover, for each $f^* \in c_0$ we have

$$\int_{c_0} f^2(\mathbf{x}) d\mu(\mathbf{x}) = \sum_{u=1}^{\infty} \sum_{v=1}^{\infty} \Gamma(1, u, v) f(\mathbf{e}_u) f(\mathbf{e}_v).$$

Hence when $V_{n,j}$ is random, the centered Gaussian measure γ_1 exists as indicated, i.e. its covariance is $\Gamma(1,\cdot,\cdot)$, and $\mu = \gamma_1$. Similarly, when the $V_{n,j}^{1/2}$ are replaced by $n^{1/2}$, then γ_2 exists as indicated, and $\mu = \gamma_2$. This last fact is easy to check by immediate simplifications of what we have done when $V_{n,j}^{1/2}$ is random, and the details are left to the reader. Hence for each choice of normalizers, there is a unique limiting Gaussian measure, and its finite dimensional distributions are centered Gaussian measures determined by the appropriate covariance function. Therefore the lemma is proved, and Theorem 6 is established.

5.2 Proof of Theorem 7

The proof parallels that for Theorem 6. We present the details for random $V_{n,j}$, and when the $V_{n,j}^{1/2}$ are replaced by $n^{1/2}$ in $\tilde{\boldsymbol{S}}_{n,n}$ or $\tilde{\boldsymbol{T}}_{n,n}$ we leave most of the details to the reader.

Our first task is to show that we have tightness in ℓ_ρ, and that $\tilde{\boldsymbol{T}}_{n,n}$ converges to zero in probability there. This is given in our next lemma.

Lemma 11. *Under the conditions (2.51) and (2.52) of Theorem 7 we have*

$$\{\mathcal{L}(\tilde{\mathbf{S}}_{n,n}) : \ n \geq 1\} \ \text{ is tight in } \ell_\rho. \tag{5.22}$$

In addition, if for each $j \geq 1$ we have $\lim_{n\to\infty} P(N_{n,1} < j) = 0$, then $\tilde{\boldsymbol{T}}_{n,n}$ converges in probability to zero in ℓ_ρ.

Proof. We first establish the uniform tightness for random $V_{n,j}$. The tightness when the $V_{n,j}^{1/2}$ are replaced by $n^{1/2}$ then follows by an immediate simplification of this argument.

If $\theta_{n,i,j} = I(j \leq N_{n,i})R_{n,i,j}$ as indicated, then $P(\theta_{n,i,j} = 1) = p_{n,j}p$, where $p_{n,j} = P(j \leq N_{n,i})$ for $n \geq 1, j \geq 1$, and for $k = 0, 1, \cdots, n$ we define the events

$$E_{k,n,j} = \cup_{I \in \mathcal{I}_{k,n,j}} F_I, \tag{5.23}$$

where $\mathcal{I}_{k,n,j}$ denotes all subsets $I = \{i_1, \cdots, i_k\}$ of size k in $\{1, \cdots, n\}$ and

$$F_I = \{\theta_{n,i,j} = 1 \text{ for all } i \in I \text{ and } \theta_{n,i,j} = 0 \text{ for } i \in \{1, \cdots, n\} \cap I^c\}.$$

Note that F_I depends on n and j, but we suppress that in our notation.

Let $Q_d(\mathbf{x}) = \sum_{j \geq d+1} x_j \mathbf{e}_j$ for $\mathbf{x} \in R^\infty$, and $d \geq 0$, and recall $V_{n,j} = \max\{1, \sum_{i=1}^n \theta_{n,i,j}\}$. Therefore, since $\sum_{i=1}^n \xi_{n,i,j}\theta_{n,i,j} = 0$ on $E_{0,n,j}$, we have for each $\delta > 0$, $n \geq 1$, and $d \geq 0$ that

$$P(||Q_d(\tilde{\mathbf{S}}_{n,n})||_\rho \geq \delta) \leq \delta^{-\rho} \sum_{j \geq d+1} \sum_{k=1}^n E((\frac{|\sum_{i=1}^n \xi_{n,i,j}\theta_{n,i,j}|}{V_{n,j}^{1/2}})^\rho I_{E_{k,n,j}}).$$

Now

$$E((\frac{|\sum_{i=1}^n \xi_{n,i,j}\theta_{n,i,j}|}{V_{n,j}^{1/2}})^\rho I_{E_{k,n,j}}) = \sum_{I \in \mathcal{I}_{k,n,j}} E((\frac{|\sum_{l=1}^k \xi_{n,i_l,j}|}{k^{1/2}})^\rho I_{F_I}), \ I = \{i_1, \cdots, i_k\}),$$

and letting

$$A_{n,d} = \sum_{j \geq d+1} \sum_{k=1}^n E((\frac{|\sum_{i=1}^n \xi_{n,i,j}\theta_{n,i,j}|}{V_{n,j}^{1/2}})^\rho I_{E_{k,n,j}})$$

we have by using the independence of the various sequences of random variables involved that

$$A_{n,d} = \sum_{j \geq d+1} \sum_{k=1}^n \sum_{I \in \mathcal{I}_{k.n,j}} E((\frac{|\sum_{l=1}^k \xi_{n,i_l,j}|}{k^{1/2}})^\rho) P(F_I, \ I = \{i_1, \ldots, i_k\}).$$

Since we are assuming (2.1) with r=2, for each integer $b \geq 1$ we have from Lemma 1 that

$$E((\frac{|\sum_{l=1}^b \xi_{n,i_l,j}|}{b^{1/2}})^\rho) = \int_0^\infty P((\frac{|\sum_{l=1}^b \xi_{n,i_l,j}|}{b^{1/2}}) > x^{1/\rho}) dx \leq 2 \int_0^\infty \exp\{-\frac{k_{n,j}}{16 c_{n,j}} x^{2/\rho}\} dx.$$

Now

$$\int_0^\infty \exp\{-\frac{k_{n,j}}{16c_{n,j}}x^{2/\rho}\}dx = (\frac{16c_{n,j}}{k_{n,j}})^{\rho/2}\eta(\rho),$$

where

$$\eta(\rho) = \frac{\rho}{2}\int_0^\infty u^{\frac{\rho}{2}-1}\exp\{-u\}du.$$

Since $2 \leq \rho < \infty,\ \eta(\rho) < \infty$, and for each integer $b \geq 1$ we have

$$E((\frac{|\sum_{l=1}^b \xi_{n,i_l,j}|}{b^{1/2}})^\rho) \leq 2(\frac{16c_{n,j}}{k_{n,j}})^{\rho/2}\eta(\rho).$$

Hence

$$A_{n,d} \leq 2\eta(\rho)\sum_{j\geq d+1}(\frac{16c_{n,j}}{k_{n,j}})^{\rho/2},$$

which implies

$$P(||Q_d(\tilde{\mathbf{S}}_{n,n})||_\infty \geq \delta) \leq 2\delta^{-\rho}\eta(\rho)\sum_{j\geq d+1}(\frac{16c_{n,j}}{k_{n,j}})^{\rho/2}.$$

Applying (2.52) we therefore have for all $\delta > 0$ that

$$\lim_{d\to\infty}\sup_{n\geq 1} P(||Q_d(\tilde{\mathbf{S}}_{n,n})||_\infty > \delta) = 0. \tag{5.24}$$

Now (2.1) with r=2 easily implies

$$E(\xi_{n,i,j}^2) \leq c_{n,j}/k_{n,j},$$

and, arguing as in the proof of Lemma 5, we then see that (2.51), (5.24) with $\delta > 0$ arbitrary, and an application of the remark on page 49 of [19] proves the tightness in (5.22) for general $V_{n,j}$.

Since the random variables $\{\xi_{n,i,j}\theta_{n,i,j} : n \geq 1, i \geq 1, j \geq 1\}$ also satisfy (2.1) with r=2, when the $V_{n,j}^{1/2}$ are replaced by $n^{1/2}$ the proof of (5.22) is immediate from a simplification of the previous argument.

To prove $\tilde{\boldsymbol{T}}_{n,n}$ converges in probability to zero in ℓ_ρ follows from what we just proved adapted to the argument for the analogue of this in Lemma 5. Thus the lemma is proven.

Now that we have tightness of $\{\mathcal{L}(\tilde{\mathbf{S}}_{n,n}) :\ n \geq 1\}$ in ℓ_ρ, the next step of the proof is to show that the finite dimensional distributions induced by $\cup_{d\geq 1}\ell_{\rho,d}^*$ are the same for every limiting measure of $\{\mathcal{L}(\tilde{\mathbf{S}}_{n,n}) :\ n \geq 1\}$. Here ℓ_ρ^* denotes the continuous linear functionals on $\ell\rho$ and

$$\ell_{\rho,d}^* = \{f \in \ell_\rho^* : f(Q_d(\mathbf{x})) = 0 \ \text{ for all } \mathbf{x} \in \ell_\rho\}. \tag{5.25}$$

We start by showing that the limiting covariance functions $\Gamma(k,\cdot,\cdot)$ given in (2.44)-(2.46) determine the limiting variance of $f(\tilde{\mathbf{S}}_{n,n})$ for each $d \geq 1$ and $f \in \ell_{\rho,d}^*$. This follows from our next lemma.

Lemma 12. *If (2.44)-(2.46) hold and $P(\min_{1\leq i\leq n} N_{n,i} < d) = o(1/n^2)$ as n tends to infinity, then for all $d \geq 1$ and $f \in \ell_{\rho,d}^*$ we have*

$$\lim_{n\to\infty} E(f^2(\tilde{\mathbf{S}}_{n,n})) = \sum_{u=1}^d\sum_{v=1}^d \Gamma(1,u,v)f(\mathbf{e}_u)f(\mathbf{e}_v). \tag{5.26}$$

If the $V_{n,j}^{1/2}$ are replaced by $n^{1/2}$ in $\tilde{\boldsymbol{S}}_{n,n}$, then (5.26) holds with $\Gamma(1,u,v)$ replaced by $\Gamma(2,u,v)$ in the right hand term.

The proof of this lemma is exactly as that of Lemma 6 in the proof of Theorem 6. Hence we immediately turn to our next task, which is to show for all $d \geq 1$ and $f \in \ell^*_{\rho,d}$ that all limit laws of $\{\mathcal{L}(f(\tilde{\mathbf{S}}_{n,n})); n \geq 1\}$ are centered Gaussian random variables with variance given by

$$\sigma^2(f) = \sum_{u=1}^{d}\sum_{v=1}^{d}\Gamma(1,u,v)f(\mathbf{e}_u)f(\mathbf{e}_v). \tag{5.27}$$

Of course, when the $V_{n,j}^{1/2}$ are replaced by $n^{1/2}$ in $\tilde{\boldsymbol{S}}_{n,n}$, then (5.27) holds with $\Gamma(1,u,v)$ replaced by $\Gamma(2,u,v)$.

To verify this step of the proof, we first prove a lemma which will put us in position to allow an application of Lyapunov's central limit theorem.

Lemma 13. *For each integer $d \geq 1$ and $\mathbf{x} \in \ell_\rho$, let*

$$\Pi_d(\mathbf{x}) = \sum_{j=1}^{d} x_j \mathbf{e}_j.$$

Under the conditions of the theorem we have for each $d \geq 1$ that

$$\lim_{n\to\infty}\sum_{i=1}^{n} E(||\Pi_d(\tilde{\mathbf{X}}_{n,i})||_\rho^4) = 0, \tag{5.28}$$

where

$$\tilde{\mathbf{X}}_{n,i} = \sum_{j\geq 1}\frac{\xi_{n,i,j}\theta_{n,i,j}}{V_{n,j}^{1/2}}\mathbf{e}_j.$$

The proof of this lemma is an immediate consequence of Lemma 9 since $||\Pi_d(\mathbf{x})||_\rho \leq d||\Pi_d(\mathbf{x})||_\infty$ for all $\mathbf{x} \in \ell_\rho$. Hence the proof of Theorem 7 will be complete once we have the following lemma.

Lemma 14. *The functions $\Gamma(k,\cdot,\cdot)$ defined for $k = 1,2$ by (2.44)-(2.46) are covariances of centered Gaussian measures γ_1 and γ_2, respectively, on ℓ_ρ. Furthermore, if the $V_{n,j}$ are random, then $\tilde{\mathbf{S}}_{n,n}$ converges weakly to γ_1 on ℓ_ρ, and if the $V_{n,j}^{1/2}$ are replaced by $n^{1/2}$, then $\tilde{\mathbf{S}}_{n,n}$ converges weakly to γ_2 on $\ell\rho$. In addition, for each $f \in \ell_\rho^*$ and $k = 1,2$ we have*

$$\int_{\ell_\rho} f^2(\mathbf{x})d\gamma_k(\mathbf{x}) = \sum_{u=1}^{\infty}\sum_{v=1}^{\infty}\Gamma(k,u,v)f(\mathbf{e}_u)f(\mathbf{e}_v).$$

Proof. First we consider random $V_{n,j}$. Applying Lemmas 12 and 13, and Lyapunov's central limit theorem, the proof of Lemma 14 follows similarly as in the proof of Lemma 10. That is, arguing as in Lemma 10 we have $\mathcal{L}(\tilde{\mathbf{S}}_{n,n})$ converging weakly to a mean zero Gaussian measure μ on ℓ_ρ. Furthermore, if $f \in \ell^*_{\rho,d}$, then

$$\int_{\ell_\rho} f^2(\mathbf{x})d\mu(\mathbf{x}) = \sum_{u=1}^{d}\sum_{v=1}^{d}\Gamma(1,u,v)f(\mathbf{e}_u)f(\mathbf{e}_v), \tag{5.29}$$

and for $f \in \ell^*_\rho$ we have

$$\int_{\ell_\rho} f^2(\mathbf{x})d\mu(\mathbf{x}) = \lim_{d\to\infty}\sum_{u=1}^{d}\sum_{v=1}^{d}\Gamma(1,u,v)f(\mathbf{e}_u)f(\mathbf{e}_v). \tag{5.30}$$

To show the previous limit equals

$$\sum_{u=1}^{\infty}\sum_{v=1}^{\infty}\Gamma(1,u,v)f(\mathbf{e}_u)f(\mathbf{e}_v),$$

and that $\gamma_1=\mu$ as indicated on ℓ_ρ, we let $f(\mathbf{x})=x_u+x_v$ and $g(\mathbf{x})=x_u-x_v$ for $\mathbf{x}=\sum_{j\geq 1}x_j\mathbf{e}_j$. Then $f,g\in\ell^*_{\rho,d}$ for $d\geq\max\{u,v\}$, and applying (5.29) we have

$$\begin{aligned}4\int_{\ell_\rho}x_ux_vd\mu(\mathbf{x}) &= \int_{\ell_\rho}(x_u+x_v)^2d\mu(\mathbf{x})-\int_{\ell_\rho}(x_u-x_v)^2d\mu(\mathbf{x})\\ &= \int_{\ell_\rho}f^2(\mathbf{x})d\mu(\mathbf{x})-\int_{\ell_\rho}g^2(\mathbf{x})d\mu(\mathbf{x})\\ &= 4\Gamma(1,u,v).\end{aligned}$$

Therefore the Gaussian measure γ_1 exists on ℓ_p as indicated, and it equals μ. In addition we will see

$$\int_{\ell_\rho}f^2(\mathbf{x})d\mu(\mathbf{x})=\sum_{u=1}^{\infty}\sum_{v=1}^{\infty}\Gamma(1,u,v)f(\mathbf{e}_u)f(\mathbf{e}_v) \tag{5.31}$$

for each $f\in\ell_\rho$. To verify (5.31) we recall that since γ_1 is a centered Gaussian measure on ℓ_ρ, it is known that

$$\sum_{u\geq 1}(\Gamma(1,u,u))^{\rho/2}<\infty.$$

Now $\Gamma(1,u,v)\leq\Gamma(1,u,u)^{1/2}\Gamma(1,v,v)^{1/2}$ and hence we have

$$\begin{aligned}\sum_{u=1}^{\infty}\sum_{v=1}^{\infty}|\Gamma(1,u,v)f(\mathbf{e}_u)f(\mathbf{e}_v)| &\leq (\sum_{u=1}^{\infty}\Gamma(1,u,u)^{1/2}|f(\mathbf{e}_u)|)^2\\ &\leq (\sum_{u=1}^{\infty}\Gamma(1,u,u)^{\rho/2})^{1/\rho}(\sum_{u=1}^{\infty}|f(\mathbf{e}_u)|^q)^{1/q})^2,\end{aligned}$$

where $1/q+1/\rho=1$. Now $f\in\ell^*_\rho=\ell_q$, so (5.30) and the dominated convergence theorem now implies (5.31) for all $f\in\ell^*_\rho$.

Hence the limiting mean zero Gaussian measure γ_1 exists on ℓ_ρ, and has covariance $\Gamma(1,\cdot,\cdot)$ with (5.31) holding. If the $V_{n,j}^{1/2}$ are replaced by $n^{1/2}$, the argument is similar. Thus Theorem 7 is proven.

5.3 Verifying Remark 11

The proof of Remark 11 parallels the proof of Theorem 7. Hence we only provide an outline.

As in the proof of Theorem 7, our first task is to verify tightness of $\{\mathcal{L}(\frac{\boldsymbol{S}_{n,n}}{n^{1/2}}):\ n\geq 1\}$ in ℓ_ρ. This follows since for each $d\geq 0$ we have

$$P(||Q_d(\boldsymbol{S}_{n,n})||_\rho\geq\delta)\leq A_\rho\delta^{-\rho}\sum_{i=1}^{n}E(||Q_d(X_{n,i})||_\rho^2)=A_\rho\delta^{-\rho}\sum_{i=1}^{n}E((\sum_{j\geq d+1}|\xi_{n,i,j}\theta_{n,i,j}|^\rho)^{2/\rho}),$$

where the inequality holds since ℓ_ρ is a type-2 Banach space for $2\leq\rho<\infty$, see [2], page 157. Now

$$E((\sum_{j\geq d+1}|\xi_{n,i,j}\theta_{n,i,j}|^\rho)^{2/\rho})\leq(\sum_{j\geq d+1}E(|\xi_{n,i,j}\theta_{n,i,j}|^\rho))^{2/\rho}\leq(\sum_{j\geq d+1}E(|\xi_{n,i,j}|^\rho))^{2/\rho},$$

where the first inequality is due to Jensen's inequality, and the second because $|\theta_{n,i,j}| \leq 1$. Finally, since $2/\rho \leq 1$ we have

$$(\sum_{j \geq d+1} E(|\xi_{n,i,j}|^\rho))^{2/\rho} \leq \sum_{j \geq d+1} (E(|\xi_{n,i,j}|^\rho)^{2/\rho},$$

which combined with the previous inequalities implies

$$P(||Q_d(\boldsymbol{S}_{n,n})||_\rho \geq \delta n^{1/2}) \leq A_\rho \delta^{-\rho} n^{-1} \sum_{i=1}^{n} \sum_{j \geq d+1} (E(|\xi_{n,i,j}|^\rho)^{2/\rho}.$$

Hence, as in the final step of Lemma 11, we see that (2.55) and (2.56) combine to imply the tightness of $\{\mathcal{L}(\frac{\boldsymbol{S}_{n,n}}{n^{1/2}}) : \ n \geq 1\}$ in ℓ_ρ.

Now that we have tightness of $\{\mathcal{L}(\frac{\boldsymbol{S}_{n,n}}{n^{1/2}}) : \ n \geq 1\}$ in ℓ_ρ, the next step of the proof is to show that the finite dimensional distributions induced by $\cup_{d \geq 1} \ell^*_{\rho,d}$ are the same for every limiting measure of $\{\mathcal{L}(\frac{\boldsymbol{S}_{n,n}}{n^{1/2}}) : \ n \geq 1\}$ We start by showing that the limiting covariance function $\Gamma(2, \cdot, \cdot)$ determined as in (2.44)-(2.46) gives the limiting variance of $f(\frac{\mathbf{S}_{n,n}}{n^{1/2}})$ for each $d \geq 1$ and $f \in \ell^*_{\rho,d}$. This follows from our next lemma.

Lemma 15. *If (2.44)-(2.46) hold and* $P(\min_{1 \leq i \leq n} N_{n,i} < d) = o(1)$ *as* n *tends to infinity, then for all* $d \geq 1$ *and* $f \in \ell^*_{\rho,d}$ *we have*

$$\lim_{n \to \infty} E(f^2(\boldsymbol{S}_{n,n}/n^{1/2})) = \sum_{u=1}^{d} \sum_{v=1}^{d} \Gamma(2, u, v) f(\mathbf{e}_u) f(\mathbf{e}_v). \tag{5.32}$$

The proof of this lemma is a simplification of the proof of Lemma 6. Hence we immediately turn to our next task, which is to show for all $d \geq 1$ and $f \in \ell^*_{\rho,d}$ that all limit laws of $\{\mathcal{L}(f(\boldsymbol{S}_{n,n})/n^{1/2}) : n \geq 1\}$ are centered Gaussian random variables with variance given by

$$\sigma^2(f) = \sum_{u=1}^{d} \sum_{v=1}^{d} \Gamma(2, u, v) f(\mathbf{e}_u) f(\mathbf{e}_v). \tag{5.33}$$

To verify this step of the proof, we first indicate a lemma which will put us in position to allow an application of Lyapunov's central limit theorem. Its proof is sketched below.

Lemma 16. *For each integer* $d \geq 1$ *and* $\mathbf{x} \in \ell_\rho$, *let*

$$\Pi_d(\mathbf{x}) = \sum_{j=1}^{d} x_j \mathbf{e}_j.$$

Under the conditions of the theorem we have for each $d \geq 1$ *and some* $\beta > 0$ *that*

$$\lim_{n \to \infty} \sum_{i=1}^{n} E(||\Pi_d(\mathbf{X}_{n,i}/n^{1/2})||_\rho^{2+\beta}) = 0. \tag{5.34}$$

If $2 < \rho < \infty$ and we take $\beta > 0$ sufficiently small that $2 + \beta < \rho$ and $\beta < \delta$ where $\delta > 0$ is as in (2.55), then we have

$$||\Pi_d(\mathbf{X}_{n,i})/n^{1/2}||_\rho^{2+\beta} = n^{-(2+\beta)/2} (\sum_{j=1}^{d} |\xi_{n,i,j}|^\rho)^{\frac{2+\beta}{\rho}} \leq n^{-(2+\beta)/2} \sum_{j=1}^{d} |\xi_{n,i,j}|^{2+\beta}, \tag{5.35}$$

where the inequality holds since $2+\beta \le \rho$. Hence (2.55) implies (5.34) holds, and the lemma is proven when $2<\rho<\infty$. If $\rho=2$ and we take $\delta>0$ as in (2.55), then

$$||\Pi_d(\mathbf{X}_{n,i})/n^{1/2}||_2^{2+\delta} = n^{-(2+\delta)/2}(\sum_{j=1}^{d}|\xi_{n,i,j}|^2)^{\frac{2+\delta}{2}} \le n^{-(2+\delta)/2}2^{(d-1)\delta/2}\sum_{j=1}^{d}|\xi_{n,i,j}|^{2+\delta}, \tag{5.36}$$

where the inequality follows since $(a+b)^c \le 2^{c-1}(a^c+b^c)$ for $a,b\ge 0, c\ge 1$, and we iterate this $d-1$ times. Hence (2.55) implies (5.34) holds, and the lemma is proven.

The proof of Remark 11 now follows once we establish the following lemma.

Lemma 17. *The function $\Gamma(2,\cdot,\cdot)$ defined by (2.44)-(2.46) is the covariance of a centered Gaussian measure γ_2 on ℓ_ρ, and $\frac{\mathbf{S}_{n,n}}{n^{1/2}}$ converges weakly to γ_2 on ℓ_ρ. In addition,*

$$\int_{\ell_\rho} f^2(\mathbf{x})d\gamma_2(\mathbf{x}) = \sum_{u=1}^{\infty}\sum_{v=1}^{\infty}\Gamma(2,u,v)f(\mathbf{e}_u)f(\mathbf{e}_v)$$

for all $f\in\ell_\rho^*$.

Applying Lemmas 15 and 16, and Lyapunov's central limit theorem, the proof of Lemma 17 follows exactly as in the proof of Lemmas 10 and 14 above. Hence the remark is established.

6 Applications

6.1 One-Sample Problem

In this section we apply the results from previous sections to test that the "mean vector" equals a specified vector. More precisely, consider testing the null hypothesis $\mathbf{H}_0 : \boldsymbol{\mu}_n = \mathbf{0}$, where $\boldsymbol{\mu}_n$ is an infinite dimensional vector whose components are $\mu_{n,j}$. The quantity $\tilde{\boldsymbol{S}}_{n,n}$, defined by (1.7), namely

$$\tilde{\boldsymbol{S}}_{n,n} = \sum_{i=1}^{n}\sum_{j\ge 1}\frac{\xi_{n,i,j}\theta_{n,i,j}}{V_{n,j}^{1/2}}\mathbf{e}_j \equiv \sum_{i=1}^{n}\sum_{j=1}^{N_{n,i}}\frac{\xi_{n,i,j}R_{n,i,j}}{V_{n,j}^{1/2}}\mathbf{e}_j, \tag{6.1}$$

can be used for developing a test of $\mathbf{H}_0$. To this end, let us denote the data vectors by $\vec{\boldsymbol{X}}_n = \{\boldsymbol{X}_{n,1},\cdots\boldsymbol{X}_{n,n}\}$. One can use the l_ρ norm for $\rho\ge 2$ and the c_0 norm to define various non-randomized test functions $\phi_\rho(\vec{\boldsymbol{X}}_n)$ as follows:

$$\phi_\rho(\vec{\boldsymbol{X}}_n) = \begin{cases} 1 & if\ \ ||\tilde{\boldsymbol{S}}_{n,n}||_\rho > c \\ 0 & otherwise, \end{cases} \tag{6.2}$$

where $c=c_\rho$ is so chosen that $E(\phi_\rho(\vec{\boldsymbol{X}}_n)|\mathbf{H}_0)\le\alpha$. The test statistics based on c_0 norm is given by

$$\phi_\infty(\vec{\boldsymbol{X}}_n) = \begin{cases} 1 & if\ \ ||\tilde{\boldsymbol{S}}_{n,n}||_\infty > c \\ 0 & otherwise, \end{cases} \tag{6.3}$$

where $c=c_\infty$ is so chosen that $E(\phi_\rho(\vec{\boldsymbol{X}}_n)|\mathbf{H}_0)\le\alpha$. Thus to perform the test one needs the distribution of $||\tilde{\boldsymbol{S}}_{n,n}||_\rho$ and $||\tilde{\boldsymbol{S}}_{n,n}||_\infty$. Their asymptotic distribution can be obtained as a corollary of Theorems 6, 7, and Remark 11 as in the following proposition.

Proposition 1. *If the null hypothesis holds, then Theorem 6 implies that* $P(||\tilde{\boldsymbol{S}}_{n,n}||_\infty > c)$ *converges to* $P(||\mathbf{G}||_\infty > c)$ *for all* $c > 0$*, where* $\mathcal{L}(\mathbf{G}) = \gamma$ *and* γ *is the Gaussian measure identified there. Furthermore, under the conditions of Theorem 7, when the null hypothesis is true, then the* $P(||\tilde{\boldsymbol{S}}_{n,n}||_\rho > c)$ *converges to* $P(||\mathbf{G}||_\rho > c)$ *for all* $c > 0$ *and* $2 \leq \rho < \infty$*, where* $\mathcal{L}(\mathbf{G}) = \gamma$ *and* γ *is the Gaussian measure identified in Theorem 7. A similar result holds for* $\mathbf{S}_{n,n}/n^{1/2}$ *under the conditions of Remark 11, when the null hypothesis is true.*

We omit the proof of the proposition as it is an immediate consequence of our central limit theorems, the continuous mapping theorem, and the fact that the norm of a Gaussian random vector with values in a separable Banach space has a continuous distribution function.

When $\rho = 2$, our test function is similar to the test function derived using Hotelling's T^2 statistic. The main difference, however, is that in our case the vectors are not finite dimensional, and we typically do not have uncorrelated coordinates variables $\{\xi_{n,i,j}\}$ with unit variances. In fact, in most infinite dimensional examples where a central limit theorem is to be expected, having non-zero constant variances for the coordinate variables is impossible. A hybrid result when $P(N_{n,i} = b(n)) = 1$, $p = 1$, $E(\xi_{n,i,j}^2) = 1$ for $n, i, j \geq 1$, and $E(\xi_{n,i,u}\xi_{n,i,v}) = 0$ for $u \neq v$ is given in Portnoy [26] under a few additional moment and regularity conditions. Under these assumptions $\tilde{\boldsymbol{S}}_{n,n} = \mathbf{S}_{n,n}/n^{1/2}$, and when $b(n)/n$ tends to zero as n tends to infinity, Portnoy shows

$$W_n \equiv \frac{||\tilde{\boldsymbol{S}}_{n,n}||_2^2 - b(n)}{\sqrt{2b(n)}} \Rightarrow G, \tag{6.4}$$

where G is a normal random variable with mean 0 and variance 1. To contrast the above result to those of Proposition 1, note that the key assumption concerning the summability of variances is violated. That is, under the assumptions in [26] the sum of the variances diverges if $b(n) \to \infty$, while our Theorem 7 and the results in Remark 11 require the sum of variances converges when $\rho = 2$. Hence a natural question to ask is whether one has an analogue of Portnoy's result under the conditions above when the identity covariance condition is replaced by simply asking the coordinates of $\tilde{\boldsymbol{S}}_{n,n}$ have variances that sum to infinity. The following example, with highly dependent coordinates, shows that the limit distribution in this case could be non-Gaussian. Of course, many other examples can be obtained in a similar way, and all fail to satisfy an analogue of Portnoy's result as they have highly correlated coordinates.

Example 1. *Let* $\{\eta_i : i \geq 1\}$ *be a collection of i.i.d. random variables with* $E(\eta_i) = 0$ *and* $E(\eta_i^2) = 1$*. Let* $\{a_j : j \geq 1\}$ *be a sequence of constants converging to* 0 *such that such that* $\sum_{j \geq 1} a_j^2 = \infty$*. Set* $\xi_{n,i,j} = \eta_i a_j$*. Then,*

$$\frac{\mathbf{S}_{n,n}}{\sqrt{n}} = \sum_{i=1}^{n} \frac{\eta_i}{\sqrt{n}} \sum_{j \geq 1} a_j \mathbf{e}_j \tag{6.5}$$

$$\Rightarrow G \sum_{j \geq 1} a_j \mathbf{e}_j, \tag{6.6}$$

where G *is a normal random variable with mean* 0 *and variance* 1*. Furthermore, the limit does not live in the space* l_2*, and*

$$W_n \equiv \sum_{j=1}^{b(n)} \{(\frac{\sum_{i=1}^n \eta_i}{\sqrt{n}})^2 a_j^2 - a_j^2\}. \tag{6.7}$$

Then, by the central limit theorem, it follows that

$$\frac{W_n}{\sum_{j=1}^{b(n)} a_j^2} \Rightarrow G^2 - 1, \tag{6.8}$$

and no constant normalizations lead to a limiting normal law.

The above example is a situation where the gene expressions would be "highly correlated," and serves to show that the identity covariance matrix assumption used in [26] represents a special situation. As far as we know it is an open problem to see what happens in the middle ground between Portnoy's assumptions and the highly correlated situation in Example 1, but that is something we leave unsettled here. Of course, both of these extremes are quite restrictive, and tend not to hold in practice, but perhaps more important is that simulation results provided in the section below (see the paragraph before (6.19)) show that the Type I error rates of tests developed using (6.4), which assumes identity covariance matrix, behave poorly compared to the nominal level when correlations are actually present.

6.2 Simulation Results

In this section we evaluate our methodology, using simulations, when the number of replications is small, but the number of variables is large. The number of replications n is fixed at 10 throughout section 6.2, but various choices of the dimension of the random vectors $b(n)$ are considered, and all our simulation results are based on 5000 independent trials of 10 replications of the various experiments being discussed. We purposely chose n small to reflect many real applications.

As a first step we need to "approximate" the limiting distribution of the random variables appearing in Proposition 1. We will work with the case $N_{n,i} = b(n)$ and assume that $\boldsymbol{X}_{n,1} \cdots \boldsymbol{X}_{n,n}$ are n i.i.d. $b(n)$ dimensional vectors with distribution $G_n(.)$ whose tails satisfy the sub-Gaussian property. Let $\hat{\Sigma}_n$ denote an estimate of the covariance matrix $\Sigma_n = ((\sigma_{n,u,v}))$, where Σ_n is a $b(n) \times b(n)$ matrix given by

$$\sigma_{n,u,v} = E(\xi_{n,1,u} - \mu_{n,u,0})(\xi_{n,1,v} - \mu_{n,v,0}). \tag{6.9}$$

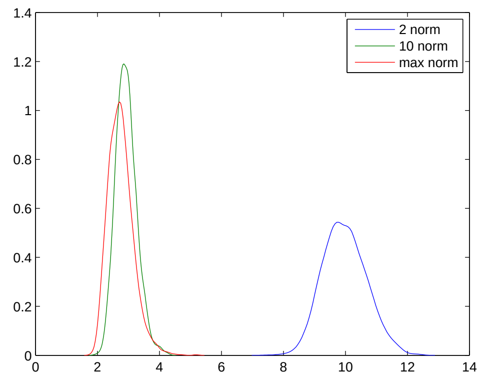
In the above definition, $\mu_{n,u,0}$ and $\mu_{n,v,0}$ are the specified values under the null hypothesis. Note that $\hat{\Sigma}_n$ is a function of the data vector $\vec{\boldsymbol{X}}_{n,n}$. One choice for $\hat{\Sigma}_n$ is the sample covariance matrix. In fact, better options are available, and we will explain them later below. If $\hat{\Sigma}_n$ is positive definite, then given $\vec{\boldsymbol{X}}_{n,n}$, we generate t i.i.d. random vectors $\boldsymbol{Y}_{n,i}$ of dimension $b(n)$ whose distribution is Gaussian with mean vector $\mathbf{0}$ and covariance matrix $\hat{\Sigma}_n$; that is

$$\boldsymbol{Y}_{n,i} | \vec{\boldsymbol{X}}_n \sim N_{b(n)}(\mathbf{0}, \hat{\Sigma}_n) \quad \text{a.s.}, \qquad 1 \leq i \leq t. \tag{6.10}$$

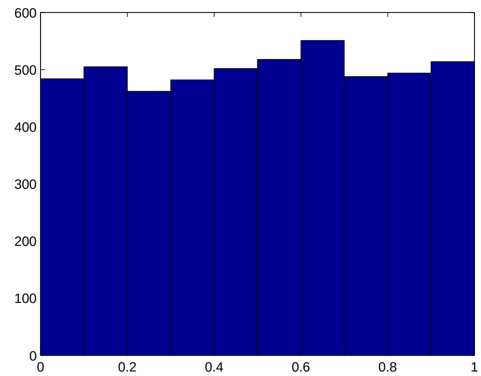
We will call $\boldsymbol{Y}_{n,i}$ the monte-carlo (MC) samples, and throughout the simulations $t = 2000$. We will use $||\cdot||$ to denote the l_ρ norm ($\rho \geq 2$) or the c_0 norm depending on the space being used. Let $||\boldsymbol{Y}_{n,1}||, \cdots, ||\boldsymbol{Y}_{n,n}||$ denote the norms of the MC samples. Furthermore, consider the following non-parametric density estimator; namely, for $x \in R$,

$$h_t(x) = \frac{1}{tc_t} \sum_{i=1}^{t} K(\frac{x - ||\boldsymbol{Y}_{n,i}||}{c_t}), \tag{6.11}$$

where c_t is a sequence of positive constants converging to 0 such that $tc_t \to \infty$, and $K(.)$ is a density function with $\int_R tK(t)dt = 0$. In the above we have suppressed the dependence on n and on ω since n and ω will be held fixed in this discussion. It follows from Devroye [5] that as $t \to \infty$, that for every $\omega \in \Omega$, $h_t(x)$ converges almost everywhere with respect to Lebesgue measure and in L_1 to the probability density of the random variable $||N(\mathbf{0}, \hat{\Sigma}_n)||$. In all our numerical experiments we will take $K(.)$ to be a standard normal density, $t = 2000$, and fix the window width c_t at 0.7. Figure 1 (a) presents the graph of the density function for the 2-norm, the 10-norm, and the sup-norm.



(a)



(b)

Figure 1: Kernel density estimates of the norms of statistics and the Histogram of p-values under the sup norm.

We use these densities to "approximate" the tail probabilities of the norms of the limiting Gaussian appearing in Proposition 1. Figure 1 (b) shows that the p-values from these hypothesis tests are uniformly distributed when we use the sup-norm. This histogram was generated using compound symmetric covariance structure, the details of which are explained later in this section. We now present various numerical experiments that describe various properties of our methodology. The nominal Type I error rate is taken to be 5% in all of our experiments.

Our first experiment studies the behavior of the Type I error rate for the testing problem H_0 : $\boldsymbol{\mu}_n = \boldsymbol{\mu}_{n,0}$, where $\boldsymbol{\mu}_{n,0} = \{\mu_{n,j,0} : j \geq 1\}$. As indicated, the sample size, or number of replications, is taken to be 10. Here the dimension of the sample vectors is taken to be 100, and the data are generated from a 100 dimensional normal distribution with mean $\boldsymbol{\mu}_n$ and covariance matrix Σ_n. The entries of the mean vector are all the same random constant obtained at the beginning of each particular experiment by randomly selecting this constant from the uniform distribution on (0,1), and then holding it fixed throughout the remainder of that experiment. Four different Σ_n matrices are considered. The first choice is $\Sigma_n = I_n$, where I_n is a 100×100 identity matrix. The second choice is Σ_n has an autoregressive structure; that is $\Sigma_n = ((\sigma_{u,v}))$, where

$$\sigma_{u,v} = \sigma^2 r^{|u-v|}. \tag{6.12}$$

We will use the terminology of AR(1) structures to describe this covariance matrix. In the experiments that yielded the data in Table 1, we chose $\sigma^2 = 3$ and $r = 0.8$. The third choice for Σ_n has

	$\Sigma = I$	AR(1)	CS	UN
Test 1	0.002	0.004	0.0078	0.0022
Test 2	0.0448	0.8478	0.9998	0.1194
Test 3	0.0442	0.0486	0.047	0.0512
Test 4	0.0458	0.0428	0.0404	0.048

Table 1: Type I error rates for the four tests under various population covariance structures.

the property that

$$\sigma_{u,v} = \begin{cases} \sigma_1^2 + \sigma^2 & \text{if } u = v, \\ \rho^\star & \text{if } u \neq v. \end{cases} \tag{6.13}$$

Covariance matrices with this property are said to compound symmetric (CS) and we will use this terminology when convenient. While in general it is not necessary for $\rho^\star$ to be positive, in several applications it turns out to be positive and in the so-called random effects models $\rho^\star = \sigma^2$. In all our simulations with the CS structure we assume that $\rho^\star = \sigma^2 > 0$. In particular, we take $\sigma^2 = 3$ and $\sigma_1^2 = 4$, although other choices could also be employed. The fourth choice for Σ_n is an unstructured (UN) symmetric positive definite 100×100 matrix. This matrix is randomly chosen at the beginning of the experiment and held fixed. We consider four different types of tests. Test 1 and Test 3 are based on our methodology described in this paper. In Test 1 we estimate the covariance matrix by the method of moments while in Test 3, we use the true covariance matrix. In Test 2 we assume that the covariance matrix is the ident! ity matrix, and when the sup-norm is used the Type I error rate determined in Proposition 1 is then equivalent to performing 100 univariate tests and rejecting the hypothesis if at least one of the 100 tests rejects the relevant one dimensional null hypothesis. Test 4 is based on a Bonferonni correction to Test 2. As explained before, to carry out the test one requires estimation of the covariance matrix. We estimate the covariance matrix by the sample covariance matrix $\hat{\Sigma}_n = ((\hat{\sigma}_{n,u,v}))$ using the formula

$$\hat{\Sigma}_n = \frac{1}{n-1} \sum_{i=1}^{n} (\boldsymbol{X}_{n,i} - \boldsymbol{\mu}_{n,0})(\boldsymbol{X}_{n,i} - \boldsymbol{\mu}_{n,0})', \tag{6.14}$$

where $\boldsymbol{\mu}_{n,0}$ is value of the mean vector under the null hypothesis and the prime denotes the transpose. Table 1 provides the observed Type I error rates when the sup-norm is used.

The primary purpose of Table 1 is to describe the basic problems that arise in statistical analyses of data sets using our methodology. It is clear from Table 1 that our methodology works, in the sense of yielding Type I error rates which are closer to the nominal error rates, if the covariance matrix is known. It is not surprising that Test 1 does poorly since one is estimating 5050 parameters from 10 data vectors, each of dimension 100. However, an important observation is that in Test 2, when the true covariance matrix Σ is not the identity matrix, assuming independence or "by-gene" analysis yields substantially inflated Type I error rates. This raises three important questions: (i) suppose one were to model the covariance structure correctly, can one improve the Type I error rate? (ii) What are the consequences of mis-modeling the covariance structure?, and (iii) is there an alternate way to estimate the covariance matrix without making any structural assumptions? Our next simulations address these questions.

	$\rho = 2$	$\rho = 4$	$\rho = \infty$
Case 1	0.0488	0.0482	0.0462
Case 2	0.0838	0.0806	0.0626
Case 3	0.2752	0.2558	0.1474

Table 2: Comparison of Type I error rates for various norms under different information concerning the covariance matrix.

We begin with an experiment addressing questions (i) and (ii). In this experiment we assume the population covariance to have AR(1) structure, as in (6.12), with parameters $\sigma^2 = 3$ and $r = 0.8$. We consider three cases. Case 1 corresponds to the situation that the user knows the true Σ_n. Case 2 corresponds to the case that the user correctly models the structure to be AR(1), while Case 3 corresponds to the situation where the user incorrectly models the structure to be compound symmetric while the true structure is autoregressive. If the covariance matrix is modeled to be compound symmetric, we estimate the covariance parameter using the formula

$$\hat{\rho^\star} = \frac{2}{b(n)(b(n)-1)} \sum_{u=1}^{b(n)} \sum_{v>u}^{b(n)} \hat{\sigma}_{n,u,v}, \tag{6.15}$$

while the common variance is estimated using the formula

$$\hat{\sigma}_n^2 = \frac{1}{nb(n)-1} \sum_{i=1}^{n} (\boldsymbol{X}_{n,i} - \boldsymbol{\mu}_{n,0})'(\boldsymbol{X}_{n,i} - \boldsymbol{\mu}_{n,0}). \tag{6.16}$$

If the covariance matrix is modeled using the autoregressive structure, then the autoregressive parameter r is estimated using the formula,

$$\hat{r}_n = \frac{1}{n} \sum_{i=1}^{n} \hat{r}_i, \tag{6.17}$$

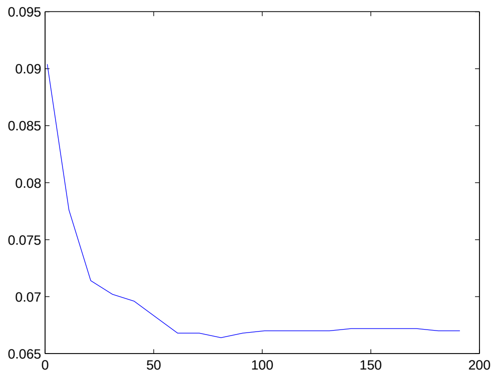
where

$$\hat{r}_i = \frac{\sum_{j=1}^{b(n)-1} (\xi_{n,i,j} - \mu_{n,j,0})(\xi_{n,i,j+1} - \mu_{n,j+1,0})}{\sum_{j=1}^{b(n)} (\xi_{n,i,j} - \mu_{n,j,0})^2}. \tag{6.18}$$

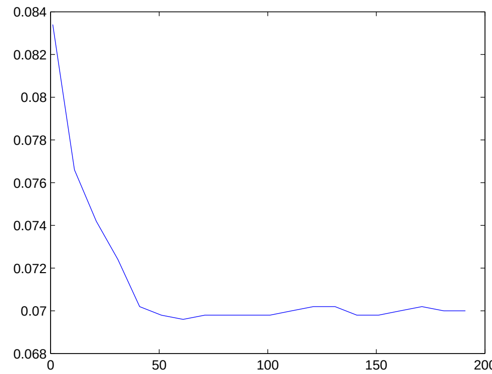
The variance parameter is estimated using the formula (6.16). In this experiment we also consider the role of the norm being used by studying the 2-norm, the 4-norm, and the sup-norm. Of course, we still have $n = 10$ and $b(n) = 100$. Table 2 gives the Type I error rates in this situation. We notice that the Type I error when the infinity norm is used is much closer to the nominal 5% rate compared to the results of Table 1 for Test 1 and Test 2 for the AR(1) column. Table 2 suggests that the effect of the norm is pronounced when the covariance matrix is estimated; furthermore, the error rates seem to get closer to the nominal values as the norm approaches infinity. When the covariance matrix is modeled incorrectly, the error rate increases substantially.

To understand the effect of the norm (used for constructing the test statistic) on the Type I error rate, we performed another numerical experiment with the compound symmetric covariance structure with $\sigma^2 = 2$ and $\sigma_1^2 = 1.5$. We studied several norms starting from $\rho = 2$ to $\rho = 200$ with increments of 1 and the final data point corresponds to $\rho = \infty$. The dimension of the data vectors used in this experiment were $b(n) = 100$, $b(n) = 500$ and $b(n) = 1000$. The value of n remained at

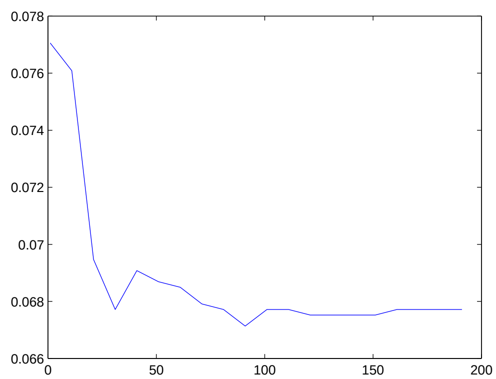
10. The variance and the covariance parameters were estimated using the formulae (6.16) and (6.15), respectively. Figure 2 gives Type I error rates for different norms for different data dimensions. The X-axis is ρ and the Y-axis is the observed Type I error rates for the tests.



(a) $b(n) = 100$



(b) $b(n) = 500$



(c) $b(n) = 1000$

Figure 2: Type I error rates as a function of the norm for various dimensions of the parameter space.

The graphs show that as ρ increases, the observed Type I error rates seem to get closer to the nominal value, even though the nominal rate is not achieved. Additional analysis is required to understand this phenomenon. Moreover, if one assumed that the covariance matrix is diagonal, that is, assuming incorrectly that the covariances are zero, then the test constructed using (6.4) yields an error rate of 0.7292 which is substantially larger than described in Figure 2. This shows that ignoring the correlations when they are actually present, yields a biased methodology as described in the last paragraph of the previous section.

We now move on to describe the interaction between the number of parameters in the covariance matrix, the dimension of the data-vectors, the effect of norms, and incorrect modeling by over specification of the parameters. For this reason we need to introduce yet another covariance structure called the heterogeneous covariance structure (HCS). This structure imposes the following conditions

Data generated using HCS covariance structure

	$b(n) = 100$	$b(n) = 500$	$b(n) = 1000$
$\rho = 2$	0.0986	0.1016	0.1042
$\rho = 4$	0.1774	0.4732	0.695
$\rho = \infty$	0.1602	0.2748	0.367

Data generated using CS covariance structure

	$b(n) = 100$	$b(n) = 500$	$b(n) = 1000$
$\rho = 2$	0.0764	0.0936	0.1042
$\rho = 4$	0.1344	0.302	0.3582
$\rho = \infty$	0.1324	0.3582	0.449

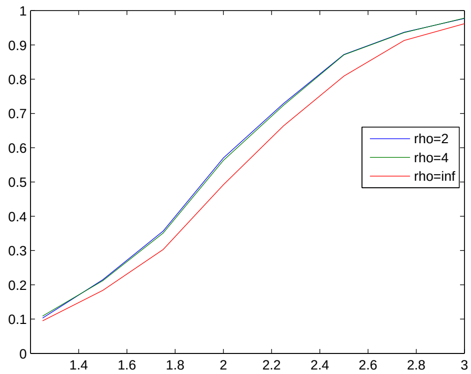
Table 3: Type I error rates under model mis-specifications.

on the covariance matrix, namely

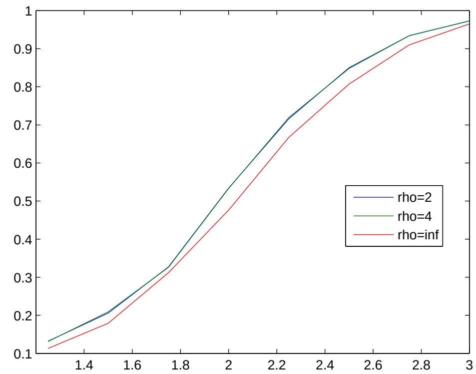
$$\sigma_{u,v} = \begin{cases} \sigma_u^2 + \sigma^2 & \text{if } u = v, \\ \rho^\star & \text{if } u \neq v. \end{cases} \tag{6.19}$$

Again, while it is not necessary that $\rho^\star > 0$, as mentioned previously, we will take it to be positive and equal to σ^2. Note that the number of parameters in this case is $b(n)+1$. We consider two cases. In the first case, the true Σ_n is HCS with $\sigma^2 = 2$ and $\sigma_u^2 = u$, but the user fits CS. The parameters in this case are estimated as described before. In the second case, the true Σ_n is CS with $\sigma^2 = 2$ and $\sigma_1^2 = 1.5$, but the user fits HCS. In this case, the variance parameters are estimated using the sample variances for each component. Table 3 provides the observed Type I error rates of the tests in these cases. Note that as the number of parameters in the covariance matrix increase, smaller values of ρ yield values closer to the nominal values. In particular, this contrasts with the results in Figure 2, where the number of parameters in each graph is constant, and then large values of ρ typically yield observed rates closer to the nominal rate of 0.05. However, also observe that in this current experiment even if the dimension of the data vector is 1000 and $n = 10$, the increase in the Type I error does not exceed 0.0542 when $\rho = 2$. This shows that our methodology is fairly stable under various perturbations of the true model. We now move on to study the power associated with our test procedure.

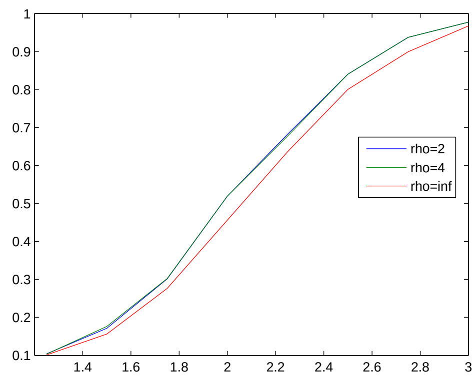
To study the power we choose $\boldsymbol{\mu}_{n,0}$ to be a $b(n)$ dimensional vector all of whose components are equal to one. The covariance structure is taken to be compound symmetric with $\sigma^2 = 2$ and $\sigma_1^2 = 1.5$ There are several alternative hypotheses but we consider the situation where $\boldsymbol{\mu} = (a, a, \cdots a)$ for values of a ranging from 1.25 to 3. As before, we study various data dimensions namely, $b(n) = 100$, $b(n) = 500$ and $b(n) = 1000$ for $\rho = 2$, $\rho = 4$, and $\rho = 1000$. In Figure 3 the X-axis represents the values of a while the Y-axis represents the observed power. The graphs clearly show that even with a sample size of 10 there is reasonable power to detect small differences. It is also clear, expectedly, that the power to detect small changes decreases as the data-dimension increases. Furthermore, the power to detect departures from the null hypothesis using the sup-norm based statistic is less than the statistics with $\rho = 2$ and $\rho = 4$. However, the size of the test is closer to nominal for the sup-norm statistic. This suggests that perhaps there is a very large value of ρ that would yield a nominal Type I error rate and substantial power to detect departures from the null hypothesis.



(a) $b(n) = 100$



(b) $b(n) = 500$



(c) $b(n) = 1000$

Figure 3: Interaction between the power and the dimension for different norms.

As one would expect, this choice may depend on the variance and the underlying data-distribution. Evaluating such a value of ρ as a function of the variance is outside the scope of this paper. However, our simulation results in the next two sections shed some light on this scenario.

Thus what is left to understand is whether it is possible to develop optimal procedures based on the techniques in this paper without making any structural assumptions concerning the covariance matrix. This is question (iii) described previously. Indeed, this is the one of the points of [11], where their techniques do not take into account the structure of the dependencies between the variables. We address this issue in the next subsection.

6.3 Unstructured Covariance Estimation and Shrinkage

It is well-documented in the statistical literature that estimation of the covariance matrix is a difficult problem. It is known that the traditional method of moments estimator is the same as the maximum likelihood estimator when the data distribution is multivariate normal. When the number of variables is larger than the sample size [14] and [28] amongst others have clearly demonstrated that the sample covariance matrix behaves poorly in terms of the mean square error. Borrowing the

Unstructured with Shrinkage

n	$\rho = 2$	$\rho = 4$	$\rho = \infty$	mean λ	var λ
10	0.0358	0.0342	0.0520	0.8886	0.0002416
20	0.0504	0.0330	0.0464	0.9448	0.0003191
30	0.0476	0.0390	0.0374	0.9621	0.0003333
40	0.0498	0.0428	0.0436	0.9694	0.0003322
50	0.0456	0.0404	0.0434	0.9732	0.0003223
60	0.0548	0.0472	0.0408	0.9743	0.0003018
70	0.0504	0.0426	0.0424	0.9762	0.0003073
80	0.0534	0.0444	0.0390	0.9756	0.0003008
90	0.0498	0.0486	0.0504	0.9762	0.000306
100	0.0458	0.0402	0.0438	0.9767	0.0003047

Table 4: Type I error rates with unstructured covariance matrix and Shrinkage

idea from shrinkage estimation, [14] developed an alternative estimator of Σ_n by taking a convex combination of the unstructured sample covariance matrix and a structured covariance matrix. Their estimator is given by

$$\Sigma_n^\star = (1-\lambda)\hat{\Sigma}_n + \lambda\tilde{\Sigma}_n, \tag{6.20}$$

where $\hat{\Sigma}_n$ is the method of moments estimator and $\tilde{\Sigma}_n$ is an estimator assuming a particular structure for the covariance matrix. The parameter λ can be estimated from the data and has a closed form expression when $\tilde{\Sigma}_n$ is taken to be an identity matrix, compound symmetric structure, heterogeneous compound symmetry structure and many other structures. The estimator $\Sigma_n^\star$ possesses the following properties: it (i) minimize the quadratic loss, (ii) has minimal asymptotic risk in a certain class, (iii) is orthogonally invariant and invertible, and (iv) is consistent in a general asymptotic framework based on what is referred to as Kolmogorov asymptotics. These properties makes attractive for use in inference.

In this section, we describe a numerical experiment to evaluate if some of the difficulties described in the previous subsection due to covariance matrix estimation can be minimized by using the shrinkage method for estimating the covariance matrix. For this reason, we first generate a random population covariance matrix Σ_n and fix this matrix. We now generate data from a normal population with mean $\boldsymbol{\mu}_n$ and covariance matrix Σ_n. In our simulations we chose for $\tilde{\Sigma}_n$, a diagonal matrix (which is referred to as Target D in [28]), namely a heterogeneous compound symmetry structure with covariances equal to 0. As explained in [28] this choice also yields a positive definite $\Sigma_n^\star$. In our numerical experiment we study the performance of our methodology when choosing this aforementioned shrinkage estimator. We also look at the role of the norms and the sample sizes. The dimension of the parameter space is fixed at $b(n) = 1!00$, and again $n = 10$. Table 4 shows the Type I error rates, the mean amount of shrinkage, and the variance in the shrinkage. It is clear that as the sample size increases, the mean of the shrinkage parameter increases to 1 thereby suggesting that the estimate gets closer to the structured matrix. We also observe from Table 4 that as n increases, small values of ρ yield Type I error rates that are closer to the nominal 5% level, while for small values of n , $\rho = \infty$ yields error rates that are closer to the 5% level. Thus, the results of this numerical experiment suggest that our methodology combined with a shrinkage estimation of

	$\rho = 2$	$\rho = 4$	$\rho = \infty$	mean λ	var λ	mean λ_v	var λ_v
$p = 100$	0.1312	0.0838	0.0686	0.8361	0.0009	0.2617	0.0029
$p = 500$	0.1914	0.0874	0.057	0.8555	0.0002	0.2485	0.0009
$p = 1000$	0.237	0.1042	0.0526	0.8493	0.0002	0.2402	0.0005

Table 5: Type I error rates for using information from leukemia data set shrinking both the variances and covariances.

the covariance matrix could yield a sound technique for data analyses. We pursue this aspect in the subsections below.

6.4 Numerical Experiment using Information from Leukemia Data

To illustrate the methodology, we now use information in a publically avaibale data set called the leukemia data set described in [7]. This data set contains the gene expression values for two types of leukemia, acute lymphoblastic leukemia (ALL) and acute myeloid leukemia (AML). We use the same pre-processing step as described in Section 3.1 of [8], We retain 3571 genes from 72 patients, 38 from ALL group and 25 from AML group. On the remaining 3571 genes, we apply the standardization technique described in Section 3.3 of [8].

In this simulation study we only use the information from the AML group. We consider $b(n) = 100, 500, 1000$ dimensions. The simulation experiments are based on data generated from $N_{b(n)}(\boldsymbol{\mu}_n, \Sigma_n)$ for a sample of size $n = 10$. We now describe how $\boldsymbol{\mu}_n$ and Σ_n are obtained from the data.

1. Fix $b(n)$ (e.g. $b(n) = 100$).
2. Randomly select $b(n)$ genes (without replacement) from the 3571 genes in the data set.
3. From the 25 AML patients estimate the mean from these $b(n)$ genes. Call the resulting estimate $\boldsymbol{\mu}_n$.
4. From the 25 AML patients estimate the covariance matrix, shrinking the covariances, from these $b(n)$ genes. Call the resulting estimate Σ_n.

Now, for a fixed $b(n)$, we use the same $\boldsymbol{\mu}_n$ and Σ_n in all 5000 simulations.

We shrunk the variances as well as the covariances; λ denotes the optimal covariance shrinkage factor where λ_v denotes the optimal variance shrinkage factor. Table 5 provides the Type I error rates for this one-sample problem.

6.5 Mixture distributions and the role of variances

Motivated by the results of the previous section, in this section we present numerical experiments that bear similarities with real data sets. For this reason, we study situations when data are from normal or mixtures of normal populations. To make our results comparable and provide insight into various phenomenon, it is necessary to understand the variances and covariances especially when the covariance matrix is assumed to be unstructured. We begin with a scenario when the data

Σ Known

$(n,\,b(n))$	$\rho = 2$	$\rho = 4$	$\rho = \infty$
$(10, 100)$	0.1714	0.1014	0.0596
$(20, 200)$	0.1856	0.096	0.0456
$(30, 300)$	0.213	0.1048	0.0454
$(40, 400)$	0.2608	0.1304	0.0456
$(50, 500)$	0.277	0.15	0.045

Σ Unknown

$(n,\,b(n))$	$\rho = 2$	$\rho = 4$	$\rho = \infty$	mean λ	var λ
$(10, 100)$	0.0532	0.043	0.0600	0.8489	5.33E-04
$(20, 200)$	0.0948	0.073	0.0724	0.8712	3.80E-04
$(30, 300)$	0.1242	0.089	0.073	0.8354	4.00E-04
$(40, 400)$	0.1314	0.1018	0.0782	0.7662	4.41E-04
$(50, 500)$	0.143	0.105	0.0714	0.7335	4.14E-04

Table 6: Type I error rates for known and unknown Σ as a function of ρ and the dimension.

are generated from a mixture of normal populations. That is, let $\boldsymbol{X}$ be a generic $b(n)$ dimensional random variable with distribution given by

$$\boldsymbol{X} \sim \begin{cases} N(\boldsymbol{\mu}_n, \Sigma_n^a) & \text{with probability } p \\ N(\boldsymbol{\mu}_n, \Sigma_n^b) & \text{with probability } q = 1 - p. \end{cases} \tag{6.21}$$

In the above, p is called the mixing proportion and was taken to be 0.5 in all our simulations. $\boldsymbol{\mu}$ is the mean of the distribution and was randomly chosen from the uniform distribution and held fixed, as described previously. The covariances matrices, of order $b(n) \times b(n)$, are given by

$$\sigma^a_{u,v} = \begin{cases} 1 & \text{if } u = v = 1, \\ \frac{1}{\log(u+1)} & \text{if } 2 \leq u = v \leq b(n), \\ .01 & \text{if } 1 \leq u \neq v \leq b(n). \end{cases}$$

$$\sigma^b_{u,v} = \begin{cases} 3 & \text{if } u = v = 1, \\ \frac{3}{\log(u+1)} & \text{if } 2 \leq u = v \leq b(n), \\ .008 & \text{if } 1 \leq u \neq v \leq b(n). \end{cases}$$

These covariance matrices were chosen so that there was ease in computation, and that the conditions are close to those required in Theorem 6. The dimension of the data vectors are chosen to be $b(n) = 10n$. Type I error rates are given in the Table 6. The covariance matrix is estimated using the shrinkage method described previously. The results show that, if Σ_n is known and satisfies the conditions indicated above, then the error rates are close to nominal; if Σ_n is unknown, then using the shrinkage technique of estimation, error rates are closer to nominal even though there is a small inflation in the error rates. The general trend again seems to suggest that large values of ρ yield results closer to the nominal value. However, since the data were generated to satisfy conditions close to those required in Theorem 6, perhaps other situations would emerge under parameter choices

Case 1: Small variances

	$b(n) = 100$	$b(n) = 500$	$b(n) = 1000$
$\frac{1}{b(n)} \sum_{u=1}^{b(n)} \sigma^{\star}_{u,u}$	2.3304	2.3238	2.328
$\min \sigma^{\star}_{u,u}$	1.8154	1.8377	1.8356
$\max \sigma^{\star}_{u,u}$	2.8074	2.8443	2.8359
$\frac{2}{b^2(n)-b(n)} \sum_{v>u} \sigma^{\star}_{u,v}$	-0.001	-0.0003	-0.0001
$\min_{u \neq v} \sigma^{\star}_{u,v}$	-0.0353	-0.0189	-0.0137
$\max_{u \neq v} \sigma^{\star}_{u,v}$	0.0336	0.0201	0.0141
$\frac{2}{b^2(n)-b(n)} \sum_{v>u} \lvert\sigma^{\star}_{u,v}\rvert$	0.0078	0.0035	0.0024

Case 2: Large variances

	$b(n) = 100$	$b(n) = 500$	$b(n) = 1000$
$\frac{1}{b(n)} \sum_{u=1}^{b(n)} \sigma^{\star}_{u,u}$	10.6312	51.9486	101.156
$\min \sigma^{\star}_{u,u}$	8.8822	47.0884	93.5997
$\max \sigma^{\star}_{u,u}$	12.1996	56.5424	107.4229
$\frac{2}{b^2(n)-b(n)} \sum_{v>u} \sigma^{\star}_{u,v}$	-0.0161	-0.0076	-0.0051
$\min_{u \neq v} \sigma^{\star}_{u,v}$	-0.5056	-0.571	-0.5793
$\max_{u \neq v} \sigma^{\star}_{u,v}$	0.5278	0.5456	0.6362
$\frac{2}{b^2(n)-b(n)} \sum_{v>u} \lvert\sigma^{\star}_{u,v}\rvert$	0.1073	0.1041	0.1012

Table 7: Properties of the covariance matrix with “small” and “large” variances.

more in line with the conditions required for the asymptotic normality results in Theorem 7 and Remark 11.

We now move on to consider the variance issue. In the previous sections, our results indicated that under mis-specifications the error rates start deviating from the nominal values. Furthermore, there were differences between the error rates for UN, HCS and CS structures (see for instance Table 1 and Table 3). It is not immediately clear if these results are comparable since the construction of the test statistic did not involve the covariance matrix, as is the case in most traditional statistical problems. However, the test performed did involve the covariance matrix. In particular since the covariance matrices used in the simulations were chosen at random and kept fixed for simulations, (especially for covariance matrices which are unstructured and with HCS structure) one cannot conclude, without further evidence, that the results “depend only on the number of parameters involved and not on the actual values of variances and covariances in a covariance matrix”. Our next numerical experiment addresses this issue, and in the process illustrates that our methodology possesses these invariance properties normally prevalent in traditional t-type statistics.

We now describe the experiment that sets the variances and covariances on a comparable scale. For this experiment, we resort to data from the normal distribution. We consider three situations: (i) the population covariance matrix is unstructured; (ii) the population covariance matrix has CS structure; and (iii) the population covariance matrix has HCS structure. In Table 7, we describe properties of the unstructured covariance matrix $\Sigma_n = ((\sigma^{\star}_{u,v}))$ with “small” and “large” variances. We use this matrix to construct covariance matrices with CS and HCS structure as described below.

To allow comparisons between results, we set the variance parameter σ_1^2 of the covariance

UN

	$\rho = 2$	$\rho = 4$	$\rho = \infty$	mean λ	var λ
$b(n) = 100$	0.0326	0.0314	0.0486	0.8889	2.29E-04
$b(n) = 500$	0.0128	0.0676	0.1698	0.8888	9.30E-06
$b(n) = 1000$	0.003	0.11	0.2732	0.8888	2.36E-06

CS

	$\rho = 2$	$\rho = 4$	$\rho = \infty$	mean λ	var λ
$b(n) = 100$	0.0374	0.0384	0.0414	0.8889	2.35E-04
$b(n) = 500$	0.04	0.0404	0.0488	0.8889	8.98E-06
$b(n) = 1000$	0.044	0.0458	0.0432	0.8889	2.39E-06

HCS

	$\rho = 2$	$\rho = 4$	$\rho = \infty$	mean λ	var λ
$b(n) = 100$	0.0452	0.0416	0.06	0.889	2.30E-04
$b(n) = 500$	0.0368	0.11	0.18	0.8889	9.19E-06
$b(n) = 1000$	0.036	0.2488	0.3018	0.8889	2.34E-06

Table 8: Type I error rates for normal model using various covariance structures-the case of small variances

matrix with CS structure to be $\frac{1}{b(n)}\sum_{i=1}^{b(n)}\sigma^{\star}_{u,u}$; we also set the covariance parameter σ^2 to be $\frac{2}{b^2(n)-b(n)}\sum_{v>u}|\sigma^{\star}_{u,v}|$. In the context of HCS, the parameters are taken as follows:

$$\sigma_{u,v} = \begin{cases} \frac{2}{b^2(n)-b(n)}\sum_{v>u}|\sigma^{\star}_{u,v}| \text{ if } u \neq v \\ \sigma^{\star}_{u,u} + \frac{2}{b^2(n)-b(n)}\sum_{v>u}|\sigma^{\star}_{u,v}| \text{ if } u = v. \end{cases} \tag{6.22}$$

Of course, the sum $\sum_{v>u}$ is actually $\sum_{b(n)\geq v>u\geq 1}$, but we suppress that to simplify our notation.

We consider three cases: (i) when the statistician models the data as unstructured and uses the shrinkage algorithm to estimate the covariance matrix; (ii) The statistician models the data correctly as CS ; and (iii) the statistician models the data correctly as HCS. In cases (ii) and (iii) the variance and covariance parameters are estimated using the appropriate formulae described in the earlier sub-sections of this section. Table 8 provides error rates for the three cases when the variance is “small” while Table 9 provides the error rates when the variance is “large”.

It is clear from these tables that as the number of parameters increase, and if the variances do not decay, both UN and HCS structures yield error rates that are different from the nominal values (compare Table 6). However, with CS structure the method yields error rates that are close to the nominal values, even for large dimensions, since such a structure requires estimation of only two parameters. This phenomenon seems to be consistent for both small and large variances in the data set.

Finally, we study the behavior of shrinkage methods when the distribution is a mixture distribution and no structure is assumed. The two covariance matrices are taken to be the Case 1 of Table 7 and the compound symmetry structure derived from Case 1, Table 7, as explained previously. The error rates are summarized in Table 10.

Our results clearly show that for smaller values of the dimension, larger ρ yields results closer to the nominal value while smaller values of ρ yield optimal results for larger dimension. The simulation

UN

	$\rho = 2$	$\rho = 4$	$\rho = \infty$	mean λ	var λ
$b(n) = 100$	0.0378	0.0344	0.055	0.8883	2.30E-04
$b(n) = 500$	0.0082	0.0718	0.1762	0.8888	9.39E-06
$b(n) = 1000$	0.0036	0.1068	0.2866	0.8889	2.40E-06

CS

	$\rho = 2$	$\rho = 4$	$\rho = \infty$	mean λ	var λ
$b(n) = 100$	0.0432	0.0454	0.0462	0.8888	2.35E-04
$b(n) = 500$	0.0414	0.0434	0.0488	0.8889	9.20E-06
$b(n) = 1000$	0.0394	0.039	0.046	0.8889	2.36E-06

HCS

	$\rho = 2$	$\rho = 4$	$\rho = \infty$	mean λ	var λ
$b(n) = 100$	0.0394	0.0352	0.0574	0.8883	2.33E-04
$b(n) = 500$	0.037	0.118	0.199	0.8889	9.12E-06
$b(n) = 1000$	0.0432	0.2578	0.3214	0.8889	2.35E-06

Table 9: Type I error rates for a normal model using various covariance structures-the case of large variances

	$\rho = 2$	$\rho = 4$	$\rho = \infty$	mean λ	var λ
$b(n) = 100$	0.0324	0.0332	0.0574	0.8876	2.55E-04
$b(n) = 500$	0.0178	0.0778	0.1786	0.8874	1.21E-05
$b(n) = 1000$	0.0114	0.1336	0.302	0.8873	3.93E-06

Table 10: Type I error rates for mixture model with covariance matrix estimated using the shrinkage algorithm

results are suggestive that this phenomenon is independent of the variances and covariances and the choice of the data distributions.

Thus based on our simulation results contained in Tables 1 through 10 and theoretical results, it is reasonable to draw the following conclusions: the methodology proposed in Section 6 combined with shrinkage estimation of covariance matrix works well (in the sense of yielding nominal Type I error rates) with small sample size and reasonably large dimension under no assumptions on the structure of the covariance matrix. If the covariance matrix is assumed to be structured with few parameters, then the methodology works even when the data dimension is very large. However, if the number of parameters in the structured matrix also is large, then the data dimension needs to be reasonable for valid inference. Furthermore, mis-modeling the covariance structure would yield substantially biased inference, the worst case being assuming the covariance structure to be the identity (Table 1). All these phenomenon hold when the sup-norm is used for constructing the test statistic. Also, since the conditions for asymptotic normality hold more generally for the sup-norm case, this norm is convenient to use for data analyses. In addition, if the conditions for the sup-norm are satisfied and the data dimension is very large relative to the sample size, use of the 2-norm for data analyses may yield more accurate results (Table 8 and Table 9). Further simulation results using structured covariances that arise in these problems are available in [13].

7 Identifying Differentially Expressed Genes

In micro array analyses, it is often the case that a scientist is concerned with identifying if two sets of genes are differentially expressed. Methods like cluster analysis are not confirmatory but does provide some insight into identifying a set of differentially expressed genes. The confirmatory tool currently in popular use, typically involves methods like $t-$test and FDR. In this section, we describe how methods of this paper can be adopted to identifying the differentially expressed genes. Since identifying differentially expressed genes relates to the mean expression profiles of two populations, it is two-sample problem. We now describe how to adopt the methodology from Section 6.1 to handle a particular case of a two-sample problem related to our data analysis. Since we have two populations we will use the superscript (i) to denote the appropriate population. For instance, $N_{n,i}^{(j)}$ will represent the number of genes from population j. We deal with the case $N_{n,i}^{(1)} = N_{n,i}^{(2)} = b(n)$ and that $R_{n,i,j} \equiv 1$. Under this model assumption, let $\boldsymbol{\mu}_n^{(j)} = E(\vec{\boldsymbol{X}}_n^{(j)})$. The null hypothesis for the two-sample problem is $\boldsymbol{\mu}_n^{(1)} = \boldsymbol{\mu}_n^{(2)}$. Now, let n_1 and n_2 denote the number of replicates from each group. Then, we use the quantity $||\mathbf{S}_{n,n_1}^{(1)} n_1^{-1/2} - \mathbf{S}_{n,n_2}^{(2)} n_2^{-1/2}||_\rho = ||\tilde{\boldsymbol{T}}_{n,n_1}^{(1)} - \tilde{\boldsymbol{T}}_{n,n_2}^{(2)}||_\rho$ as the test statistic and compare it to its distribution obtained via monte carlo methods.

We now apply this methgodology to the leukemia data set described in [7], which we also used in the previous section. We analyzed the sixteen genes given in Table 2 of [31]. The value of the test statistics, using our methods, were determined to be 4.4992, 2.7396, and 2.0250 for 2-norm, 4-norm, and sup-norm respectively. The corresponding p-values were all less than 10^{-6} showing that the genes are differentially expressed. The same conclusion was also obtained by Xiting *et. al.* ([31]) using different methods. More importantly, as explained in [31], these genes have biological significance and the three existing statistical methods in popular use did not identify them to be differentially expressed.

8 Concluding Remarks

In this paper we developed results concerning the joint consistency and joint asymptotic normality of estimators of parameters of several variables whose dimension increases as the sample size increases by viewing it as an infinite dimensional problem. We used these results to develop test statistics and their asymptotic limit distributions. In particular, existence of the limiting infinite dimensional Gaussian distributions justifies the methodology described in Section 6. Our simulation results bring out the importance of estimating the covariance matrix well. Furthermore, although our central limit theorems require technical conditions, when applied to i.i.d. random vectors in the various settings, some of these conditions are extremely close to being optimal. It is also important to note that our simulations suggest they yield useful results under a variety of simplifications of the precise assumptions of these results. Indeed, apart from establishing the infinite dime! nsional results under various missing mechanisms, our results help develop procedures for joint inference in multivariate problems where the dimension of the parameter space increases with the sample size. In particular, our results provide asymptotic justification of the joint marginal inference in a one-way random effects model with incomplete data and when the number of parameters increase with the sample size. Extensions of our methodology to more general models and related questions are under consideration.

Acknowledgements: Authors would like to thank Mr. Bret Hanlon for help with simulations. Mr. Hanlon is a student of the second author.

References

[1] Anderson, T.W. (2003). *An introduction to Multivariate Statistical Analysis*,John Wiley & Sons, New Jersey.

[2] Araujo, A. and Giné, E. (1980). *The central limit theorem for real and banach valued random variables*, John Wiley & Sons, New York.

[3] Buldygin, V.V. and Dončenko, V.S.(1977). Convergence to zero of Gaussian sequences, *Mat. Zemetki*, 21, 531-538.

[4] Chung, K. L. (1974) *A Course in Probability Theory*, 2^{nd} edition, Academic Press, New York.

[5] Devroye, L. and Gyorfi, L. (1985). *Nonparametric density estimation: The L_1 view*, John Wiley & Sons, New York.

[6] Feller, W. (1966). *An introduction to probability theory and its applications*, John Wiley & Sons, New York.

[7] Golub, T. R., Slonim, D. K., Tamayo, P., Huard, C., Gaasenbeek, M.,Mesirov, J. P., Coller, H., Loh, M., Downing, J. R., Caligiuri, M.A., Bloomfield, C. D., and Lander, E. S. (1999). Molecular Classification of Cancer: Class Discovery and Class Prediction by Gene Expression Monitoring, *Science*, 286, 531-537.

[8] Dudoit, Sandrine, Fridlyand, Jane, and Speed, Terence P. (2002).Comparison of discrimination methods for the classification of tumors using gene expression data, *J. Amer. Statist. Assoc.*, 97, 77–87.

[9] Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables, *Journal of American Statistical Association*, 58, 13-30.

[10] Huber, P.(1973). Robust regression:asymptotics, conjectures and Monte Carlo, *The Annals of Statistics*, 1, 799-821.

[11] Kosorok, M. and Ma, S. (2007). Marginal asymptotics for the large p, small n paradigm: With application to Microarray data, *The Annals of Statistics*, 35, 1456-1486.

[12] Kuelbs, J. (1978). Some exponential moments of sums of independent random variables, *Transactions of the American Mathematical society*, 240, 145-162.

[13] Kuelbs, J. and Vidyashankar, A.N. (2008). Simulation Report using Structured Covariances, *Preprint*.

[14] Ledoit, O. and Wolf, M. (2004). A well-conditioned estimator for large dimensional covariance matrices, *Journal of Multivariate Analysis*, 88, 365-411.

[15] Lu, Yan, Liu, Peng-Yuan, Xiao, Peng, and Deng, Hong-Wen. (2005). Hotelling's T^2 multivariate profiling for detecting differential expression in microarrays, *Bioinformatics*, 21, 3105-3113.

[16] Massart, P. (1990). The tight constant in the Dvoretzky-Kiefer-Wolfowitz inequality, *The Annals of Probability*, 3, 1269-1283.

[17] Muirhead, R.J. (1982).*Aspects of multivariate statistical theory*, John Wiley & Sons, New Jersey.

[18] Okamato, Masashi(1958). Some inequalities relating to the partial sum of binomial probabilities, *Annals of the Institute of Statistical Mathematics*, 10, 29-35.

[19] Parthasarathy, K. R. (1967). *Probability measures on metric spaces*, Academic Press, New York.

[20] Paulauskas, V. (1984). On the central limit theorem in c_0, *Probability and Mathematical Statistics*, 3, 127-141.

[21] Paulauskas, V., Rackauskas, A., Sakalauskas, V. (1983). Central limit theorems in the space of sequences converging to zero, *Litovskii Matematicheskii Sbornik*, 23, 163-174.

[22] Portnoy, S. (1984). Asymptotic behavior of M-estimators of p regression parameters when $\frac{p^2}{n}$ is large. I. Consistency, *The Annals of Statistics*, 12, 1298-1309.

[23] Portnoy, S. (1985). Asymptotic behavior of M-estimators of p regression parameters when $\frac{p^2}{n}$ is large. II. Normal Approximation, *The Annals of Statistics*, 12, 1298-1309.

[24] Portnoy, S. (1986). On the central limit theorem in R^p when $p \to \infty$, *Probability Theory and Related Fields*, 4, 571-583.

[25] Portnoy, S. (1987). A central limit theorem applicable to robust regression estimators, *Journal of Multivariate Analysis*, 1, 24-50.

[26] Portnoy, S. (1988). Asymptotic behavior of likelihood methods for exponential families when the number of parameters tends to infinity, *The Annals of Statistics*, 15, 356-366.

[27] Ross, Sheldon (2006). *A First Course in Probability*, 7^{th} edition, Pearson Prentice Hall, New Jersey.

[28] Schaffer, J. and Strimmer, K.(2005). A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics, *Statistical Applications in Genetics and Molecular Biology*, 4, 1-30.

[29] van der Lann, M.J., and Bryan, J. (2001). Gene expression analysis with parametric bootstrap, *Biostatistics*, 2, 445-461.

[30] van der Vaart, A. W., and Wellner, J.A. (1996). *Weak Convergence and Empirical Processes*, Springer, New York.

[31] Xiting Yana, Minghua Denga, Wing K. Fungb and Minping Qian (2005). Detecting differentially expressed genes by relative entropy, *Journal of Theoretical Biology*, Volume 234, Issue 3, Pages 395-402.