Strong approximations of level exceedences related to multiple hypothesis testing

PETER HALL[1] and QIYING WANG[2]

[1]*Department of Mathematics and Statistics, The University of Melbourne, VIC 3010, Australia.*
E-mail: halpstat@ms.unimelb.edu.au
[2]*School of Mathematics and Statistics F07, University of Sydney, NSW 2006, Australia.*
E-mail: qiying@maths.usyd.edu.au

Particularly in genomics, but also in other fields, it has become commonplace to undertake highly multiple Student's t-tests based on relatively small sample sizes. The literature on this topic is continually expanding, but the main approaches used to control the family-wise error rate and false discovery rate are still based on the assumption that the tests are independent. The independence condition is known to be false at the level of the joint distributions of the test statistics, but that does not necessarily mean, for the small significance levels involved in highly multiple hypothesis testing, that the assumption leads to major errors. In this paper, we give conditions under which the assumption of independence is valid. Specifically, we derive a strong approximation that closely links the level exceedences of a dependent "studentized process" to those of a process of independent random variables. Via this connection, it can be seen that in high-dimensional, low sample-size cases, provided the sample size diverges faster than the logarithm of the number of tests, the assumption of independent t-tests is often justified.

Keywords: false discovery rate; family-wise error rate; genomic data; large deviation probability; moving average; Poisson approximation; Student's t-statistic; upper tail dependence; upper tail independence

1. Introduction

Today it is commonplace to undertake highly multiple hypothesis testing, generally in genomics and very often using tests based on Student's t-statistic; see, for example, Benjamini and Yekutieli (2001), Efron and Tibshirani (2002), Cui and Churchill (2003), Amaratunga and Cabrera (2004), page 114, Scheid and Spang (2005), Shaffer (2005), Fox and Dimmic (2006), Hu and Willsky (2006), Qiu and Yakovlev (2006), Efron (2007a), Liu and Hwang (2007) and van de Wiel and Kim (2007). This popularity of multiple t-testing also extends to other fields (e.g., Pawluk-Kołc *et al.* (2006)). The principal methods used to control the family-wise error rate and false discovery rate are founded on the assumption of independence among tests. Alternative approaches are generally based either on Bonferroni bounds, which are unsatisfactory for a variety of reasons (see, e.g., Perneger (1998)), or on the hope that, despite ample evidence of non-independence in terms of correlation analysis, independence can be assumed in practice.

The latter hope tends to be pinned either on work of Benjamini and Yekutieli (2001), who argued that in some settings, the absence of independence can give conservative results, or on experience with the analysis of financial data, which suggests that in some circumstances, it might be reasonable to assume that the upper tails of the test statistics are independent, even if

the joint distributions are not. Upper tail independence, as it is sometimes called (for discussion, see, e.g., Wu (1994), Falk and Reiss (2001), R. Schmidt (2002), Li (2006), R. Schmidt and Stadtmüller (2006), T. Schmidt (2007)), is generally assumed to be non-asymptotic in nature. That is, tails of joint distributions are often taken to be perfectly independent beyond a certain threshold.

However, this type of model is not really appropriate for the analysis of genomic data. In particular, it is difficult to determine a biological reason for, or the actual location of, a threshold. It is of greater practical interest to consider the possibility that the strength of dependence in upper tails could become successively weaker as the number of simultaneous tests, and the number of data vectors, increases. If this could be established in the context of tests based on Student's t-statistic, it would lend immediate justification to the often-made assumption (see the articles cited in the first paragraph of this paper) that highly multiple t-statistics can be taken to be independent.

The present paper will establish such a result. The mechanism for our model involves the critical points for tests becoming more extreme as the number, p, of tests diverges (in fact, the increase in critical points is a direct consequence of p diverging) so that the tests are conducted further into the tails; furthermore, the tails of the distributions of test statistics becoming successively lighter as the number of degrees of freedom of the test statistics increases.

We impose particularly weak conditions on the marginal distributions of components. In particular, the distributions need only three finite moments. With this assumption, and permitting the size, n, of the group sample to increase a little faster than the logarithm of the number of tests, it follows from our results that the joint distributions of test statistics enjoy an asymptotic form of the upper tail independence property.

This result would not be so striking if the statistics had normal distributions, but it fails for heavy-tailed distributions such as those for which not all moments are finite. Of course, Student's t-distribution is itself in this category, yet our results show that asymptotic independence holds in a particularly strong sense for Student's t-statistic, even if it is computed from relatively heavy-tailed data. The reason this is possible is that we permit the group sample size to increase at a rate that is just sufficient to convert heavy tails to tails that are sufficiently light, to enable approximate independence at high levels.

It can be seen from this property that the availability of upper-tail asymptotic independence is a bonus of working with highly multiple hypothesis testing, that is, with "large p and small n" problems. It is not available in more conventional, "small p and large n" problems, where there is a very large literature on modelling dependence in highly multiple hypothesis testing.

There is a literature on comparing studentized means when the variances used for studentizing are computed from pooled data and so are common to each test statistic. However, in our experience, that approach is used less frequently, in practice, than the "local" standardization treated in the present paper. When using the latter method, each mean is divided by the standard deviation of the sample from which it was computed. A major motivation is that the true variances may be different in each instance. Even if the variances can reasonably be assumed to be the same, it can be desirable to use the local approach since it confers greater robustness. For example, when applied to the mean alone, rather than its locally studentized form, the large-deviation properties that underpin the analysis of high-level exceedences require the data to have lighter tails.

Statistical literature on highly multiple hypothesis testing is outlined in helpful reviews by Hochberg and Tamhane (1987), Pigeot (2000), Dudoit *et al.* (2003), Bernhard *et al.* (2004) and

Lehmann and Romano (2005), Chapter 9. Benjamini and Hochberg (1995) introduced an approach, which has become very popular, to the controlling of false discovery rates; see also Simes (1986), Hommel (1988), Hochberg (1988), Sarkar and Chang (1997), Sarkar (1998), Sen (1999), Hochberg and Benjamini (1990) and Lehmann *et al.* (2005). Benjamini and Yekutieli (2001) specified conditions under which simultaneous, dependent hypothesis tests, conducted as though they were independent, give conservative results; Benjamini and Yekutieli (2005) addressed similar issues in the context of false coverage-statement rate. Sarkar (2002) extended the work of Benjamini and Yekutieli (2001). Efron (2007b) suggested correlation corrections for large-scale simultaneous hypothesis testing. Blair *et al.* (1996) proposed methods for controlling family-wise error rates in multiple procedures, Holland and Cheung (2002) discussed robustness of family-wise error rates and Clarke and Hall (2009) discussed robustness of testing procedures based on means.

2. Results and applications

2.1. Model and main results

Given $p, n \geq 1$, assume that for $1 \leq i \leq p$ and $1 \leq j \leq n$, we observe data U_{ij}, which we use to construct t-statistics $T_i = n^{1/2} \bar{U}_i / S_i$, where $\bar{U}_i = n^{-1} \sum_j U_{ij}$ and $S_i^2 = n^{-1} \sum_j U_{ij}^2 - \bar{U}_i^2$. In practice, the statistic T_i is used to test the hypothesis that the ith group has zero mean, against a one-sided alternative. When controlling the level of family-wise error rate (FWER) for step-down tests, we require the values of probabilities $P(T_i > t$ for $i = i_1, \ldots, i_k)$ for different levels t and different subsets $\{i_1, \ldots, i_k\}$ of $\{1, \ldots, p\}$. Theorem 1 below will enable us to compute these through approximation by the case where the T_i's are all independent; see Section 2.3 for further details.

We standardize S_i^2 by dividing by n, rather than $n - 1$, since the former is more common in nonparametric problems, but the results below are unaffected by this issue. Since we studentize, there is no loss of generality in assuming that the variance of each component equals 1. More particularly, we ask that

$$0 \leq E(U_{i1}) = d_i, \qquad \text{var}(U_{i1}) = 1 \qquad \text{for all } i, \qquad \sup_{i \geq 1} E(|U_{i1}|^3) < \infty, \qquad (2.1)$$

where $d_1, d_2, \ldots$ is a sequence of constants. The assumption that $d_i \geq 0$ is made here because, in the great majority of practical applications, the hypothesis alternative to the null entails the zero level being exceeded. Accordingly, the tests are one-sided, hence our preoccupation with exceedences of a level. However, minor modifications of our arguments permit the two-sided case to be treated.

Further, we assume that for an integer $\kappa > 0$,

the random vectors $(U_{1i}, U_{2i}, \ldots)$, for $i \geq 1$, are independent and identically distributed, the sequence of random variables $U_{11}, U_{21}, \ldots$ is κ-dependent and $\max_{i_1, i_2: i_1 \neq i_2} \rho_{i_1 i_2} < 1$,   (2.2)

where $\rho_{i_1 i_2} = \text{corr}(U_{i_1 1}, U_{i_2 1})$. The third moment condition in (2.1) permits the variables U_{ij} to have relatively heavy-tailed distributions, for example, a Pareto distribution with tail exponent greater than 3.

The assumption of short-range correlation in (2.2) is, of course, an oversimplification, but it reflects the low level of correlation that is often observed in practice. For example, Messer and Arndt (2006) argue that correlation decays from about 0.08, at a separation of approximately two base pairs, to about 0.01 for a separation of ten base pairs. Results reported by Mansilla *et al.* (2004) corroborate these figures if we assume that their data are normally distributed. More generally, Almirantis and Provata (1999) give evidence of both short-range and long-range correlation, depending on the nature of the DNA or RNA under investigation.

The relationship between the group size, n, and the number of hypothesis tests, p, is assumed to satisfy

$$\log p = \mathrm{o}(n). \tag{2.3}$$

This allows the group size to be very much smaller than the number of tests. In the absence of more detailed assumptions about the distributions of the U_{ij}'s, (2.3) is necessary for the theorem we shall give below. To appreciate why, note that if the U_{ij}'s are independent and identically distributed with an atom at zero and, in particular, if $\delta \equiv P(U_{ij} = 0) > 0$, then, with probability at least δ^n, the t-statistic T_1 assumes the indeterminate value $0/0$. In such cases, we shall take $T_1 = 1$, but in order for the theorem to have a meaningful interpretation when t is the $(1 - p^{-1})$-level quantile of the standard normal distribution, it is essential that the probability that $T_1 = 0/0$ be of smaller order than p^{-1}. Therefore, we require $\delta^n = \mathrm{o}(p^{-1})$ for all $0 < \delta < 1$ and this assumption is equivalent to (2.3).

Define

$$\alpha = \frac{1}{4} \min_{i_1, i_2 : \, i_1 \neq i_2} (1 - \rho_{i_1 i_2}) \tag{2.4}$$

and $\gamma = \alpha + 1$. Condition (2.2) implies that $\alpha > 0$ and, of course, $\alpha \leq \frac{1}{2}$. Given $\eta > 0$, let $t = t(p)$ satisfy

$$(1 + \eta)\sqrt{2\gamma^{-1} \log p} \leq t = \mathrm{O}(\sqrt{\log p}), \qquad \max_{1 \leq i \leq p} d_i = \mathrm{o}(t/\sqrt{n}), \tag{2.5}$$

where $d_1, d_2, \ldots$ are as in (2.1). If t satisfies the first part of (2.5), then any function ϕ which satisfies, as $t \to \infty$,

$$\phi(t) = \exp\{\mathrm{o}(t^2)\}\{\exp(-\tfrac{1}{4}t^2) + p \exp(-\gamma t^2/2)\} \tag{2.6}$$

converges to zero as $p \to \infty$. In the arguments in Section 3, we shall use this notation generically; while ϕ will satisfy (2.6), it will alter from one appearance to another. Strictly speaking, it is not essential to take p to diverge. Although that condition motivates the assumption of divergent t and is, in turn, motivated by the contemporary high-dimensional problems that led to this work, it is not necessary for the theorem below.

Theorem 1. *If* (2.2)–(2.6) *hold, then there exists a probability space on which are defined random variables* $T_1^{\mathrm{new}}, \ldots, T_p^{\mathrm{new}}$ *and* $T_1', \ldots, T_p'$ *such that* (i) *the joint distribution of*

$T_1^{\text{new}}, \ldots, T_p^{\text{new}}$ *is identical to that of* $T_1, \ldots, T_p$; (ii) *the random variables* $T_1', \ldots, T_p'$ *are independent and distributed, respectively, as* $T_1, \ldots, T_p$; *and* (iii) *with probability equal to* $1 - \phi(t)$, *the exceedences of t by* $T_1^{\text{new}}, \ldots, T_p^{\text{new}}$ *occur at the same indices and take the same values as the exceedences of t by* $T_1', \ldots, T_p'$.

To interpret the theorem, note that we would normally expect the dependent data set $T_1, \ldots, T_p$ to exhibit clusters of level exceedences, rather than the single, isolated exceedences associated with the independent sequence $T_1', \ldots, T_p'$. The fact that the T_i process (or, equivalently, the T_i^{new} process) behaves like the T_i' process in the case of large exceedences reflects the fact that, since the marginal distribution of a t-statistic is relatively light-tailed (if n is sufficiently large – see (2.3)), exceedences of a high level are rare and so are unlikely to occur together. The case of low-level exceedences is a very different matter, of course, and so we would expect the theorem to fail if the lower bound for t, in the first part of (2.5), were relaxed too far.

2.2. Applications

In this section, we treat the case of the null hypothesis, where $d_i = 0$ for each i. This would be assumed in most applications of Theorem 1 since it represents the setting that is conventionally used for calibration.

The theorem implies that, in a strong sense, exceedences down to those of the level $(1 + \eta)(2\gamma^{-1} \log p)^{1/2}$ are identical to the ones that would occur in the case of independent tests. Now, the probability associated with an exceedence of $(1 + \eta)(2\gamma^{-1} \log p)^{1/2}$ is, for small η, approximately $p^{-1/\gamma}$. Therefore, false discoveries at probability levels of approximately $p^{-1/\gamma}$, and at lower levels, can be adequately controlled by assuming that the tests are independent, even when they are not. Note that $\gamma^{-1} < 1$ and that the false-discovery level controlled by the conventional family-wise error rate is only p^{-1}.

Next, we discuss the sorts of calculations that are enabled by Theorem 1. Let Q_j denote the number of indices $i \in [1, p]$ for which T_i lies in the interval $(t_j, t_{j-1}]$, where $j \geq 1$ and t_j is determined by $P(T_1 > t_j) = j\beta/p$, with $\beta > 0$ held fixed. (We take $t_0 = \infty$.) If $T_1, \ldots, T_p$ were fixed, then the joint distribution of the $k + 1$ random variables $Q_1, \ldots, Q_k, p - \sum_j Q_j$ would be exactly multinomial with parameters p and $q_1, \ldots, q_k, 1 - \sum_j q_j$, where $q_j = P(T_1 \in (t_j, t_{j-1}])$. Theorem 1 implies that for the dependent process $T_1, \ldots, T_p$, and for any $k_0 = k_0(p)$ for which t_{k_0} satisfies (2.5), any simultaneous probability calculation based on the multinomial result, but applied to the actual T_j process rather than an idealized process with independent marginals, is valid, provided that $k \leq k_0$ and the final computed probability is quantified by adding an error which is stated to be of order $\exp\{o(t_k^2)\}\{\exp(-t_k^2/4) + p \exp(-\gamma t_k^2/2)\}$. The latter probability converges to zero, even if $k = k_0$ is taken as large as $p^{(\gamma_1 - 1)/\gamma_1}$, where $\gamma_1 \in (1, \gamma)$.

From this point, simultaneous multinomial probability calculations based on $Q_1, \ldots, Q_k$, familiar from the well-understood case of independent test statistics, can be used to construct rules for controlling FWER or false discovery rate (FDR); see, for example, Benjamini and Hochberg (1995). Wang and Hall (2009) have shown that, under the assumption of finite third

moments, highly accurate approximations are available for the marginal distribution of T_1. Such calculations, which justify standard normal, Student's t- or bootstrap approximations to the marginal distribution of T_1, are already widely used in practice (see Section 1), in conjunction with the independence assumption, when controlling false discovery rates. Our paper provides justification for these methods.

More generally, Theorem 1 implies that if a probability statement about what the process $T_1, \ldots, T_p$ does above the level t is founded on the assumption of independence, then, no matter how complex or convoluted the statement might be, the claimed probability level is accurate to within $\phi(t)$.

To give an example of calculations based on Theorem 1, take $p \le p_0 = 10^6$, $n = 100$ and $t = 5.052$, the latter denoting the upper $(1 - p_0^{-1})$-level quantile of Student's t-distribution with $n - 1 = 99$ degrees of freedom. Reflecting empirical evidence given in Section 2.1, take $\gamma = \frac{1}{4}(1 - 0.1) + 1 = 1.225$. Then, (2.5) is in order; the probability that at least one value of p independent t statistics, each on 99 degrees of freedom, exceeds $t = 5.052$ equals 0.010, 0.095 and 0.63 for $p = 10^4$, 10^5 and 10^6, respectively; and (2.6) suggests that the errors in these levels are in error by less than 30%, 20% and 0.25%, respectively. Most likely, the errors are much less than these since the asymptotic bound is derived only as an upper bound. If we were to make a general probability statement about exceedences of the level 5.052 by the stochastic process of t statistics, under the assumption of independence, then, despite the process actually being κ-dependent rather than independent, we would expect to make errors no greater than these respective values. In the same general setting, relative error decreases to zero as p and t increase. For example, in cases where t solves $1 - \Phi(t) \asymp p^{-1}$, with Φ denoting the standard normal distribution function, we have $\{\exp(-\frac{1}{4}t^2) + p\exp(-\frac{1}{2}\gamma t^2)\}/(1 - p)^p = O[\exp(-\frac{1}{4}t^2) + t\exp\{-\frac{1}{2}(\gamma - 1)t^2\}] \to 0$ as $t \to \infty$.

2.3. Generalizations

Theorem 1 can be extended to other settings, in particular, to those where (a) a wider range of dependence, obtained by allowing κ in (2.2) to diverge with p, is allowed; (b) the value of n for the ith group equals n_i, depending on i, and (2.3) is altered by requiring that $\log p = o(\min_{i \le p} n_i)$; (c) weights w_{ij} are incorporated into the construction of the t-statistics T_i, by defining $\bar{U}_i = n_i^{-1} \sum_j w_{ij} U_{ij}$, $S_i^2 = n_i^{-1} \sum_j w_{ij}^2 U_{ij}^2 - \bar{U}_i^2$ and, as before, $T_i = n_i^{1/2} \bar{U}_i / S_i$. Provided the weights satisfy

$$\sup_{i,j} |w_{ij}| \le C_1, \qquad \inf_{1 \le i \le p} n_i^{-1} \#\{j: |w_{ij}| \ge C_2\} \ge C_3,$$

where C_1, C_2, C_3 are positive constants not depending on p, the proof in this more general case is as in Section 3. However, the statement of the theorem is then less elegant and less transparent, so we do not give the more general version here. Incorporation of the weights w_{ij} permits the scope of the example above to be extended to hypothesis-testing problems involving linear regression.

To indicate the types of results that can be achieved under longer ranges of dependence, we shall discuss the case of a moving average,

$$U_{ij} = \kappa^{-1/2} \sum_{k=1}^{\kappa} \varepsilon_{j,i+k},$$

where $\kappa = \kappa(p)$ is permitted to diverge to infinity at a rate not exceeding $\log p$ and the independent disturbances ε_{ji} are all distributed as ε, for which $E(\varepsilon) = 0$ and $E|\varepsilon|^3 < \infty$. In this setting, (2.2) holds. We strengthen (2.3) by asking that $\log p = O(n^{1/3})$. The definition of t implicit in (2.5) can now be refined to

$$t = \sqrt{2\gamma^{-1}(\log p + A \log\log p)},$$

where $A > 0$ denotes a sufficiently large absolute constant. The conclusions of Theorem 1 continue to hold, with a similar proof if we replace $1 - \phi(t)$ by $1 - o(1)$.

3. Proof of Theorem 1

3.1. Step 1: Preliminaries

The notation $D_1, D_2, \ldots$ will denote constants not depending on n or p. Let $Q_i = n^{-1} \sum_j (U_{ij} - d_i)^2$, $R_i = n^{1/2} \bar{U}_i / Q_i^{1/2}$ and note that

for each $t > 0$, the events $T_i > t$ and $R_i > t/(1 + n^{-1}t^2)^{1/2}$ are identical. (3.1)

Also, note that $R_i = (\sum_j V_{ij} + n d_i)/(\sum_j V_{ij}^2)^{1/2}$, where $V_{ij} = U_{ij} - d_i$, $1 \leq j \leq n$, are independent and identically distributed random variables satisfying

$$0 \leq E(V_{i1}) = 0, \qquad E(V_{i1}^2) = 1 \qquad \text{for all } i, \qquad \sup_{i \geq 1} E(|V_{i1}|^3) < \infty; \qquad (3.2a)$$

cf. (2.1).

3.2. Step 2: Probabilities of exceedences in ones and twos

Using results of Wang and Hall (2009) (see also Wang (2005)), it can be shown that, for constants $D_1, D_2, D_3 > 0$, and whenever $0 < s < D_1 n^{1/2}$,

$$P(R_i > s) \leq D_2 s^{-1} \exp\left(D_3 s^3 n^{-1/2} - \tfrac{1}{2}s^2 + \sqrt{n}d_i s\right). \qquad (3.2)$$

We also wish to prove the following related result for pairs of exceedences.

Lemma. *Assume the conditions of Theorem* 1. *There then exist* $D_4, D_5 > 0$ *such that for all* i_1, i_2 *with* $i_1 \neq i_2$, *and for all* $0 < s < D_4 n^{1/2}$, *we have*

$$
\sup_{1 \le |i_1 - i_2| \le k_2 - k_1} P(R_{i_1} > s, R_{i_2} > s)
$$

$$
\le 5 \exp\left\{-\tfrac{1}{2}(1+\alpha)s^2 + D_5 n^{-1/2} s^3 + 2\sqrt{n}(d_{i_1} + d_{i_2})s\right\},
\tag{3.3}
$$

where α *is as in* (2.4).

To establish the lemma, we write

$$
U_{i_1 1}^{(1)} = V_{i_1 1} I(|V_{i_1 1}| \le n^{1/2}/s, |V_{i_2 1}| \le n^{1/2}/s), \qquad U_{i_1 1}^{(2)} = V_{i_1 1} - U_{i_1 1}^{(1)},
$$

$$
U_{i_2 1}^{(1)} = V_{i_2 1} I(|V_{i_1 1}| \le n^{1/2}/s, |V_{i_2 1}| \le n^{1/2}/s), \qquad U_{i_2 1}^{(2)} = V_{i_2 1} - U_{i_1 1}^{(1)}.
$$

By virtue of (3.2a), simple calculations show that

$$
\left|E\left(U_{i_1 1}^{(1)} + U_{i_2 1}^{(1)}\right)\right| \le D_6 s^2/n,
$$

$$
E\left\{\left(U_{i_1 1}^{(1)}\right)^2 + \left(U_{i_2 1}^{(1)}\right)^2\right\} = 2 + O(1)s/\sqrt{n},
$$

$$
E\left(U_{i_1 1}^{(1)} + U_{i_2 1}^{(1)}\right)^2 = E\left\{(V_{i_1 1} + V_{i_2 1}) - \left(U_{i_1 1}^{(2)} + U_{i_2 1}^{(2)}\right)\right\}^2
$$

$$
\le 2(1 + \rho_{ij}) + D_6 s/\sqrt{n},
$$

$$
E\left|U_{i_1 1}^{(1)} + U_{i_2 1}^{(1)}\right|^3 \le D_6.
$$

These results, and the bound $e^x \le 1 + x + \tfrac{1}{2}x^2 + \tfrac{1}{6}|x|^3 e^x$ (valid for all real x), imply that, with $h = s/\sqrt{n}$,

$$
E\left[\exp\left\{\tfrac{1}{2}h(V_{i_1 1} + V_{i_2 1}) - \tfrac{1}{4}h^2(V_{i_1 1}^2 + V_{i_2 1}^2)\right\} I(|V_{i_1 1}| \le n^{1/2}/s, |V_{i_2 1}| \le n^{1/2}/s)\right]
$$

$$
\le 1 + (\rho_{ij} - 1)\frac{s^2}{4n} + D_7 \frac{s^3}{n^{3/2}},
\tag{3.4}
$$

$$
E\left[\exp\left\{\tfrac{1}{2}h(V_{i_1 1} + V_{i_2 1}) - \tfrac{1}{4}h^2(V_{i_1 1}^2 + V_{i_2 1}^2)\right\} I(|V_{i_1 1}| \ge n^{1/2}/s, \text{ or } |V_{i_2 1}| \ge n^{1/2}/s)\right]
$$

$$
\le \exp\{P(|V_{i_1 1}| > n^{1/2}/s) + P(|V_{i_2 1}| > n^{1/2}/s)\}
\tag{3.5}
$$

$$
\le D_8 (s/\sqrt{n})^3.
$$

Results (3.4) and (3.5), together with the independence of V_{i_1k} for each i_1, imply that, for $s \leq D_9\sqrt{n}$, with D_9 sufficiently small,

$$E\left[\exp\left\{\frac{h}{2}\sum_k(V_{i_1k} + V_{i_2k}) - \frac{h^2}{4}\sum_k(V_{i_1k}^2 + V_{i_2k}^2)\right\}\right]$$

$$\leq \left\{1 + (\rho_{ij} - 1)\frac{s^2}{4n} + D_{10}\frac{s^2}{n^{3/2}}\right\}^n \leq \exp\left\{(\rho_{ij} - 1)\frac{s^2}{4} + D_{11}\frac{s^3}{\sqrt{n}}\right\}. \tag{3.6}$$

Define $\varepsilon^2 = (1 - \rho_{ij})/8$. It follows from (3.6) that whenever $s \leq D_9\sqrt{n}$, with D_9 sufficiently small, we have

$$\pi_{1n} \equiv P\left\{2h\sum_k(V_{i_1k} + V_{i_2k}) - h^2\sum_k(V_{i_1k}^2 + V_{i_2k}^2)\right.$$

$$\left. + 2\sqrt{n}(d_{i_1} + d_{i_2})s \geq 2s^2(1 - \varepsilon^2)\right\}$$

$$\leq \exp\left\{2\sqrt{n}(d_{i_1} + d_{i_2})s - \frac{1}{2}s^2(1 - \varepsilon^2)\right\} \tag{3.7}$$

$$\times E\left[\exp\left\{\frac{1}{2}h\sum_k(V_{i_1k} + V_{i_2k}) - \frac{1}{4}h^4\sum_k(V_{i_1k}^2 + V_{i_2k}^2)\right\}\right]$$

$$\leq \exp\left\{-\frac{1}{2}(1 + \alpha)s^2 + D_{12}\frac{s^3}{\sqrt{n}} + 2\sqrt{n}(d_{i_1} + d_{i_2})s\right\},$$

where α is as defined in (2.4). Write $\Omega_n = (1 - \varepsilon, 1 + \varepsilon)$ and note that if $0 < \varepsilon < \frac{1}{2}$, then

$$\left\{\sum_k V_{i_2k} + nd_{i_2} \geq s(nQ_{i_2}^2)^{1/2}, Q_{i_2} \in \Omega_n\right\}$$

$$\subseteq \left\{2h\sum_k V_{i_2k} - h^2\sum_k V_{i_2k}^2 + 2\sqrt{n}d_{i_2}s \geq s^2(1 - \varepsilon^2)\right\},$$

where $h = s/\sqrt{n}$. It can be shown that

$$P(R_{i_1} > s, R_{i_2} > s)$$

$$= P\left\{\sum_k V_{i_1k} + nd_{i_1} \geq s(nQ_{i_1}^2)^{1/2}, \sum_k V_{i_2k} + nd_{i_2} \geq s(nQ_{i_2}^2)^{1/2}\right\} \tag{3.8}$$

$$\leq \pi_{1n} + \pi_{2n} + \pi_{3n} + \pi_{4n} + \pi_{5n},$$

where

$$\pi_{2n} = P\left\{\sum_k V_{i_1k} + nd_{i_1} \geq s(nQ_{i_1}^2)^{1/2}, Q_{i_1} \geq 1 + \varepsilon\right\},$$

$$\pi_{3n} = P\left\{\sum_k V_{i_2k} + nd_{i_2} \geq s(nQ_{i_2}^2)^{1/2}, Q_{i_2} \geq 1 + \varepsilon\right\},$$

$$\pi_{4n} = P\left\{\sum_k V_{i_1k} + nd_{i_1} \geq s(nQ_{i_1}^2)^{1/2}, Q_{i_1} \leq 1 - \varepsilon\right\},$$

$$\pi_{5n} = P\left\{\sum_k V_{i_2k} + nd_{i_2} \geq s(nQ_{i_2}^2)^{1/2}, Q_{i_2} \leq 1 - \varepsilon\right\}.$$

Property (3.3) will follow from (3.7) and (3.8) if we prove that there exists $D_{13} > 0$ such that, for $s \leq D_{13}n^{1/2}$,

$$\pi_{kn} \leq \exp\left\{-\frac{1}{2}(1+\alpha)s^2 + D_{12}\frac{s^3}{\sqrt{n}} + 2\sqrt{n}(d_{i_1} + d_{i_2})s\right\} \tag{3.9}$$

for $k = 2, 3, 4, 5$.

Our proof of (3.9) is based on arguments of Shao (1999) (see also the proof of Proposition 4.2 of Wang and Hall (2009)) and uses the following result: if $EX = 0$, $EX^2 = 1$ and $E|X|^3 < \infty$, then for any $\lambda > 0$, $\theta > 0$ and $x > 0$,

$$E[\exp\{\lambda bX - \theta(bX)^2\}] = 1 + (\lambda^2 - \theta)n^{-1}x^2 + A(\lambda, \theta)n^{-3/2}x^3 E|X|^3, \tag{3.10}$$

where $b = x/\sqrt{n}$ and $A(\lambda, \theta)$ depends only on λ and θ. This result is a special case of Lemma 1 of Shao (1999). Also, note that

$$\pi_{2n} \leq \pi_{2n}^{(1)} + P\left\{\sum_k V_{i_1k} + nd_{i_1} \geq s(nQ_{i_1}^2)^{1/2}, Q_{i_1} \geq 3\right\}$$

$$\leq \pi_{2n}^{(1)} + \pi_{2n}^{(2)} + \pi_{2n}^{(3)},$$

where, noting that $\sqrt{n}|d_i| \leq s/5$, we define

$$\pi_{2n}^{(1)} = P\left\{\sum_k V_{i_1k} + nd_{i_1} \geq s(nQ_{i_1}^2)^{1/2}, 1 + \varepsilon \leq Q_{i_1} < 3\right\},$$

$$\pi_{2n}^{(2)} = P\left\{\sum_k V_{i_1k}I(|V_{i_1k}| > n^{1/2}/s) \geq s\left(\sum_k V_{i_1k}^2\right)^{1/2}\right\},$$

$$\pi_{2n}^{(3)} = P\left\{\sum_k V_{i_1k}I(|V_{i_1k}| \leq n^{1/2}/s) \geq 3s\sqrt{n}/2\right\}.$$

If the random variable H has the Bi(n, p) distribution and if $a > 0$, then $P(H > an) \le (ep/a)^{an}$ and so

$$\pi_{2n}^{(2)} \le P\left\{\sum_k I(|V_{i_1k}| > n^{1/2}/s) \ge s^2\right\}$$

$$\le \{s^{-2}12n P(|V_{i_1k}| > n^{1/2}/s)\}^{s^2} \le \frac{1}{2}e^{-s^2}$$

for $s \le D_{14}\sqrt{n}$, with D_{14} sufficiently small. Arguments similar to those in the proof of (3.7) yield that $\pi_{2n}^{(3)} \le \frac{1}{2}e^{-s^2}$ for $s \le D_{14}\sqrt{n}$ with D_{14} sufficiently small. To estimate $\pi_{2n}^{(1)}$, we write $S_1 = \{(x, y): x \ge s\sqrt{y}, s^2(1 + \varepsilon)^2 \le y \le 9s^9\}$. It follows from (3.10) with $\lambda = 1$, $\theta = \frac{1}{6}$ and $X = V_{i_11}$ that, with $h = s/\sqrt{n}$,

$$\pi_{2n}^{(1)} = P\left\{\left(h\sum_k V_{i_1k} + \sqrt{n}d_is, h^2\sum_k V_{i_1k}^2\right) \in S_1\right\}$$

$$\le E\left[\exp\left(h\sum_k V_{i_1k} - \frac{1}{6}h^2\sum_k V_{i_1k}^2 + \sqrt{n}d_is\right)\exp\left\{-\inf_{(x,y)\in S_1}(x - y/6)\right\}\right]$$

$$\le \exp\left\{\left(\frac{1}{2} - \frac{1}{6}\right)s^2 - s^2(1 + \varepsilon) + \frac{1}{6}s^2(1 + \varepsilon)^2 + \sqrt{n}d_is + D_{15}s^3n^{-1/2}\right\}$$

$$\le \exp\left\{-\frac{1}{2}s^2 - (5\varepsilon s^2/8) + \sqrt{n}d_is + \left(D_{15}s^3/\sqrt{n}\right)\right\}$$

$$\le \exp\left\{-\frac{1}{2}(1 + \alpha)s^2 + \sqrt{n}d_is + \left(D_{15}s^3/\sqrt{n}\right)\right\},$$

where we have used the fact that the function $f(y) = s\sqrt{y} - \frac{1}{6}y$ is increasing in $s^2(1 + \varepsilon)^2 \le y \le 9s^2$. Combining all of the above estimates, we obtain

$$\pi_{2n} \le \exp\left\{-\frac{1}{2}(1 + \alpha)s^2 + \sqrt{n}d_is + \left(D_{15}s^3/\sqrt{n}\right)\right\}.$$

Similarly, we may prove (3.9) for $k = 3$.

Put $S_2 = \{(x, y): x \ge s\sqrt{y}, y \le (1 - \varepsilon)^2 s^2\}$. It follows from (3.10) with $\lambda = 1$, $\theta = 2$ and $X = V_{i_11}$ that, with $h = s/\sqrt{n}$,

$$\pi_{4n} = P\left\{\left(h\sum_k V_{i_1k} + \sqrt{n}d_is, h^2\sum_k V_{i_1k}^2\right) \in S_2\right\}$$

$$\le E\left[\exp\left(h\sum_k V_{i_1k} - 2h^2\sum_k V_{i_1k}^2 + \sqrt{n}d_is\right)\exp\left\{-\inf_{(x,y)\in S_2}(x - 2y)\right\}\right]$$

$$\le \exp\left\{-1.5s^2 - s^2(1 - \varepsilon) + 2s^2(1 - \varepsilon)^2 + \sqrt{n}d_is + \left(D_{16}s^3/\sqrt{n}\right)\right\}$$

$$\leq \exp\left\{-\frac{1}{2}s^2 - 2\varepsilon s^2 + \sqrt{n}d_i s + \left(D_{16}s^3/\sqrt{n}\right)\right\}$$

$$\leq \exp\left\{-\frac{1}{2}(1+\alpha)s^2 + \sqrt{n}d_i s + \left(D_{16}s^3/\sqrt{n}\right)\right\}.$$

Similarly, we may prove (3.9) for $k = 5$. This completes the derivation of (3.9) and, hence, also the proof of the lemma.

3.3. Step 3: Blocks and expected numbers of level exceedences

Partition the set of positive integers into small blocks, each of length $\kappa + 1$, where κ is as in (2.2), and large blocks, each of length ℓ, where ℓ is a divergent function of p. We shall take

$$\ell \sim \exp\left(\tfrac{1}{4}s^2\right), \tag{3.11}$$

where $s \to \infty$ as p increases. The integers in each block are consecutive, each consecutive pair of large blocks is separated by a small block and the block furthest to the left is a large block. Let the small blocks be $b_1, b_2, \ldots$ and the large blocks be $B_1, B_2, \ldots$, indexed such that the order of the blocks is $B_1, b_1, B_2, b_2, \ldots$. Let $B = B_1 = \{1, \ldots, \ell\}$ denote the first large block and let N_1 be the number of indices $i \in B$ for which $R_i > s$. We wish to develop a bound for $E\{N_1 I(N_1 \geq 2)\}$. Identical bounds can be derived, uniformly in the block indices, for the versions of N_1 in the case of blocks $B_2, B_3, \ldots$; for notational simplicity, we focus solely on B_1.

By Hölder's inequality,

$$E\{N_1 I(N_1 \geq 2)\} \leq (E N_1^{a_1})^{1/a_1} P(N_1 \geq 2)^{1/a_2}, \tag{3.12}$$

where $a_1, a_2 > 1$ satisfy $a_1^{-1} + a_2^{-1} = 1$. Define $d^0 = \sqrt{n}\max_{1 \leq i \leq p} d_i$. In view of (3.2) and (3.3),

$$
\begin{aligned}
P(N_1 \geq 2) &= P(\text{for some } i_1, i_2 \in B \text{ with } i_1 < i_2, R_{i_1}, R_{i_2} > s) \\
&\leq \sum_{i_1=1}^{\ell-1} \sum_{i_2=i_1+1}^{\ell} P(R_{i_1} > s, R_{i_2} > s) \\
&= \sum_{i_1=1}^{\ell-1} \sum_{i_2=i_1+1}^{\min(i_1+\kappa+1,\ell)} P(R_{i_1} > s, R_{i_2} > s) \\
&\quad + \sum_{i_1=1}^{\ell-1} \sum_{i_2=\min(i_1+\kappa+2,\ell)}^{\ell} P(R_{i_1} > s) P(R_{i_2} > s) \\
&\leq D_{17} \exp(D_{18}s^3 n^{-1/2} + D_{19}d^0 s) \\
&\quad \times \left[\ell \exp\left\{-\frac{1}{2}(\alpha+1)s^2\right\} + \ell^2 \exp(-s^2)\right].
\end{aligned}
\tag{3.13}
$$

Noting that N_1 can be written as $\kappa + 1$ sums of $\ell/(\kappa + 1)$ independent and identically distributed random variables and using calculations based on the binomial distribution, it can be shown that, for the choice of ℓ at (3.11), $E(N_1^{a_1})$ is bounded as $p \to \infty$ for each $a_1 > 0$. Hence, using (3.12) and (3.13), we deduce that for each $\eta_2 \in (0, 1)$,

$$E\{N_1 I(N_1 \geq 2)\} \leq D_{20} \exp(D_{21} s^3 n^{-1/2} + D_{22} d^0 s)$$
$$\times \left[\ell \exp\{-\tfrac{1}{2}(\alpha + 1)s^2\} + \ell^2 \exp(-s^2)\right]^{1-\eta_2}. \tag{3.14}$$

Write N_2 for the number of exceedences of s that occur in the union of the small blocks b_j that intersect the interval $[1, p]$. There are $O(p/\ell)$ such small blocks and each is of length $\kappa + 1$, so, by (3.2),

$$E(N_2) \leq D_{23} p\ell^{-1} P(R_1 > s) \leq D_{24} p\ell^{-1} \exp\left(D_3 s^3 n^{-1/2} - \tfrac{1}{2}s^2 + d^0 s\right). \tag{3.15}$$

Provided we choose $s = s(p)$ to diverge to infinity in such a manner that

$$s = O\left(\sqrt{\log p}\right), \qquad d^0 = o(s), \tag{3.16}$$

it follows from (2.3) that $s^3 n^{-1/2} + d^0 s = o(s^2)$ and so (3.14) entails that

$$E\{N_1 I(N_1 \geq 2)\} = \exp\{o(s^2)\}\left[\ell \exp\{-\tfrac{1}{2}(\alpha + 1)s^2\} + \ell^2 \exp(-s^2)\right]^{1-\eta_2}.$$

Since this is true for each $\eta_2 > 0$, we have

$$E\{N_1 I(N_1 \geq 2)\} = \exp\{o(s^2)\}\left[\ell \exp\{-\tfrac{1}{2}(\alpha + 1)s^2\} + \ell^2 \exp(-s^2)\right]. \tag{3.17}$$

3.4. Step 4: Bound for $P(N_1 \geq 1)$, and related bounds

Let N_3 denote the number of exceedences of s which come from large blocks B_j, $1 \leq j \leq m$, that have two or more exceedences. Write $\sum_j \pi_j$ for the sum over $1 \leq j \leq m$ of the probability π_j that $R_i > s$ for some $i \in B_j$. Then (a) the expected number of exceedences of s by $R_1, \ldots, R_p$ equals $\sum_{i \leq p} P(R_i > s)$ and is less than or equal to $\sum_j \pi_j + E(N_2) + E(N_3)$; (b) the expected number of exceedences in (a) is greater than or equal to $\sum_{j \leq m-1} \pi_j$; and (c) since $P(N_1 \geq 1) \leq E(N_1) = \ell P(R_1 > s)$ and $P(R_1 > s)$ satisfies (3.2), we have

$$\pi_1 = P(N_1 \geq 1) \leq P^0 \equiv D_2 s^{-1} \ell \exp\left(D_3 s^3 n^{-1/2} - \tfrac{1}{2}s^2 + d^0 s\right) \tag{3.18}$$

and an identical bound holds for $\pi_1, \ldots, \pi_m$, in particular, (d) $\pi_m \leq P^0$. Results (a)–(d) imply that

$$\left|\sum_{j=1}^m \pi_j - \sum_{i=1}^p P(R_i > s)\right| \leq E(N_2) + E(N_3) + P^0. \tag{3.19}$$

Since $E(N_3) \leq m E\{N_1 I(N_1 \geq 2)\}$, $m = \mathrm{O}(p/\ell)$ and bounds for $E\{N_1 I(N_1 \geq 2)\}$, $E(N_2)$ and $P(N_1 \geq 1)$ are given by (3.17), (3.15) and (3.18), it follows that (3.19) entails, on taking ℓ as in (3.11),

$$\left| \sum_{j=1}^{m} \pi_j - \sum_{i=1}^{p} P(R_i > s) \right| = \exp\left\{ -\frac{1}{4}s^2 + \mathrm{o}(s^2) \right\} \left\{ 1 + p \exp\left(-\frac{1}{4}s^2 - \frac{1}{2}\alpha s^2 \right) \right\}. \quad (3.20)$$

3.5. Step 5: Probabilities of level exceedences

Let $\mathcal{F}$ denote the event that (a) there are no exceedences of s in any of the small blocks that are wholly contained within $[1, p]$; (b) in each of the large blocks that is wholly contained within $[1, p]$, there is at most one exceedence of s; and (c) there are no exceedences of s in any block fragment that overlaps the end point p. Write $\mathcal{G}$ for the complement of $\mathcal{F}$. Results (3.15), (3.17) and (3.18) imply that, with ℓ given by (3.11) and assuming that (3.16) holds,

$$P(\mathcal{G}) \leq \exp\left\{ -\tfrac{1}{4}s^2 + \mathrm{o}(s^2) \right\} \left\{ 1 + p \exp\left(-\tfrac{1}{4}s^2 - \tfrac{1}{2}\alpha s^2 \right) \right\}. \quad (3.21)$$

Therefore, in order for $P(\mathcal{G}) \to 0$, it is sufficient that for some $\eta_3 \in (0, 1)$ and all sufficiently large p, we have

$$(1 + \eta_3)\sqrt{2\gamma^{-1} \log p} \leq s = \mathrm{O}\left(\sqrt{\log p} \right), \quad (3.22)$$

where γ is as defined in Section 2. This choice of s satisfies (3.16) and so if s is given by (3.22), then $P(\mathcal{G})$ satisfies (3.21).

3.6. Step 6: Strong approximation

Let M_j, $1 \leq j \leq m$, be the number of times that $R_i > s$ for $i \in B_j$. Then, the number, N, say, of blocks B_j for which $M_j \geq 1$ is distributed as $\sum_j I_j$, where the random variables I_j are independent, $I_j = 1$ if $M_j \geq 1$ and $I_j = 0$ otherwise. As before, we define $\pi_j = P(M_j \geq 1)$. Conditional on N and on the events "$M_{j_1} \geq 1$" and "$M_{j_2} \geq 1$," where $1 \leq j_1 < j_2 \leq m$, the sequences $\{R_i : i \in B_{j_1}\}$ and $\{R_i : i \in B_{j_2}\}$ are independent.

Order the blocks B_j for which $M_j \geq 1$, giving $B_{J_1}, \ldots, B_{J_N}$, where $1 \leq J_1 < \cdots < J_N \leq m$, and let W_k denote a value of R_i for which $R_i > s$, randomly chosen among such values for which $i \in B_{J_k}$. Write $i = I_k$ for the index of the value of R_i that is chosen as W_k. Then, conditional on N, the random variables $W_1, \ldots, W_N$ are independent and identically distributed as $R(s)$, $J_1, \ldots, J_N$ is a set of integers chosen independently and randomly from $1, \ldots, m$ and I_k is uniformly distributed among indices in B_{J_k}.

Let $R'_1, \ldots, R'_p$ be independent random variables having the distributions of $R_1, \ldots, R_p$, respectively, let M'_j denote the number of times that R'_i exceeds s for $i \in B_j$ and put

$\pi'_j = P(M'_j \geq 1)$. The numbers N' of blocks B_j for which $M'_j \geq 1$ are distributed as $\sum_j I'_j$, where the random variables I'_j are independent and $I'_j = 1$ if $M'_j \geq 1$, $I'_j = 0$ otherwise. An argument similar to, but simpler than, that leading to (3.20) shows that

$$\sum_{j=1}^m |\pi_j - \pi'_j| \leq \exp\left\{-\frac{1}{4}s^2 + o(s^2)\right\}\left\{1 + p\exp\left(-\frac{1}{4}s^2 - \frac{1}{2}\alpha s^2\right)\right\}. \qquad (3.23)$$

By enlarging the probability space if necessary, we can think of N as denoting the number out of m independent and random variables $U_1, \ldots, U_j$, each uniformly distributed on $[0, 1]$, which lie in the respective intervals $[0, \pi_j]$. Take N' to be the number of U_i's that lie in $[0, \pi'_j]$. Then,

$$P(N = N') \geq 1 - \sum_{j=1}^m |\pi_j - \pi'_j|. \qquad (3.24)$$

We have already constructed sequences $W_1, \ldots, W_N$, $I_1, \ldots, I_N$ and $J_1, \ldots, J_N$. If $N' > N$, then, conditional on these quantities and on N and N', we select new values $W_{N+1}, \ldots, W_{N'}$, $I_{N+1}, \ldots, I_{N'}$ and $J_{N+1}, \ldots, J_{N'}$ which are independent of $W_1, \ldots, W_N$, $I_1, \ldots, I_N$ and $J_1, \ldots, J_N$, with $W_{N+1}, \ldots, W_{N'}$ independently distributed as $R(s)$, the values of $J_{N+1}, \ldots, J_{N'}$ independently and uniformly distributed among $\{1, \ldots, m\} \setminus \{J_1, \ldots, J_N\}$ and the values of $I_{N+1}, \ldots, I_{N'}$ uniformly distributed within the blocks $B_{J_{N+1}}, \ldots, B_{J_{N'}}$, respectively. In this instance, we take $W'_1, \ldots, W'_{N'}$ and $I'_1, \ldots, I'_{N'}$ to be identical to $W_1, \ldots, W_{N'}$ and $I_1, \ldots, I_{N'}$, respectively. If $N' < N$, then we take $(W'_1, J'_1), \ldots, (W'_{N'}, J'_{N'})$ to be the (exceedence, block index) pairs that remain after randomly and independently deleting $N - N'$ pairs from the sequence $(W_1, J_1), \ldots, (W_N, J_N)$.

Let N_0 denote the number of exceedences of s by $R'_1, \ldots, R'_p$ and let N' represent the number of large blocks B_j in which there is at least one exceedence of s by the sequence $R'_1, \ldots, R'_p$. Then, $P(N_0 \geq N') = 1$. Conditional on N_0 and N', let $W'_{N'+1}, \ldots, W'_{N_0}$ denote independent and identically random variables, all distributed as $R(s)$, and distribute the locations $I'_{N'+1}, \ldots, I'_{N_0}$ of these exceedences independently and uniformly over the points $\{1, \ldots, p\} \setminus \{I'_1, \ldots, I'_{N'}\}$, conditional on all of the variables N', N_0, $W'_1, \ldots, W'_{N'}$ and $J'_1, \ldots, J'_{N'}$. Take the values of $R'_1, \ldots, R'_p$ that exceed s to be the variables $W'_1, \ldots, W'_{N_0}$ and let the locations of those exceedences be the points $I'_1, \ldots, I'_{N_0}$. By construction, $W'_1, \ldots, W'_{N_0}$ are distributed as the exceedences of s by p independent and identically distributed random variables distributed as $R(s)$; conjointly, $I'_1, \ldots, I'_{N_0}$ are distributed as the locations of those exceedences and the probability that $N_0 = N' = N$, $M_j \in \{0, 1\}$ for each $j \in [1, m]$ and there are no exceedences of s in any of the small blocks b_j for any $j \in [1, m]$ is bounded below by $1 - \tau(s)$, where $\tau(s)$ satisfies (2.6); see also (3.20), (3.21), (3.23) and (3.24).

Hence, provided that s satisfies (3.22), we may construct a sequence $R'_1, \ldots, R'_p$ of independent variables with the same marginal distribution as R_1 and such that, with probability bounded below by $1 - \tau(s)$, the exceedences of $R_1, \ldots, R_p$ over s are identical to those of $R'_1, \ldots, R'_p$. The theorem follows from this property, (2.3) and (3.1), on taking $s = t$.

Acknowledgments

References

Almirantis, Y. and Provata, A. (1999). Long- and short-range correlations in genome organization. *J. Statist. Phys.* **97** 233–262.

Amaratunga, D. and Cabrera, J. (2004). *Exploration and Analysis of DNA Microarray and Protein Array Data*. New York: Wiley.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. Roy. Statist. Soc. Ser. B* **57** 289–300. MR1325392

Benjamini, Y. and Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Ann. Statist.* **29** 1165–1188. MR1869245

Benjamini, Y. and Yekutieli, D. (2005). False discovery rate-adjusted multiple confidence intervals for selected parameters (with discussion). *J. Amer. Statist. Assoc.* **100** 71–93. MR2156820

Bernhard, G., Klein, M. and Hommel, G. (2004). Global and multiple test procedures using ordered *p*-values – a review. *Statist. Papers* **45** 1–14. MR2019782

Blair, R.C., Troendle, J.F. and Beck, R.W. (1996). Control of familywise errors in multiple endpoint assessments via stepwise permutation tests. *Statist. Med.* **15** 1107–1121.

Clarke, S. and Hall, P. (2009). Robustness of multiple testing procedures against dependence. *Ann. Statist.* **37** 332–358.

Cui, X. and Churchill, G.A. (2003). Statistical tests for differential expression in cDNA microarray experiments. *Genome Biol.* **4** 210 (electronic).

Dudoit, S., Shaffer, J.P. and Boldrick, J.C. (2003). Multiple hypothesis testing in microarray experiments. *Statist. Sci.* **18** 73–103. MR1997066

Efron, B. (2007a). Size, power and false discovery rates. *Ann. Statist.* **35** 1351–1377. MR2351089

Efron, B. (2007b). Correlation and large-scale simultaneous significance testing. *J. Amer. Statist. Assoc.* **102** 93–103. MR2293302

Efron, B. and Tibshirani, R. (2002). Empirical Bayes methods and false discovery rates for microarrays. *Genetic Epidemiol.* **23** 70–86.

Falk, M. and Reiss, R.-D. (2001). Estimation of canonical dependence parameters in a class of bivariate peaks-over-threshold models. *Statist. Probab. Lett.* **52** 233–242. MR1838210

Fox, R.J. and Dimmic, M.W. (2006). A two-sample Bayesian *t*-test for microarray data. *BMC Bioinformatics* **7** 126 (electronic).

Hochberg, Y. (1988). A sharper Bonferroni procedure for multiple tests of significance. *Biometrika* **75** 800–802. MR0995126

Hochberg, Y. and Benjamini, Y. (1990). More powerful procedures for multiple testing. *Statist. Med.* **9** 811–818.

Hochberg, Y. and Tamhane, A.C. (1987). *Multiple Comparison Procedures*. New York: Wiley. MR0914493

Holland, B. and Cheung, S.H. (2002). Familywise robustness criteria for multiple-comparison procedures. *J. Roy. Statist. Soc. Ser. B* **64** 63–77. MR1881845

Hommel, G. (1988). A comparison of two modified Bonferroni procedures. *Biometrika* **76** 624–625. MR1040659

Hu, Z. and Willsky, G.R. (2006). Utilization of two sample *T*-test statistics from redundant probe sets to evaluate different probe set algorithms in GeneChip studies. *BMC Bioinformatics* **7** 12 (electronic).

Lehmann, E.L. and Romano, J.P. (2005). *Testing Statistical Hypotheses*, 3rd ed. New York: Springer. MR2135927

Lehmann, E.L., Romano, J.P. and Shaffer, J.P. (2005). On optimality of stepdown and stepup multiple test procedures. *Ann. Statist.* **33** 1084–1108. MR2195629

Li, J. (2006). Modelling dependency between different lines of business with copulas. Research Paper 146, Centre for Actuarial Studies, Department of Economics, The University of Melbourne.

Liu, P. and Hwang, J.T.G. (2007). Quick calculation for sample size while controlling false discovery rate with application to microarray analysis. *Bioinformatics* **23** 739–746.

Mansilla, R., De Castillo, N., Govezensky, T., Miramontes, P., José, M. and Coho, G. (2004). Long-range correlation in the whole human genome. Available at http://arxiv.org/pdf/q-bio/0402043v1.

Messer, P.W. and Arndt, P.F. (2006). CorGen-measuring and generating long-range correlations for DNA sequence analysis. *Nucleic Acids Research* **34** W692–W695.

Pawluk-Kolc, M., Zieba-Palus, J. and Parczewski, A. (2006). Application of false discovery rate procedure to pairwise comparisons of refractive index of glass fragments. *Forensic Sci. Internat.* **160** 53–58.

Perneger, T.V. (1998). What's wrong with Bonferroni adjustments. *BMJ* **316** 1236–1238.

Pigeot, I. (2000). Basic concepts of multiple tests – A survey. *Statist. Papers* **41** 3–36. MR1746085

Qiu, X. and Yakovlev, A. (2006). Some comments on instability of false discovery rate estimation. *J. Bioinform. Comput. Biol.* **4** 1057–1068.

Sarkar, S.K. (1998). Some probability inequalities for ordered MTP$_2$ random variables: A proof of the Simes conjecture. *Ann. Statist.* **26** 494–504. MR1626047

Sarkar, S.K. (2002). Some results on false discovery rate in stepwise multiple testing procedures. *Ann. Statist.* **30** 239–257. MR1892663

Sarkar, S.K. and Chang, C.K. (1997). The Simes method for multiple hypothesis testing with positively dependent test statistics. *J. Amer. Statist. Assoc.* **92** 1601–1608. MR1615269

Scheid, S. and Spang, R. (2005). Twilight; a Bioconductor package for estimating the local false discovery rate. *Bioinformatics* **21** 2921–2922.

Schmidt, R. (2002). Tail dependence for elliptically contoured distributions. *Math. Methods Oper. Res.* **55** 301–327. MR1919580

Schmidt, R. and Stadtmüller, U. (2006). Nonparametric estimation of tail dependence. *Scand. J. Statist.* **33** 307–335. MR2279645

Schmidt, T. (2007). Coping with copulas. In *Copulas – From Theory to Applications in Finance* (J. Rank, ed.). Risk Books.

Sen, P.K. (1999). Some remarks on Simes-type multiple tests of significance. *J. Statist. Plann. Inference* **82** 139–145. MR1736438

Shaffer, J.P. (2005). Comment on "False discovery rate-adjusted multiple confidence intervals for selected parameters" by Y. Benjamini and D. Yekutieli. *J. Amer. Statist. Assoc.* **100** 82–84. MR2156820

Shao, Q.-M. (1999). A Cramér type large deviation result for Student's t-statistic. *J. Theoret. Probab.* **12** 385–398.

Simes, R.J. (1986). An improved Bonferroni procedure for multiple tests of significance. *Biometrika* **73** 751–754. MR0897872

van de Wiel, M.A. and Kim, K.I. (2007). Estimating the false discovery rate using nonparametric deconvolution. *Biometrics* **63** 806–815. MR2395718

Wang, Q. (2005). Limit Theorems for self-normalized large deviation. *Electronic J. Probab.* **10** 1260–1285. MR2176384

Wang, Q. and Hall, P. (2009). Relative errors in central limit theorem for Student's t statistic, with applications. *Statist. Sinica* **19** 343–354. MR2487894

Wu, G. (1994). An empirical test of ordinal independence. *J. Risk and Uncertainty* **9** 39–60.