Berry Esseen Bounds for Combinatorial Central Limit Theorems and Pattern Occurrences, using Zero and Size Biasing * †

Larry Goldstein
University of Southern California

August 13, 2005

Abstract

Berry Esseen type bounds to the normal, based on zero- and size-bias couplings, are derived using Stein's method. The zero biasing bounds are illustrated with an application to combinatorial central limit theorems where the random permutation has either the uniform distribution or one which is constant over permutations with the same cycle type and having no fixed points. The size biasing bounds are applied to the occurrences of fixed relatively ordered sub-sequences (such as rising sequences) in a random permutation, and to the occurrences of patterns, extreme values, and subgraphs on finite graphs.

1 Introduction

Berry Esseen type bounds for normal approximation are developed using Stein's method, based on zero and size bias couplings. The results are applied to bound the proximity to the normal in combinatorial central limit theorems where the random permutation has either a uniform distribution, or one which is constant over permutations with the same cycle type, with no fixed points; to counting the number of occurrences of fixed, relatively ordered sub-sequences, such as rising sequences, in a random permutation; and to counting on finite graphs the number of occurrences of patterns, local extremes, and subgraphs.

Stein's method ([26], [28]) uses characterizing equations to obtain bounds on the error when approximating distributions by a given target. For the normal [27], $X \sim \mathcal{N}(\mu, \sigma^2)$ if and only if

$$\mathbb{E}(X-\mu)f(X) = \sigma^2 \mathbb{E} f'(X) \tag{1}$$

for all absolutely continuous f for which $\mathbb{E}|f'(X)| < \infty$. From such a characterizing equation, a difference or differential equation can be set up to bound the difference between the

*AMS 2000 subject classifications. Primary 60F05, 60C05 .
†Key words and phrases: smoothing inequality, Stein's method, permutations, graphs

expectation of a test function h evaluated on a given variable Y, and on the variable X having the target distribution. For the normal, with X having the same mean μ and variance σ^2 as Y, the characterizing equation (1) leads to the differential equation

$$h((y-\mu)/\sigma) - Nh = \sigma^2 f'(y) - (y-\mu)f(y), \tag{2}$$

where $Nh = \mathbb{E}h(Z)$ with $Z \sim \mathcal{N}(0,1)$, the standard normal mean of the test function h. At Y, the expectation of the left hand side can be evaluated by calculating the expectation of the right hand side using the bounded solution f of (2) for the given h. By this device, Stein's method can handle various kinds of dependence through the use of coupling constructions.

We consider and compare two couplings of a given Y to achieve normal bounds. First, for Y with mean zero and variance $\sigma^2 \in (0,\infty)$, we say that Y^* has the Y-zero biased distribution if

$$\mathbb{E}Yf(Y) = \sigma^2 \mathbb{E}f'(Y^*) \tag{3}$$

for all absolutely continuous functions f for which the expectation of either side exists. This 'zero bias transformation' from Y to Y^* was introduced in [15], and it was shown there that Y^* exists for every mean zero Y with finite variance. Similarly, for Y non-negative with finite mean $\mathbb{E}Y = \mu$, we say that Y^s has the Y-size biased distribution if

$$\mathbb{E}Yf(Y) = \mu \mathbb{E}f(Y^s) \tag{4}$$

for all f for which the expectation of either side exists. The size biased distribution exists for any non-negative Y with finite mean, and was used for normal approximation in [17].

A coupling (Y, Y^*) where Y^* has the Y-zero biased distribution lends itself for use in the Stein equation (2) in the following way; by (3), with $\sigma^2 = 1$ say, we have

$$\mathbb{E}h(Y) - Nh = \mathbb{E}\left[f'(Y) - Yf(Y)\right] = \mathbb{E}\left[f'(Y) - f'(Y^*)\right]. \tag{5}$$

Therefore, the difference between Y and the normal, as tested on h, equals the difference between Y and Y^*, as tested on f'. Additionally, as observed in [15] and seen directly from (5), Y is normal if and only if $Y =_d Y^*$. It is therefore natural that the distance from Y to the normal can be expressed in terms of distance from Y to Y^*. Theorem 1.1 makes this statement precise, showing that the distance from the standardized Y to the normal as measured by δ in (6) depends on the distribution of Y only through a bound on $|Y^* - Y|$. A similar phenomenon is seen in [14] with d_{W} the Wasserstein distance, where it is shown that, for any mean zero variance σ^2 variable Y, and $X \sim \mathcal{N}(0,\sigma^2)$,

$$d_{\mathrm{W}}(Y,X) \le 2d_{\mathrm{W}}(Y,Y^*).$$

The use of size bias couplings in the Stein equation in (67), (68) and subsequent calculations depends on the following identity, which is applied in a less direct manner than (5); for $Y \ge 0$ with mean μ and variance σ^2,

$$\mathbb{E}(Y-\mu)f(Y) = \mu\mathbb{E}\left(f(Y^s) - f(Y)\right) \quad \text{and therefore} \quad \sigma^2 = \mu\mathbb{E}(Y^s - Y).$$

With $W = (Y-\mu)/\sigma$, many authors (e.g. [7], [19], [8], [24], [25], [10]) have been successful in obtaining bounds on the distance

$$\delta = \sup_{h \in \mathcal{H}} |\mathbb{E}h(W) - Nh| \tag{6}$$

to the normal, over classes of non-smooth functions $\mathcal{H}$, using Stein's method. Here we take the smoothing inequality approach, following [25]. In particular, $\mathcal{H}$ is a class of measurable functions on the real line such that

(i) The functions $h \in \mathcal{H}$ are uniformly bounded in absolute value by a constant, which we take to be 1 without loss of generality,

(ii) For any real numbers c and d, and for any $h(x) \in \mathcal{H}$, the function $h(cx+d) \in \mathcal{H}$,

(iii) For any $\epsilon > 0$ and $h \in \mathcal{H}$, the functions h_ϵ^+, h_ϵ^- are also in $\mathcal{H}$, and

$$\mathbf{E}\tilde{h}_\epsilon(Z) \le a\epsilon \tag{7}$$

for some constant a which depends only on the class $\mathcal{H}$, where

$$h_\epsilon^+(x) = \sup_{|y|\le\epsilon} h(x+y), \quad h_\epsilon^-(x) = \inf_{|y|\le\epsilon} h(x+y), \quad \text{and} \quad \tilde{h}_\epsilon(x) = h_\epsilon^+(x) - h_\epsilon^-(x). \tag{8}$$

The collection of indicators of all half lines, and indicators of all intervals, for example, each form classes $\mathcal{H}$ which satisfy (8) and (7) with $a = \sqrt{2/\pi}$ and $a = 2\sqrt{2/\pi}$ respectively (see e.g. [25]).

Since the bound on δ in Theorem 1.1 depends only the size of $|Y^* - Y|$, it may be computed without the need for the calculation of the variances of certain conditional expectations that arise in other versions of Stein's method, $\sqrt{\mathrm{Var}\{\mathbf{E}[(Y''-Y')^2|Y']\}}$ for the exchangeable pair method, or the term (13) for the size bias coupling studied here.

Theorem 1.1 *Let Y be a mean zero random variable with variance $\sigma^2 \in (0,\infty)$, and Y^* be defined on the same space having the Y-zero biased distribution. If $|Y^* - Y| \le 2B$ for some $B \le \sigma/24$, then for δ as in (6) and a as in (7),*

$$\delta \le A\,(37 + 12A + 112a)\,, \tag{9}$$

for $A = 2B/\sigma$. For indicators of all half lines, and the indicators of all intervals, using $a = \sqrt{2/\pi}$ and $a = 2\sqrt{2/\pi}$, we have respectively

$$\delta \le A\,(127 + 12A) \quad \text{and} \quad \delta \le A\,(216 + 12A)\,. \tag{10}$$

See (75) and (76) for some variations on the bound (9) here, and (12) below, respectively. We note that Theorem 1.1 immediately provides a bound on δ of order σ^{-1} whenever $|Y^* - Y|$ is bounded. In Section 2, we apply Theorem 1.1 to random variables of the form

$$Y = \sum_{i=1}^n a_{i,\pi(i)}, \tag{11}$$

depending on a fixed array of real numbers $\{a_{ij}\}_{i,j=1}^n$ and a random permutation $\pi \in \mathcal{S}_n$, the symmetric group. In Section 2.1 we consider π having the uniform distribution on $\mathcal{S}_n$, and in Section 2.2 distributions constant on cycle type having no fixed points (conditions (25) and (27) respectively).

For a size bias coupling (Y, Y^s), Theorem 1.2 gives a bound on δ which depends on the size of $|Y^s - Y|$, and additionally on Δ in (13). While Δ may be difficult to calculate precisely in many cases, size bias couplings can be more easily constructed for a broader range of examples than the zero bias couplings.

Theorem 1.2 *Let $Y \geq 0$ be a random variable with mean μ and variance $\sigma^2 \in (0, \infty)$, and let Y^s be defined on the same space, with the Y-size biased distribution. If $|Y^s - Y| \leq B$ for some $B \leq \sigma^{3/2}/\sqrt{6\mu}$, then for δ as in (6) and a as in (7),*

$$\delta \leq \frac{aA}{2} + \frac{\mu}{\sigma}\left((19 + 56a)A^2 + 4A^3\right) + \frac{23\mu\Delta}{\sigma^2}, \tag{12}$$

where

$$\Delta = \sqrt{Var(\mathbf{E}(Y^s - Y|Y))} \tag{13}$$

and $A = B/\sigma$. For indicator functions of all half lines and the indicators functions of all intervals, by using $a = \sqrt{2/\pi}$ and $a = 2\sqrt{2/\pi}$, we respectively find that

$$\delta \leq 0.4A + \frac{\mu}{\sigma}\left(64A^2 + 4A^3\right) + \frac{23\Delta\mu}{\sigma^2} \quad \text{and} \quad \delta \leq 0.8A + \frac{\mu}{\sigma}\left(109A^2 + 4A^3\right) + \frac{23\Delta\mu}{\sigma^2}.$$

If the mean μ is of order σ^2, B is bounded and $\Delta = \sigma^{-1}$, then δ will have order $O(\sigma^{-1})$. The application of Theorem 1.2 to counting the occurrences of fixed relatively ordered subsequences, such as rising sequences, in a random permutation, and to counting the occurrences of color patterns, local maxima, and sub-graphs in finite graphs is illustrated in Section 3. The proofs of Theorems 1.2 and 1.1 are given in Section 4.

Nothing should be inferred from the fact that the zero bias applications presented here involve global dependence, and that the dependence in the examples used to illustrate the size bias approach is local; the exchangeable pair coupling on which our zero biased constructions are based can also be applied in cases of local dependence, and the size bias approach was applied in [17] to variables having global dependence.

In both zero and size biasing, a sum $Y = \sum_{\alpha \in \mathcal{A}} X_\alpha$ of independent variables on a finite index set $\mathcal{A}$ is biased by choosing a summand at random and replacing it with its biased version. To describe the zero biasing coupling, let $\{X_\alpha\}_{\alpha \in \mathcal{A}}$ be a collection of mean zero variables with finite variance, and I an independent random index with distribution

$$P(I = \alpha) = \frac{w_\alpha}{\sum_{\beta \in \mathcal{A}} w_\beta}, \tag{14}$$

where $w_\alpha = \text{Var}(X_\alpha)$. It was shown in [15] that replacing X_I by a variable X_I^* having the X_I-zero bias distribution, independent of $\{X_\alpha, \alpha \neq I\}$, gives

$$Y^* = Y - X_I + X_I^*, \tag{15}$$

a variable having the Y-zero biased distribution. Hence, when a sum of many independent variables of the same order is coupled this way to its zero biased version, the magnitude of $(Y^* - Y)/\sqrt{\text{Var}(Y)}$, and therefore of distance measures such as δ, are small.

The construction of the size biased coupling in the independent case is similar. Let $\{X_\alpha\}_{\alpha \in \mathcal{A}}$ be a collection of non-negative variables with finite mean. Then, with I a random index independent of all others variables, having distribution (14) with $w_\alpha = \mathbf{E}\, X_\alpha$, the replacement of X_I by a variable X_I^s with the X_I-size bias distribution, independent of the remaining variables, gives a variable with the Y-size biased distribution.

Zero biased couplings of Y^* to a sum Y of non-independent variables $X_1, \ldots, X_n$ is presently not very well understood. A construction in the presence of the weak global dependence of simple random sampling was given in [15]. Based on a remark in [15], we here exploit a connection between the zero bias coupling and the exchangeable pair (Y', Y'') of [28] with distribution $dP(y', y'')$ satisfying $\mathbb{E}(Y''|Y') = (1-\lambda)Y'$ for some $\lambda \in (0,1]$; in particular, we make use of a pair $(Y^\dagger, Y^\ddagger)$ with distribution proportional to $(y'-y'')^2 dP(y', y'')$.

The construction of Y and Y^s on a common space for the sum of non-independent variables $X_1, \ldots, X_n$ is more direct, and was described in Lemma 2.1 of [17]; we choose a summand with probability proportional to its expectation, replace it by one from its size-biased distribution, and then adjust the remaining variables according to the conditional distribution given the value of the newly chosen variable. This construction is applied in Section 3, and a 'squared' zero biasing form of it in Section 2.

The mappings of a distribution Y to its zero biased Y^* or size biased Y^s versions are special cases of distributional transformations from Y to some $Y^{(m)}$ which are specified by a function H and characterizing equation

$$\mathbb{E} H(Y) f(Y) = \eta \mathbb{E} f^{(m)}(Y^{(m)}) \quad \text{for all smooth } f,$$

where $f^{(m)}$ denotes the m^{th} derivative of f, and η is, necessarily, $(m!)^{-1}\mathbb{E} H(Y) Y^m$ when this expectation exists. The zero bias and size bias transformation correspond to $m=1$ and $H(x) = x$, and $m = 0$ and $H(x) = x^+$, respectively. In general, such a $Y^{(m)}$ exists when H and Y satisfy certain sign change and orthogonality properties, as discussed in [16].

2 Zero Biasing: Combinatorial Central Limit Theorems

In this section, we illustrate the use of Theorem 1.1 to obtain Berry Esseen bounds in combinatorial central limit theorems, that is, for variables Y as in (11), in Section 2.1 we do so for permutations having the uniform distribution over the symmetric group and, in Section 2.2, we do so for permutations with distribution constant on those having the same cycle type, with no fixed points. First we present Proposition 2.1, which suggests a method for the construction of zero bias couplings based on the existence of exchangeable pairs; its statement appears in [15].

Proposition 2.1 *Let Y' and Y'' be an exchangeable pair, with distribution $dP(y', y'')$ and $Var(Y') = \sigma^2 \in (0, \infty)$, which satisfies the linearity condition*

$$\mathbb{E}(Y''|Y') = (1-\lambda)Y' \quad \textit{for some } \lambda \in (0,1]. \tag{16}$$

Then

$$\mathbb{E} Y' = 0 \quad \textit{and} \quad \mathbb{E}(Y' - Y'')^2 = 2\lambda\sigma^2, \tag{17}$$

and if $Y^\dagger$ and $Y^\ddagger$ have distribution

$$dP^\dagger(y', y'') = \frac{(y'-y'')^2}{\mathbb{E}(Y'-Y'')^2} dP(y', y''), \tag{18}$$

and $U \sim \mathcal{U}[0,1]$ is independent of $Y^\dagger$ and $Y^\ddagger$, then the variable

$$Y^* = UY^\dagger + (1-U)Y^\ddagger \quad \textit{has the } Y' \textit{ zero biased distribution.} \tag{19}$$

Proof: The claims in (17) follow from (16) and exchangeability. Hence we need only show that Y^* in (19) satisfies (3). For a differentiable test function f,

$$\begin{aligned}\sigma^2 \mathbb{E} f'(UY^\dagger + (1-U)Y^\ddagger) &= \sigma^2 \mathbb{E}\left(\frac{f(Y^\dagger) - f(Y^\ddagger)}{Y^\dagger - Y^\ddagger}\right) \\ &= \sigma^2 \mathbb{E}\left(\frac{(Y' - Y'')(f(Y') - f(Y''))}{\mathbb{E}(Y' - Y'')^2}\right).\end{aligned}$$

Now if we use (16) to obtain $\mathbb{E}\, Y'' f(Y') = (1-\lambda)\mathbb{E}\, Y' f(Y')$, followed by (17), expanding yields

$$\sigma^2 \mathbb{E}\left(\frac{Y'f(Y') - Y''f(Y') - Y'f(Y'') + Y''f(Y'')}{\mathbb{E}(Y' - Y'')^2}\right) = \frac{2\lambda\sigma^2 \mathbb{E}\, Y' f(Y')}{\mathbb{E}(Y' - Y'')^2} = \mathbb{E}\, Y' f(Y'). \blacksquare$$

Example 2.2 Given a mean zero finite variance Y', let Y'' be an independent copy of Y'. The pair (Y', Y'') satisfies the conditions of Proposition 2.1 with $\lambda = 1$, and hence, Y^* as in (19) has the Y' zero bias distribution with $(Y^\dagger, Y^\dagger)$ as in (18). However, by coupling Y' close to Y'', so that $Y^\dagger$ is close to $Y^\ddagger$, causes B, and therefore, the bound δ of Theorem 1.1 to be small.

Remark 2.3 *The following construction of $(Y^\dagger, Y^\ddagger)$ suggested by Proposition 2.1 is similar to the one used for size biasing (see Lemma 2.1 of [17] and Section 3). Given Y', first construct an exchangeable Y'' close to Y' satisfying (16), and then, independently construct the variables appearing in the 'square biased' term $(Y' - Y'')^2$. Lastly, adjust the remaining variables that make up (Y', Y'') to have their original conditional distribution, given the newly generated variables.*

Example 2.4 Let $\{X_i', X_i''\}_{i=1,\ldots,n}$ be i.i.d. mean zero variables with finite variancess, let $Y' = \sum_{i=1}^n X_i'$, and let I be an independent random index with uniform distribution over $\{1, \ldots, n\}$. Letting $Y'' = Y' - X_I' + X_I''$, the pair (Y', Y'') is exchangeable and satisfies the conditions of Proposition 2.1, with $\lambda = 1/n$. Set $S = \sum_{i \neq I} X_i'$ and $(T', T'') = (X_I', X_I'')$. Applying Example 2.2 to (T', T''), and forming $(T^\dagger, T^\ddagger)$ independently of $\{X_i', X_i''\}_{i \neq I}$, gives $UT^\dagger + (1-U)T^\ddagger = X_I^*$. By their independence from X_I', X_I'', $\{X_i', X_i''\}_{i \neq I}$ already have their original conditional distribution, given $(T^\dagger, T^\ddagger)$; hence $Y^* = S + X_I^*$, in agreement with (15).

Applying this construction in the presence of dependence results in S, a function of the variables which can be kept fixed, and variables $T', T^\dagger, T^\ddagger$, on a joint space, such that

$$Y' = S + T', \quad Y^\dagger = S + T^\dagger, \quad \text{and} \quad Y^\ddagger = S + T^\ddagger. \tag{20}$$

When $T', T^\dagger$ and $T^\ddagger$ are all bounded by B, (19) gives

$$|Y^* - Y'| = |UT^\dagger + (1-U)T^\ddagger - T'| \leq U|T^\dagger| + (1-U)|T^\ddagger| + |T'| \leq 2B. \tag{21}$$

Let an array $\{a_{ij}\}_{i,j=1}^n$ of real numbers satisfy

$$\sum_{j=1}^n a_{ij} = 0 \quad \text{for all } i, \text{ and set} \quad C = \max_{i,j} |a_{ij}|. \tag{22}$$

By replacing Y in (11) by $Y - \mathbb{E}Y$ we assume, without loss of generality, that $\mathbb{E}a_{i,\pi(i)} = 0$ for every i. In Theorem 2.5, below, where π is uniformly distributed over $\mathcal{S}_n$, this assumption is equivalent to (22). In Theorem 2.6, since π has no fixed points, by (27), without loss of generality we have $a_{ii} = 0$ for all i in (26). In addition, since the distribution of π is constant on permutations having the same cycle type, by (25), $\mathbb{E}\, a_{i,\pi(i)} = (1/(n-1))\sum_{j\neq i} a_{ij}$, and the mean zero assumption is again equivalent to (22). Avoiding trivial cases, we also assume that $\mathrm{Var}(Y) = \sigma^2 > 0$. For ease of notation we write Y' and π' interchangeably for Y and π, respectively, in the remainder of this section .

In Sections 2.1 and 2.2 the construction above produces variables $Y', Y^\dagger$ and $Y^\ddagger$, given by (11) (with π replaced by $\pi', \pi^\dagger$ and $\pi^\ddagger$, respectively), and a set of indices $\mathcal{I}$ outside of which these permutations agree, such that (20) holds with

$$S = \sum_{i \notin \mathcal{I}} a_{i,\pi'(i)}, \; T' = \sum_{i \in \mathcal{I}} a_{i,\pi'(i)}, \; T^\dagger = \sum_{i \in \mathcal{I}} a_{i,\pi^\dagger(i)}, \text{ and } T^\ddagger = \sum_{i \in \mathcal{I}} a_{i,\pi^\ddagger(i)}. \tag{23}$$

Therefore B in (21) can be set equal to C in (22) times a worst case bound on the size of $\mathcal{I}$. The specifications of $\pi', \pi'', \pi^\dagger$, and $\pi^\ddagger$ are given in terms of transpositions τ_{ij}, those permutations satisfying $\tau_{ij}(i) = j, \tau_{ij}(j) = i$ and $\tau_{ij}(k) = k$ for all $k \notin \{i, j\}$.

2.1 Uniform permutation distribution

Many authors (e.g. [29] [7], [20]) have considered normal approximation to the distribution of (11) when π is a permutation chosen uniformly from $\mathcal{S}_n$. In Theorem 2.5, the dependence of δ on C is not as refined as the bound in [7], which depends on an (unspecified) universal constant times the normalized absolute third moments of the $\{a_{ij}\}_{i,j=1}^n$. Here, on the other hand, an explicit constant is provided.

Theorem 2.5 *With $n \geq 3$, let $\{a_{ij}\}_{i,j=1}^n$ satisfy (22) and let π be a random permutation with uniform distribution over $\mathcal{S}_n$. Then, with C as in (22), conclusions (9) and (10) of Theorem 1.1 hold for the sum $Y = \sum_{i=1}^n a_{i,\pi(i)}$ with $A = 8C/\sigma$ when $A \leq 1/12$.*

Proof: Given π', take (I, J) to be independent of π', uniformly over all pairs with $1 \leq I \neq J \leq n$, and set $\pi'' = \pi'\tau_{I,J}$. In particular, $\pi''(i) = \pi'(i)$ for $i \notin \{I, J\}$; the variables Y' and Y'', given by (11) with π' and π'' respectively, are exchangeable; and

$$Y' - Y'' = (a_{I,\pi'(I)} + a_{J,\pi'(J)}) - (a_{I,\pi'(J)} + a_{J,\pi'(I)}). \tag{24}$$

The linearity condition (16) is satisfied with $\lambda = 2/(n-1)$, since, from (24) and (22),

$$\begin{aligned} E\left(Y' - Y''|\pi'\right) &= 2\left(\frac{1}{n}\sum_{i=1}^n a_{i,\pi'(i)} - \frac{1}{n(n-1)}\sum_{i\neq j} a_{i,\pi'(j)}\right) \\ &= 2\left(\frac{1}{n}\sum_{i=1}^n a_{i,\pi'(i)} + \frac{1}{n(n-1)}\sum_{i=1}^n a_{i,\pi'(i)}\right) = \frac{2}{n-1}Y'. \end{aligned}$$

To construct $(Y^\dagger, Y^\ddagger)$ with distribution proportional to $(y'-y'')^2 dP(y', y'')$, choose $I^\dagger, K^\dagger, J^\dagger, L^\dagger$ with distribution proportional to the squared difference $(Y' - Y'')^2$, that is,

$$P(I^\dagger = i, K^\dagger = k, J^\dagger = j, L^\dagger = l) \sim \left[(a_{ik} + a_{jl}) - (a_{il} + a_{jk})\right]^2,$$

and let

$$\pi^\dagger = \begin{cases} \pi\tau_{\pi^{-1}(K^\dagger),J^\dagger} & \text{if } L^\dagger = \pi(I^\dagger), K^\dagger \neq \pi(J^\dagger) \\ \pi\tau_{\pi^{-1}(L^\dagger),I^\dagger} & \text{if } L^\dagger \neq \pi(I^\dagger), K = \pi(J^\dagger) \\ \pi\tau_{\pi^{-1}(K^\dagger),I^\dagger}\tau_{\pi^{-1}(L^\dagger),J^\dagger} & \text{otherwise,} \end{cases}$$

and $\pi^\ddagger = \pi^\dagger\tau_{I^\dagger,J^\dagger}$. Then (20) and (23) hold with $\mathcal{I} = \{I^\dagger, \pi^{-1}(K^\dagger), J^\dagger, \pi^{-1}(L^\dagger)\}$, a set of size at most 4, so by (21), $|Y^* - Y| \leq 8C$. ■

2.2 Permutations with distribution constant over cycle type

In this section we focus on the normal approximation of Y as in (11) when the distribution of the random permutation π is a function only of its cycle type. Our framework includes the case considered in [22], the uniform distribution over permutations with a single cycle.

Consider a permutation $\pi \in \mathcal{S}_n$ represented in cycle form; in $\mathcal{S}_7$ for example, $\pi = ((1,3,7,5),(2,6,4))$ is the permutation consisting of one 4 cycle in which $1 \to 3 \to 7 \to 5 \to 1$ and one 3 cycle where $2 \to 6 \to 4 \to 2$. For $q = 1,\ldots,n$, let $c_q(\pi)$ be the number of q cycles of π. We say permutations π and σ are of the same cycle type if $c_q(\pi) = c_q(\sigma)$ for all $q = 1,\ldots,n$; π and σ are of the same cycle type if and only if π and σ are conjugate, i.e. if and only if there exists a permutation ρ such that $\pi = \rho^{-1}\sigma\rho$. Hence, we say a probability measure $\mathbf{P}$ on $\mathcal{S}_n$ is constant over cycle type if

$$\mathbf{P}(\pi) = \mathbf{P}(\rho^{-1}\pi\rho) \quad \text{for all } \pi, \rho \in \mathcal{S}_n. \tag{25}$$

In [18], the authors consider a statistical test for determining when a given pairing of $n = 2m$ observations shows an unusually high level of similarity; the test statistic is of the form (11), and, under the null hypothesis of no distinguished pairing, the distribution $\mathbf{P}$ satisfies (25) with $\mathbf{P}(\pi)$ equal to a constant if π has m 2-cycles, and $\mathbf{P}(\pi) = 0$ otherwise; that is, under the null, $\mathbf{P}$ is uniform over permutations having m 2-cycles. Bounds between the normal and the null distribution of Y were determined in [18] using a construction in which an exchangeable π'' is obtained from π by a transformation which preserves the m 2-cycle structure. The construction in Theorem 2.6 preserves the cycle structure in general and, when there are m 2-cycles, specializes to one similar, but not equivalent, to that of [18].

Theorem 2.6 *With $n \geq 4$, let an array $\{a_{ij}\}_{ij=1}^n$ of real numbers satisfy (22), let*

$$a_{ij} = a_{ji} \quad \text{and} \quad a_{ii} = 0, \tag{26}$$

and let $\pi \in \mathcal{S}_n$ be a random permutation with distribution $\mathbf{P}$ constant on cycle type, with no fixed points. That is, $\mathbf{P}$ satisfies (25), (26), and

$$\mathbf{P}(\pi) = 0 \quad \text{if } c_1(\pi) \neq 0. \tag{27}$$

Then, with C as in (22), conclusions (9) and (10) of Theorem 1.1 hold for the sum $Y = \sum_{i=1}^n a_{i,\pi(i)}$ with $A = 40C/\sigma$ when $A \leq 1/12$.

Proof: To fully highlight the reason for the imposition of the conditions (26) and (27), and also to make the complete case analysis easier to follow, we initially consider an array satisfying only the consequence $\sum_{1\le i,j\le n} a_{ij} = 0$ of (22), and a $\mathbf{P}$ not necessarily satisfying (27).

Again, using the construction outlined in Remark 2.3, we first construct π'' from the given π'. Let I and J, $1 \le I \ne J \le n$ be chosen uniformly and independently of π', and let $\pi'' = \tau_{IJ}\pi'\tau_{IJ}$; that is, π'' is obtained by interchanging I and J in the cycle representation of π'. We claim the pair π', π'' is exchangeable. For fixed permutations σ'', σ', if $\sigma' \ne \tau_{IJ}\sigma''\tau_{IJ}$ then

$$\mathbf{P}(\pi'' = \sigma'', \pi' = \sigma') = 0 = \mathbf{P}(\pi' = \sigma'', \pi'' = \sigma').$$

Otherwise, $\sigma' = \tau_{IJ}\sigma''\tau_{IJ}$ and, using (25) for the second equality, we have

$$\mathbf{P}(\pi'' = \sigma'', \pi' = \sigma') = \mathbf{P}(\pi' = \sigma') = \mathbf{P}(\pi' = \tau_{IJ}\sigma'\tau_{IJ}) = \mathbf{P}(\pi'' = \sigma') = \mathbf{P}(\pi' = \sigma'', \pi'' = \sigma').$$

Consequently, Y and Y'', given by (11) with permutations π and π'', respectively, are exchangeable. By conditioning on π, we show Y', Y'' satisfies the linearity condition (16) with $\lambda = 4/n$.

Let S be the size of the set $\{I, J, \pi(I), \pi(J)\}$, and, for $i \in \{1,\dots,n\}$ let $|i|$ denote the number of elements in the cycle of π that contains i. Since $I \ne J$, we have $2 \le S \le 4$. When $S = 2$, either $\pi(I) = I$ and $\pi(J) = J$, or $\pi(I) = J$ and $\pi(J) = I$; in the both cases $\pi'' = \pi$. There are four cases for $S = 3$; either $A_{I,J} = \{|I| = 1, |J| \ge 2\}$ or I and J are interchanged (denoted by $A_{J,I}$); or I, J and $\pi(J)$ are three consecutive distinct values of π, indicated by $B_{I,J}$, or I and J are interchanged (denoted by $B_{J,I}$). The case $S = 4$ is indicated by F. Hence,

$$\begin{aligned} Y' - Y'' \;=\; & \left(a_{I,I} + a_{\pi^{-1}(J),J} + a_{J,\pi(J)} - (a_{J,J} + a_{\pi^{-1}(J),I} + a_{I,\pi(J)})\right) A_{I,J} \\ + \; & \left(a_{J,J} + a_{\pi^{-1}(I),I} + a_{I,\pi(I)} - (a_{I,I} + a_{\pi^{-1}(I),J} + a_{J,\pi(I)})\right) A_{J,I} \\ + \; & \left(a_{\pi^{-1}(I),I} + a_{I,J} + a_{J,\pi(J)} - (a_{\pi^{-1}(I),J} + a_{J,I} + a_{I,\pi(J)})\right) B_{I,J} \\ + \; & \left(a_{\pi^{-1}(J),J} + a_{J,I} + a_{I,\pi(I)} - (a_{\pi^{-1}(J),I} + a_{I,J} + a_{J,\pi(I)})\right) B_{J,I} \\ + \; & \left(a_{\pi^{-1}(I),I} + a_{I,\pi(I)} + a_{\pi^{-1}(J),J} + a_{J,\pi(J)} - (a_{\pi^{-1}(I),J} + a_{J,\pi(I)} + a_{\pi^{-1}(J),I} + a_{I,\pi(J)})\right) F. \end{aligned} \tag{28}$$

For example, using the fact that the sum of $a_{\pi^{-1}(J),J}$ is the same as that of $a_{J,\pi(J)}$ over a given cycle, the contribution to $n(n-1)\mathbf{E}(Y' - Y''|\pi)$ from $A_{I,J} = \{|I| = 1, |J| \ge 2\}$, added to the equal one from $A_{J,I}$, simplifies to

$$2(n - 3c_1(\pi)) \sum_{|i|=1} a_{i,i} + 4c_1(\pi) \sum_{|i|\ge 1} a_{i,\pi(i)} - 2c_1(\pi) \sum_{|i|\ge 2} a_{i,i} - 2 \sum_{|i|=1,|j|\ge 2} a_{i,j} - 2 \sum_{|i|\ge 2,|j|=1} a_{i,j}. \tag{29}$$

Next, the equal contributions from $B_{I,J} = \mathbf{1}(\pi(I) = J, |I| \ge 3)$ and $B_{J,I}$ sum to

$$6 \sum_{|i|\ge 3} a_{i,\pi(i)} - 4 \sum_{|i|\ge 3} a_{\pi^{-1}(i),\pi(i)} - 2 \sum_{|i|\ge 3} a_{\pi(i),i}. \tag{30}$$

On $F = \mathbf{1}(|I| \ge 2, |J| \ge 2, I \ne J, \pi(I) \ne J, \pi(J) \ne I)$, the contribution from $a_{\pi^{-1}(I),I}$ is

$$\sum_{|i|,|j|\ge 2} a_{\pi^{-1}(i),i} \mathbf{1}(i \ne j, \pi(i) \ne j, \pi(j) \ne i). \tag{31}$$

Let $i \cong j$ denote the fact that i and j are elements of the same cycle. When $i \cong j$ and $\{i, j, \pi(i), \pi(j)\}$ are distinct, we have $|i| \geq 4$ and $|i|-3$ possible choices for $j \cong i$ that satisfy the conditions in the indicator in (31). Hence, the case $i \cong j$ contributes

$$\sum_{|i|\geq 4} a_{\pi^{-1}(i),i} \sum_{j \cong i} \mathbf{1}(i \neq j, \pi(i) \neq j, \pi(j) \neq i) = \sum_{|i|\geq 4} a_{\pi^{-1}(i),i}(|i|-3) = \sum_{|i|\geq 3}(|i|-3)a_{i,\pi(i)}.$$

When $i \ncong j$ the conditions in the indicator function (31) are satisfied if and only if $|i| \geq 2, |j| \geq 2$. For $|i| \geq 2$ there are $n - |i| - c_1(\pi)$ choices for j, so the case $i \ncong j$ contributes

$$\sum_{|i|\geq 2} a_{\pi^{-1}(i),i} \sum_{j \ncong i, |j|\geq 2} 1 = \sum_{|i|\geq 2}(n-|i|-c_1(\pi))a_{i,\pi(i)}.$$

The next three terms on F give the same as the first, so in total we have

$$4(n-2-c_1(\pi)) \sum_{|i|=2} a_{i,\pi(i)} + 4(n-3-c_1(\pi)) \sum_{|i|\geq 3} a_{i,\pi(i)}. \tag{32}$$

Decomposing the contribution from the fifth term, according to whether $i \cong j$ or $i \ncong j$, gives

$$\begin{aligned}
& -\sum_{|i|,|j|\geq 2} a_{\pi^{-1}(i),j}\mathbf{1}(i \neq j, \pi(i) \neq j, \pi(j) \neq i) \\
& -\sum_{|i|\geq 4}\sum_{j\cong i} a_{\pi^{-1}(i),j}\mathbf{1}(i \neq j, \pi(i) \neq j, \pi(j) \neq i) - \sum_{|i|,|j|\geq 2}\sum_{j \ncong i} a_{\pi^{-1}(i),j} \\
= & -\sum_{|i|\geq 4}\sum_{j\cong i} a_{\pi^{-1}(i),j} + \sum_{|i|\geq 4}\left(a_{\pi^{-1}(i),i} + a_{\pi^{-1}(i),\pi(i)} + a_{\pi^{-1}(i),\pi^{-1}(i)}\right) - \sum_{|i|,|j|\geq 2}\sum_{j\ncong i} a_{i,j} \\
= & -\sum_{|i|\geq 4}\sum_{j\cong i} a_{i,j} + \sum_{|i|\geq 4}\left(a_{i,\pi(i)} + a_{\pi^{-1}(i),\pi(i)} + a_{i,i}\right) - \sum_{|i|,|j|\geq 2}\sum_{j\ncong i} a_{i,j}.
\end{aligned} \tag{33}$$

To simplify (33), let $a \wedge b = \min(a, b)$ and consider the decomposition

$$\sum_{i,j=1}^{n} a_{i,j} = \sum_{|i|\geq 4}\sum_{j\cong i} a_{i,j} + \sum_{|i|\leq 3}\sum_{j\cong i} a_{i,j} + \sum_{|i|,|j|\geq 2}\sum_{j\ncong i} a_{i,j} + \sum_{|i|\wedge|j|=1}\sum_{j\ncong i} a_{i,j}. \tag{34}$$

Since $\sum_{i,j} a_{ij} = 0$, we may replace the sum of the first and last terms in (33) by the sum of the second and fourth terms on the right hand side of (34), respectively, resulting in

$$\begin{aligned}
& \sum_{|i|\leq 3}\sum_{j\cong i} a_{i,j} + \sum_{|i|\wedge|j|=1}\sum_{j\ncong i} a_{i,j} + \sum_{|i|\geq 4}\left(a_{i,\pi(i)} + a_{\pi^{-1}(i),\pi(i)} + a_{i,i}\right) \\
= & \sum_{|i|\leq 2}\sum_{j\cong i} a_{i,j} + \sum_{|i|\wedge|j|=1}\sum_{j\ncong i} a_{i,j} + \sum_{|i|\geq 3}\left(a_{i,\pi(i)} + a_{\pi^{-1}(i),\pi(i)} + a_{i,i}\right),
\end{aligned}$$

where we have used the fact that $\pi^2(i) = \pi^{-1}(i)$ when $|i| = 3$. Similarly shifting the $|i| = 2$ term we obtain

$$\begin{aligned}
& \sum_{|i|=1} a_{i,i} + \sum_{|i|\wedge|j|=1}\sum_{j\ncong i} a_{i,j} + \sum_{|i|\geq 2}\left(a_{i,\pi(i)} + a_{i,i}\right) + \sum_{|i|\geq 3} a_{\pi^{-1}(i),\pi(i)} \\
= & \sum_{|i|\wedge|j|=1}\sum_{j\ncong i} a_{i,j} + \sum_{|i|\geq 2} a_{i,\pi(i)} + \sum_{|i|\geq 1} a_{i,i} + \sum_{|i|\geq 3} a_{\pi^{-1}(i),\pi(i)}.
\end{aligned}$$

Combining this with the next three terms of F, each of which yields the same contribution, gives

$$4\sum_{|i|\geq 2} a_{i,\pi(i)} + 4\sum_{|i|\geq 3} a_{\pi^{-1}(i),\pi(i)} + 4\sum_{|i|\geq 1} a_{i,i} + 4\sum_{|i|\wedge|j|=1}\sum_{j\not\cong i} a_{i,j}. \tag{35}$$

Combining (35) with the contribution (32) of the first four terms in F, the $A_{I,J}$ and $A_{J,I}$ terms in (29) and the $B_{I,J}$ and $B_{J,I}$ terms (30), yields $n(n-1)\mathbb{E}(Y'-Y''|\pi')$; after cancelling the terms involving $a_{\pi^{-1}(i),\pi(i)}$ in (30) and (35) and grouping like terms, we obtain

$$4(n-1)\sum_{|i|=2} a_{i,\pi(i)} + (4n-2)\sum_{|i|\geq 3} a_{i,\pi(i)} - 2\sum_{|i|\geq 3} a_{\pi(i),i} \tag{36}$$

$$+ \quad 2(n-c_1(\pi)+2)\sum_{|i|=1} a_{i,i} - 2(c_1(\pi)-2)\sum_{|i|\geq 2} a_{i,i} \tag{37}$$

$$+ \quad 4\sum_{|i|\wedge|j|=1, j\not\cong i} a_{i,j} - 2\sum_{|i|=1,|j|\geq 2} a_{i,j} - 2\sum_{|i|\geq 2,|j|=1} a_{i,j}. \tag{38}$$

The assumption that $a_{i,i}=0$ causes the contribution from (37) to vanish, the assumption that there are no 1-cycles causes the contribution from (38) to vanish, and the assumption that $a_{i,j}$ is symmetric causes the combination of the second and third terms in (36) to yield $\mathbb{E}(Y'-Y''|\pi') = (4/n)\sum_{i=1}^n a_{i,\pi'(i)} = (4/n)Y'$. Hence, the linearity condition (16) is satisfied.

Since $\pi'' = \tau_{IJ}\pi\tau_{IJ}$, the terms that multiply the indicator functions in the difference $Y'-Y''$ in (28) depend only on values in a subset of $\{\pi^{-1}(I), I, \pi(I), \pi^{-1}(J), J, \pi(J)\}$ determined by the event indicated; for example, on $B_{I,J}$ the difference only depends on $\{\pi^{-1}(I), I, J, \pi(J)\}$. For each event we tabulate such values in a vector $\mathbf{i}$. Likewise, with $\pi^\dagger$ and $\pi^\ddagger$ constructed according to $\pi^\ddagger = \tau_{I^\dagger J^\dagger}\pi^\dagger\tau_{I^\dagger J^\dagger}$, the difference $Y^\dagger - Y^\ddagger$ depends only on a subset of $\{P^\dagger, I^\dagger, K^\dagger, Q^\dagger, J^\dagger, L^\dagger\}$, the corresponding values in the $\pi^\dagger$ cycle, which we will tabulate in a vector $\mathbf{i}^\dagger$. Since $Y'-Y''$ in (28) is a sum of terms multiplied by indicator functions of disjoint events, $(Y'-Y'')^2$ is a sum of those terms squared, multiplied by the same indicator functions. Hence to generate $(\pi^\dagger, \pi^\ddagger)$ such that $(Y^\dagger, Y^\ddagger)$ has a distribution proportional to $(y'-y'')^2 dF(y', y'')$, on each event we generate the elements of $\mathbf{i}^\dagger$ with square weighted probability appropriate to the set indicated. Once the values in $\mathbf{i}^\dagger$ are chosen, in order for $\pi^\dagger$ to have the conditional distribution of π given these values, the remaining values of $\pi^\dagger$ are obtained by interchanging $\mathbf{i}$ with $\mathbf{i}^\dagger$ in the cycle structure of π. That is, in each case we specify $\pi^\dagger$ in terms of π by

$$\pi^\dagger = \tau_{\mathbf{i},\mathbf{i}^\dagger}\pi\tau_{\mathbf{i},\mathbf{i}^\dagger} \quad \text{where} \quad \tau_{\mathbf{i},\mathbf{i}^\dagger} = \prod_{k=1}^{\kappa} \tau_{i_k, i_k^\dagger}, \tag{39}$$

and $\mathbf{i} = (i_1, \ldots, i_\kappa)$ and $\mathbf{i}^\dagger = (i_1^\dagger, \ldots, i_\kappa^\dagger)$ are vectors of disjoint indices, of some length κ.

For $\rho \in \mathcal{S}_\kappa$ and $\mathbf{l} = (l_1, \ldots, l_\kappa)$ any κ-dimensional vector of indices, let $\rho(\mathbf{l}) = \{\rho(l_k) : k = 1, \ldots, \kappa\}$, and let ι be the identity permutation. Since the values of $\tau_{ii^\dagger}\pi\tau_{ii^\dagger}$ may differ from those of π only at $i, i^\dagger, \pi^{-1}(i)$ and $\pi^{-1}(i^\dagger)$, (20) will hold for the variables given by (23), with

$$\mathcal{I} = \iota(\mathbf{i}) \cup \iota(\mathbf{i}^\dagger) \cup \pi^{-1}(\mathbf{i}) \cup \pi^{-1}(\mathbf{i}^\dagger).$$

The construction in each case proceeds as follows. Since 1-cycles are excluded, $A_{I,J}$ and $A_{J,I}$ are null. On $B_{I,J}$, where I, J and $\pi(J)$ are three distinct, consecutive values of π, if $|I| = 3$ then the symmetry of $a_{i,j}$ gives $Y'' = Y'$, an event on which the distribution of $(Y^\dagger, Y^\ddagger)$, proportional to $(Y'' - Y')^2$, puts mass zero. Otherwise, $|I| \geq 4$ and $Y' - Y''$ depends only on $\mathbf{i} = (\pi^{-1}(I), I, J, \pi(J))$, and we choose $\mathbf{i}^\dagger = (P^\dagger, I^\dagger, J^\dagger, L^\dagger)$, the corresponding values for $\pi^\dagger$, according to the distribution

$$(P^\dagger, I^\dagger, J^\dagger, L^\dagger) \quad \sim \quad [(a_{p,i} + a_{j,l}) - (a_{p,j} + a_{i,l})]^2 \mathbf{1}(p, i, j, \text{ and } l \text{ are distinct}),$$

noting that $a_{i,j}$ cancels with $a_{j,i}$ by symmetry. Now set $\pi^\dagger$ as specified in (39). In this case $\mathcal{I}$ has size at most thirteen. Reversing the roles of I and J gives the construction on $B_{J,I}$.

Next consider F, where $I, \pi(I), J$ and $\pi(J)$ are distinct. If $|I| = |J| = 2$ then take

$$(I^\dagger, K^\dagger, J^\dagger, L^\dagger) \sim [(a_{i,k} + a_{j,l}) - (a_{i,l} + a_{j,k})]^2 \mathbf{1}(i, k, j \text{ and } l \text{ are distinct}),$$

and set $\pi^\dagger$ as specified in (39), with $\mathbf{i} = (I, \pi(I), J, \pi(J))$ and $\mathbf{i}^\dagger = (I^\dagger, K^\dagger, J^\dagger, L^\dagger)$, and with the size of $\mathcal{I}$ at most twelve. For $|I| \geq 3$ and $|J| = 2$, take

$$(P^\dagger, I^\dagger, K^\dagger, J^\dagger, L^\dagger) \sim [(a_{p,i} + a_{i,k} + 2a_{j,l}) - (a_{p,j} + a_{j,k} + 2a_{i,l})]^2 \mathbf{1}(p, i, k, j, \text{ and } l \text{ are distinct}),$$

and set $\pi^\dagger$ as specified in (39), with $\mathbf{i} = (\pi^{-1}(I), I, \pi(I), J, \pi(J))$ and $\mathbf{i}^\dagger = (P^\dagger, I^\dagger, K^\dagger, J^\dagger, L^\dagger)$, and with the size of $\mathcal{I}$ at most sixteen. Reversing the roles of I and J gives the case in which $|J| = 2$ but $|I| \geq 3$. For $|I| \geq 3, |J| \geq 3$, take

$$\begin{aligned}(P^\dagger, I^\dagger, K^\dagger, Q^\dagger, J^\dagger, L^\dagger) \quad &\sim \quad [(a_{p,i} + a_{i,k} + a_{q,j} + a_{j,l}) - (a_{p,j} + a_{j,k} + a_{q,i} + a_{i,l})]^2 \\ &\times \quad \mathbf{1}(p, i, k, q, j, \text{ and } l \text{ distinct}),\end{aligned}$$

and set $\pi^\dagger$ as specified in (39), with $\mathbf{i} = (\pi^{-1}(I), I, \pi(I), \pi^{-1}(J), J, \pi(J))$ and $\mathbf{i}^\dagger = (P^\dagger, I^\dagger, K^\dagger, Q^\dagger, J^\dagger, L^\dagger)$. In this case, the size of $\mathcal{I}$ is at most twenty and, by (21), $|Y^* - Y| \leq 40C$ in all cases. ■

3 Size Biasing: Permutations and Patterns

In this section we derive corollaries of Theorem 1.2 to obtain Berry Esseen bounds for the number of occurrences of fixed, relatively ordered sub-sequences, such as rising sequences, in a random permutation, and of color patterns, local maxima, and sub-graphs in finite graphs.

Following [17], given a finite collection $\mathbf{X} = \{X_\alpha, \alpha \in \mathcal{A}\}$ of non-negative random variables with index set $\mathcal{A}$, for $\alpha \in \mathcal{A}$ we say the collection $\mathbf{X}^\alpha = \{X^\alpha_\beta, \beta \in \mathcal{A}\}$ has the $\mathbf{X}$-size-biased distribution in direction α if

$$\mathbb{E} X_\alpha f(\mathbf{X}) = \mathbb{E} X_\alpha \mathbb{E} f(\mathbf{X}^\alpha) \tag{40}$$

for all functions f on $\mathbf{X}$ for which these expectations exist. For the given $\mathbf{X}$, the collection $\mathbf{X}^\alpha$ exists for any $\alpha \in \mathcal{A}$ and has distribution $dP^\alpha(\mathbf{x}) = x_\alpha dP(\mathbf{x})/\mathbb{E} X_\alpha$, where $dP(\mathbf{x})$ is the distribution of $\mathbf{X}$. Specializing (40) to the coordinate function $f(\mathbf{X}) = g(X_\alpha)$, we see that X^α_α has the X_α-size-biased distribution X^s_α, defined in (4).

Corollary 3.1 *Let $\mathbf{X} = \{X_\alpha, \alpha \in \mathcal{A}\}$ be a finite collection of random variables with values in $[0, M]$ and let $Y = \sum_{\alpha \in \mathcal{A}} X_\alpha$. Assume, for each $\alpha \in \mathcal{A}$, there exists a dependency neighborhood $\mathcal{B}_\alpha \subset \mathcal{A}$ such that*

$$X_\alpha \quad \textit{and} \quad \{X_\beta : \beta \notin \mathcal{B}_\alpha\} \quad \textit{are independent.} \tag{41}$$

Furthermore, let $p_\alpha = \mathbb{E}\, X_\alpha / \sum_{\beta\in\mathcal{A}} \mathbb{E}\, X_\beta$ *and* $\max_\alpha |\mathcal{B}_\alpha| = b$*. For each* $\alpha \in \mathcal{A}$*, let* $(\mathbf{X}, \mathbf{X}^\alpha)$ *be a coupling of* $\mathbf{X}$ *to an* $\mathbf{X}^\alpha$ *with the* $\mathbf{X}$*-size-biased distribution in direction* α*, and let* $\mathcal{D} \subset \mathcal{A}\times\mathcal{A}$ *and* $\mathcal{F} \supset \sigma\{Y\}$ *be such that if* $(\alpha_1, \alpha_2) \notin \mathcal{D}$ *then*

$$Cov(\mathbb{E}\,(X_{\beta_1}^{\alpha_1} - X_{\beta_1}|\mathcal{F}), \mathbb{E}\,(X_{\beta_2}^{\alpha_2} - X_{\beta_2}|\mathcal{F})) = 0 \quad \textit{for all } (\beta_1, \beta_2) \in \mathcal{B}_{\alpha_1} \times \mathcal{B}_{\alpha_2}. \tag{42}$$

Then Theorem 1.2 may be applied with

$$B = bM \quad \textit{and} \quad \Delta \le M \sqrt{\sum_{(\alpha_1,\alpha_2)\in\mathcal{D}} p_{\alpha_1} p_{\alpha_2} |\mathcal{B}_{\alpha_1}||\mathcal{B}_{\alpha_2}|} \le (\max_\alpha p_\alpha) bM \sqrt{|\mathcal{D}|}. \tag{43}$$

Proof: Assuming, without loss of generality, that $\mathbb{E} X_\alpha > 0$ for each $\alpha \in \mathcal{A}$, the factorization

$$P^\alpha(\mathbf{X} \in d\mathbf{x}) = \left(\frac{x_\alpha P(X_\alpha \in dx_\alpha)}{\mathbb{E}\, X_\alpha} \right) P(\mathbf{X} \in d\mathbf{x} \,|\, X_\alpha = x_\alpha)$$

shows that we can construct $\mathbf{X}^\alpha$ by first choosing X_α^α from the X_α-size-bias distribution, and then choosing the remaining variables from the conditional distribution of $\mathbf{X}$, given the chosen value of X_α^α. Note that $X_\beta^\alpha \in [0, M]$ for all α, β and, by (41), that we may take $X_\beta^\alpha = X_\beta$ for $\beta \notin \mathcal{B}_\alpha$. By Lemma 2.1 of [17], $Y^s = \sum_{\beta\in\mathcal{A}} X_\beta^I$ has the Y-size-biased distribution, where the random index I has distribution $P(I = \alpha) = p_\alpha$, and is independent of both $(\mathbf{X}, \mathbf{X}^\alpha)$ and $\mathcal{F}$. Hence

$$Y^s - Y = \sum_{\beta\in\mathcal{B}_I} (X_\beta^I - X_\beta), \quad \text{and therefore,} \quad |Y^s - Y| \le bM. \tag{44}$$

Since $\sigma\{Y\} \subset \mathcal{F}$, $\Delta^2 = \text{Var}(\mathbb{E}\,(Y^s - Y|Y)) \le \text{Var}(\mathbb{E}\,(Y^s - Y|\mathcal{F}))$. Taking conditional expectation with respect to $\mathcal{F}$ in (44) yields,

$$\mathbb{E}\,(Y^s - Y|\mathcal{F}) = \sum_{\alpha\in\mathcal{A}} p_\alpha \sum_{\beta\in\mathcal{B}_\alpha} \mathbb{E}\,(X_\beta^\alpha - X_\beta|\mathcal{F})$$

and, therefore,

$$\text{Var}(\mathbb{E}\,(Y^s - Y|\mathcal{F})) \;=\; \mathbb{E} \sum_{\substack{(\alpha_1,\alpha_2)\in\mathcal{A}\times\mathcal{A} \\ (\beta_1,\beta_2)\in\mathcal{B}_{\alpha_1}\times\mathcal{B}_{\alpha_2}}} p_{\alpha_1} p_{\alpha_2} \text{Cov}(\mathbb{E}\,(X_{\beta_1}^{\alpha_1} - X_{\beta_1}|\mathcal{F}), \mathbb{E}\,(X_{\beta_2}^{\alpha_2} - X_{\beta_2}|\mathcal{F})).$$

Using (42), we may replace the sum over $(\alpha_1, \alpha_2) \in \mathcal{A}\times\mathcal{A}$ by the sum over $(\alpha_1, \alpha_2) \in \mathcal{D}$, and subsequent application of the Cauchy Schwarz inequality yields the bound (43) for Δ. ■

If, in some asymptotic regime, the X_α are comparable in expectation in such a way that $p_\alpha \sim |\mathcal{A}|^{-1}$; if μ and σ^2 grow like $|\mathcal{A}|$; if b remains bounded; and if $|\mathcal{D}|$ is of order $|\mathcal{A}|$, then, in Theorem 1.2, A and Δ and, therefore, δ are of order $1/\sigma$.

Corollary 3.2 *Let* $\mathcal{G}$ *be an index set, let* $\{C_g, g \in \mathcal{G}\}$ *be a collection of independent random elements taking values in an arbitrary set* $\mathcal{C}$*, let* $\{\mathcal{G}_\alpha, \alpha \in \mathcal{A}\}$ *be a finite collection of subsets of* $\mathcal{G}$*, and, for* $\alpha \in \mathcal{A}$*, let*

$$X_\alpha = X_\alpha(C_g : g \in \mathcal{G}_\alpha)$$

be a function of the variables $\{C_g, g \in \mathcal{G}_\alpha\}$, taking values in $[0, M]$. Then Theorem 1.2 may be applied to $Y = \sum_\alpha X_\alpha$ with B and Δ as in (43), taking $p_\alpha = \mathbb{E} X_\alpha / \sum_\beta \mathbb{E} X_\beta$,

$$\mathcal{B}_\alpha = \{\beta \in \mathcal{A} : \mathcal{G}_\beta \cap \mathcal{G}_\alpha \neq \emptyset\} \quad \textit{for } \alpha \in \mathcal{A}, \tag{45}$$

and any $\mathcal{D}$ for which

$$\mathcal{D} \supset \{(\alpha_1, \alpha_2) : \textit{there exists } (\beta_1, \beta_2) \in \mathcal{B}_{\alpha_1} \times \mathcal{B}_{\alpha_2} \textit{ with } \mathcal{G}_{\beta_1} \cap \mathcal{G}_{\beta_2} \neq \emptyset\}. \tag{46}$$

Proof: Since X_α and X_β are functions of disjoint sets of independent variables when $\mathcal{G}_\alpha \cap \mathcal{G}_\beta = \emptyset$, (41) holds with the dependency neighborhoods given by (45). Now, for each $\alpha \in \mathcal{A}$, consider the following $(\mathbf{X}, \mathbf{X}^\alpha)$ coupling. Let $\{C_g^{(\alpha)}, g \in \mathcal{G}_\alpha\}$ be independent of $\{C_g, g \in \mathcal{G}\}$ and have distribution

$$dP^{(\alpha)}(c_g, g \in \mathcal{G}_\alpha) = \frac{X_\alpha(c_g, g \in \mathcal{G}_\alpha)}{\mathbb{E} X_\alpha(c_g, g \in \mathcal{G}_\alpha)} dP(c_g, g \in \mathcal{G}_\alpha).$$

Then, by direct verification of (40), the collection

$$X_\beta^\alpha = X_\beta(C_g, g \in \mathcal{G}_\beta \cap \mathcal{G}_\alpha^c,\ C_g^{(\alpha)}, g \in \mathcal{G}_\beta \cap \mathcal{G}_\alpha), \quad \beta \in \mathcal{A}$$

has the $\mathbf{X}^\alpha$ distribution. Taking $\mathcal{F} = \{C_g : g \in \mathcal{G}\}$, we have $\mathbb{E}(X_\beta^\alpha | \mathcal{F}) = \mathbb{E}(X_\beta^\alpha | C_g, g \in \mathcal{G}_\beta)$ and, since $\mathbb{E}(X_\beta | \mathcal{F}) = X_\beta$, the conditional expectation $\mathbb{E}(X_\beta^\alpha - X_\beta | \mathcal{F})$ is a function of $\{C_g, g \in \mathcal{G}_\beta\}$ only. In particular, if $(\alpha_1, \alpha_2) \notin \mathcal{D}$ then, for all $\beta_1 \in \mathcal{B}_{\alpha_1}$ and $\beta_2 \in \mathcal{B}_{\alpha_2}$ we have $\mathcal{G}_{\beta_1} \cap \mathcal{G}_{\beta_2} = \emptyset$ and, consequently, $\mathbb{E}(X_{\beta_1}^{\alpha_1} - X_{\beta_1} | \mathcal{F})$ and $\mathbb{E}(X_{\beta_2}^{\alpha_2} - X_{\beta_2} | \mathcal{F})$ are independent, yielding (42), and all conditions of Corollary 3.1 hold. ∎

With the exception of Example 3.5, in the remainder of this section we consider graphs $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ having random elements $\{C_g\}_{g \in \mathcal{V} \cup \mathcal{E}}$ assigned to their vertices and edges, and applications of Corollary 3.2 to the sum $Y = \sum_{\alpha \in \mathcal{A}} X_\alpha$ of bounded functions $X_\alpha = X_\alpha(C_g, g \in \mathcal{V}_\alpha \cup \mathcal{E}_\alpha)$, where $\mathcal{G}_\alpha = (\mathcal{V}_\alpha, \mathcal{E}_\alpha), \alpha \in \mathcal{A}$ is a given finite family of subgraphs of $\mathcal{G}$; we abuse notation slightly in that a graph $\mathcal{G}$ is replaced by $\mathcal{V} \cup \mathcal{E}$ when used as an index set for the underlying variables C_g. When $\{C_g\}_{g \in \mathcal{G}}$ are independent, Corollary 3.2 applies and, in (45) and (46), the intersection of the two graphs $(\mathcal{V}_1, \mathcal{E}_1)$ and $(\mathcal{V}_2, \mathcal{E}_2)$ is the graph $(\mathcal{V}_1 \cap \mathcal{V}_2, \mathcal{E}_1 \cap \mathcal{E}_2)$.

Furthermore, if $\mathcal{A} \subset \mathcal{V}$ and there is a distance $d(\alpha, \beta)$ defined on $\mathcal{A}$, then letting

$$\rho = \inf\{\varrho : \mathcal{V}_\alpha \cap \mathcal{V}_\beta = \emptyset \text{ for all } \alpha, \beta \in \mathcal{A} \text{ with } d(\alpha, \beta) > \varrho\}, \tag{47}$$

we may use

$$\mathcal{B}_\alpha = \{\beta : d(\alpha, \beta) \leq \rho\} \quad \text{and} \quad \mathcal{D} = \{(\alpha_1, \alpha_2) : d(\alpha_1, \alpha_2) \leq 3\rho\} \tag{48}$$

in (45) and (46), respectively, since rearranging $d(\alpha_1, \alpha_2) \leq d(\alpha_1, \beta_1) + d(\beta_1, \beta_2) + d(\beta_2, \alpha_2)$ gives,

$$d(\beta_1, \beta_2) \geq d(\alpha_1, \alpha_2) - (d(\alpha_1, \beta_1) + d(\alpha_2, \beta_2)) \geq d(\alpha_1, \alpha_2) - 2\rho > \rho,$$

for $(\alpha_1, \alpha_2) \notin \mathcal{D}$ and $(\beta_1, \beta_2) \in \mathcal{B}_{\alpha_1} \times \mathcal{B}_{\alpha_2}$.

For $v \in \mathcal{V}$ and $r \geq 0$ let $\mathcal{G}_{v,r}$ be the restriction of $\mathcal{G}$ to the vertices at most a distance r from v; that is $\mathcal{G}_{v,r}$ has vertex set $\mathcal{V}_{v,r} = \{w \in \mathcal{V} : d(v, w) \leq r\}$ and edge set $\mathcal{E}_{v,r} = \{\{w, u\} \in \mathcal{E} : w, u \in \mathcal{V}_{v,r}\}$. We say that a graph $\mathcal{G}$ is *distance r-regular* if $\mathcal{G}_{v,r}$ is isomorphic to some graph $(\mathcal{V}_r, \mathcal{E}_r)$ for all v. For example, a graph of constant degree is distance 1-regular. This notion of distance r-regular is related to, but not the same as, the notion of a distance-regular graph as given in [6] and [9]. For a distance r-regular graph let

$$V(r) = |\mathcal{V}_r|. \tag{49}$$

Corollary 3.3, below, follows from Corollary 3.2 as a consequence of the remarks above, and by noting that the given assumptions imply that $|\mathcal{D}| = |\mathcal{A}|V(3\rho)$ and that $\mathbf{E}X_\alpha$ is constant, yielding $p_\alpha = 1/|\mathcal{A}|$.

Corollary 3.3 *Let $\mathcal{G}$ be a graph with a finite family of isomorphic subgraphs $\{\mathcal{G}_\alpha, \alpha \in \mathcal{A}\}, \mathcal{A} \subset \mathcal{V}$, let $d(\cdot,\cdot)$ be a distance on $\mathcal{A}$, and define ρ as in (47). For each $\alpha \in \mathcal{A}$, let X_α be given by*

$$X_\alpha = X(C_g, g \in \mathcal{G}_\alpha) \tag{50}$$

for a fixed function X taking values in $[0, M]$, and let the elements of $\{C_g\}_{g\in\mathcal{G}}$ be independent, with $\{C_g : g \in \mathcal{G}_\alpha\}$ identically distributed. If $\mathcal{G}$ is a distance-3ρ-regular graph, then Theorem 1.2 may be applied to $Y = \sum_{\alpha\in\mathcal{A}} X_\alpha$ with $V(r)$ as given in (49) and

$$B = V(\rho)M, \quad \Delta \le M|\mathcal{A}|^{-1/2}V(\rho)\sqrt{V(3\rho)}. \tag{51}$$

Natural families of examples in $\mathbb{R}^p$ can be generated using the vertex set $\mathcal{V} = \{1, \ldots, n\}^p$ with componentwise addition modulo n, and $d(\alpha, \beta)$ given by e.g. the L^1 distance $||\alpha - \beta||$.

Example 3.4 (*Sliding m-window.*) For $n \ge m \ge 1$, let $\mathcal{A} = \mathcal{V} = \{1, \ldots, n\}$ considered modulo n, $\{C_g : g \in \mathcal{G}\}$ i.i.d. real valued random variables, and for each $\alpha \in \mathcal{A}$

$$\mathcal{G}_\alpha = (\mathcal{V}_\alpha, \mathcal{E}_\alpha), \quad \text{where} \quad \mathcal{V}_\alpha = \{v \in \mathcal{V} : \alpha \le v \le \alpha + m - 1\} \quad \text{and} \quad \mathcal{E}_\alpha = \emptyset. \tag{52}$$

Then for $X : \mathbb{R}^m \to [0, 1]$, Corollary 3.3 may be applied to the sum $Y = \sum_{\alpha\in\mathcal{A}} X_\alpha$ of the m-dependent sequence $X_\alpha = X(C_\alpha, \ldots, C_{\alpha+m-1})$, formed by applying the function X to the variables in the 'm-window' $\mathcal{V}_\alpha$. In this example, taking $d(\alpha, \beta) = |\alpha - \beta|$ gives $\rho = m - 1$ and $V(r) = 2r + 1$. Hence, from (51), $B = (2m - 1)$ and $\Delta \le n^{-1/2}(2m - 1)(6m - 5)^{1/2}$.

In Example 3.5 the underlying variables are not independent, and Corollaries 3.2 and 3.3 cannot be directly applied.

Example 3.5 (*Relatively ordered sub-sequences of a random permutation.*) For $n \ge m \ge 1$, let π be a uniform random permutation of the integers $\mathcal{V} = \{1, \ldots, n\}$, taken modulo n. For a permutation τ on $\{1, \ldots, m\}$, let $\mathcal{G}_\alpha$ and $\mathcal{V}_\alpha$ be as specified in (52), and let X_α the indicator function requiring that the pattern τ appears on $\mathcal{V}_\alpha$; that is, that the values $\{\pi(v)\}_{v\in\mathcal{V}_\alpha}$ and $\{\tau(v)\}_{v\in\mathcal{V}_1}$ are in the same relative order. Equivalently, the pattern τ appears on $\mathcal{V}_\alpha$ if and only if $\pi(\tau^{-1}(v) + \alpha - 1), v \in \mathcal{V}_1$ is an increasing sequence, and we write

$$X_\alpha(\pi(v), v \in \mathcal{G}_\alpha) = \mathbf{1}(\pi(\tau^{-1}(1) + \alpha - 1) < \cdots < \pi(\tau^{-1}(m) + \alpha - 1)).$$

With $\mathcal{A} = \mathcal{V}$, the sum $Y = \sum_{\alpha\in\mathcal{A}} X_\alpha$ counts the number of m-element-long segments of π that have the same relative order as τ.

For $\alpha \in \mathcal{A}$, we generate $\mathbf{X}^\alpha = \{X^\alpha_\beta, \beta \in \mathcal{A}\}$ by reordering the values of $\pi(\gamma)$ for $\gamma \in \mathcal{V}_\alpha$, to be in the same relative order as τ, and let X^α_β be the indicator requiring τ to appear at position β in the reordered permutation. Letting $\mathcal{F} = \sigma\{\pi\}$, we have $\mathbb{E}(X^\alpha_\beta|\mathcal{F})$ and X_β depend only on the relative order of $\{\pi(\gamma) : -(m-1) \leq \gamma - \beta \leq 2(m-1)\}$. Since the relative order of the non-overlapping segments of the values of π are independent, (41) and (42) hold when $\mathcal{B}_\alpha$ and $\mathcal{D}$ are as in (48), for $d(\alpha, \beta) = |\alpha - \beta|$ and $\rho = m-1$; hence, Theorem 1.2 may be applied with the same value for B and bound on Δ as in Example 3.4.

When $\tau = \iota_m$, the identity permutation of length m, we say that π has a rising sequence of length m at position α if $X_\alpha = 1$. Rising sequences were studied in [4] in connection with card tricks and card shuffling. Due to the regular-self-overlap property of rising sequences, namely that a non-empty intersection of two rising sequences is again a rising sequence, some improvement on the constant in the bound can be obtained by a more careful consideration of the conditional variance.

Example 3.6 (*Coloring patterns and subgraph occurrences on a finite graph $\mathcal{G}$*). For illustration, take $\mathcal{V} = \mathcal{A} = \{1, \ldots, n\}^p$, considered modulo n, let $d(\alpha, \beta) = ||\alpha - \beta||$ with $||\cdot||$ the sup norm, let $\mathcal{E} = \{\{w, v\} : d(w, v) = 1\}$, and, for each $\alpha \in \mathcal{A}$, let $\mathcal{G}_\alpha = (\mathcal{V}_\alpha, \mathcal{E}_\alpha)$ where

$$\mathcal{V}_\alpha = \{\alpha + (e_1, \ldots, e_p) : e_i \in \{0, 1\}\} \quad \text{and} \quad \mathcal{E}_\alpha = \{\{v, w\} : v, w \in \mathcal{V}_\alpha,\ d(w, v) = 1\}.$$

Let $\mathcal{C}$ be a set (of e.g. colors) from which is formed a given pattern $\{c_g : g \in \mathcal{G}_\mathbf{0}\}$, let $\{C_g, g \in \mathcal{G}\}$ be independent variables in $\mathcal{C}$ with $\{C_g : g \in \mathcal{G}_\alpha\}_{\alpha \in \mathcal{A}}$ identically distributed, and let

$$X(C_g, g \in \mathcal{G}_\mathbf{0}) = \prod_{g \in \mathcal{G}_\mathbf{0}} \mathbf{1}(C_g = c_g), \tag{53}$$

and X_α given by (50). Then $Y = \sum_{\alpha \in \mathcal{A}} X_\alpha$ counts the number of times the pattern appears in the subgraphs $\mathcal{G}_\alpha$. Corollary 3.3 may be applied with $M = 1$, $\rho = 1$ (by (47)), $V(r) = (2r+1)^p$, and (by (51)) $B = 3^p$ and $\Delta \leq (63/n)^{p/2}$.

Such multi-dimensional pattern occurrences are a generalization of the well-studied case in which one-dimensional sequences are scanned for pattern occurrences; see, for instance, [13] and [23] for scan and window statistics, see [21] for applications of the normal approximation in this context to molecular sequence data, and see also [11] and [12], where higher-dimensional extensions are considered.

Occurrences of subgraphs can be handled as a special case. For example, with $(\mathcal{V}, \mathcal{E})$ the graph above, let G be the random subgraph with vertex set $\mathcal{V}$ and random edge set $\{e \in \mathcal{E} : C_e = 1\}$ where $\{C_e\}_{e \in \mathcal{E}}$ are independent and identically distributed Bernoulli variables. Then say, taking the product in (53) over edges $e \in \mathcal{E}_0$ and setting $c_e = 1$, the sum $Y = \sum_{\alpha \in \mathcal{A}} X_\alpha$ counts the number of times that copies of $\mathcal{E}_0$ appear in the random graph G; the same bounds hold as above.

The authors of [3] studied the related problem of counting the number of small cliques that occur in the random binomial graph, a case in which the dependence is not local; the technique applied is the Chen-Stein method.

Example 3.7 (*Local extremes.*) Let $\mathcal{G}_\alpha, \alpha \in \mathcal{A}$, be a collection of subgraphs of $\mathcal{G}$ isomorphic to $\mathcal{G}_0$, let $v \in \mathcal{V}_0$ be a distinguished vertex, let $\{C_g, g \in \mathcal{V}\}$ be a collection of independent and identically distributed random variables, and let X_α be defined by (50) with

$$X(C_\beta, \beta \in \mathcal{V}_0) = \mathbf{1}(C_v \geq C_\beta, \beta \in \mathcal{V}_0).$$

Then the sum $Y = \sum_{\alpha \in \mathcal{A}} X_\alpha$ counts the number of times the vertex in $\mathcal{G}_\alpha$ which corresponds under the isomorphism to the distinguished vertex $v \in \mathcal{V}_0$, is a local maxima. Corollary 3.3 holds with $M = 1$; the other quantities determining the bound begin dependent on the structure of $\mathcal{G}$.

For example, consider the hypercube $\mathcal{V} = \{0,1\}^p$ and $\mathcal{E} = \{\{v,w\} : ||v-w|| = 1\}$, where $||\cdot||$ is the Hamming distance (see also [1] and [2]). Take $v = \mathbf{0}$, $\mathcal{A} = \mathcal{V}$, and, for each $\alpha \in \mathcal{A}$, let $\mathcal{V}_\alpha = \{\beta : ||\beta - \alpha|| \le 1\}$ and $\mathcal{E}_\alpha = \{\{v,w\} : v,w \in \mathcal{V}_\alpha, ||v-w|| = 1\}$. Corollary 3.3 applies with $\rho = 2$ (by (47)), $V(r) = \sum_{j=0}^{r} \binom{p}{j}$, and (by (51))

$$B = 1 + p + \binom{p}{2} \quad \text{and} \quad \Delta \le 2^{-p/2} \sum_{j=0}^{2} \binom{p}{j} \sqrt{\sum_{j=0}^{6} \binom{p}{j}}.$$

4 Proofs of Theorems 1.1 and 1.2

In this section, $\mathcal{H}$ denotes a class of measurable functions satisfying properties (i),(ii), and (iii) (as described in Section 1), and h denotes an element of $\mathcal{H}$. Recall that δ is given by (6), let $\phi(t)$ denote the standard normal density, and, for $t \in (0,1)$, define

$$h_t(x) = \int h(x+ty)\phi(y)dy \quad \text{and} \quad \delta_t = \sup\{|\mathbf{E} h_t(W) - Nh_t| : h \in \mathcal{H}\}. \tag{54}$$

Lemma 4.1 *For a random variable W on $\mathbb{R}$, we have*

$$\delta \le 2.8\delta_t + 4.7at \quad \text{for all } t \in (0,1), \tag{55}$$

where a is as in (7). Furthermore, for all $A > 0$ and $\tilde{h}_\epsilon$ as in (8),

$$\mathbf{E}\left(\int \tilde{h}_{A+t|y|}(W)|\phi'(y)|dy\right) \le 2\delta + a(A+t). \tag{56}$$

Proof: Inequality (55) is Lemma 4.1 of [25], following Lemma 2.11 of [19], which stems from [5]. As in [25], adding and subtracting to the left hand side of (56) we have

$$\begin{aligned} &\mathbf{E}\left(\int (\tilde{h}_{A+t|y|}(W) - \tilde{h}_{A+t|y|}(Z))\,|\phi'(y)|dy + \int \tilde{h}_{A+t|y|}(Z)|\phi'(y)|dy\right) \\ \le\ & \int |\mathbf{E}\tilde{h}_{A+t|y|}(W) - \mathbf{E}\tilde{h}_{A+t|y|}(Z)|\,|\phi'(y)|dy + \int \mathbf{E}\tilde{h}_{A+t|y|}(Z)|\phi'(y)|dy \qquad (57) \\ \le\ & \left(2\delta + \int a(A+t|y|)|\phi'(y)|dy\right) \le 2\delta + a(A+t), \end{aligned}$$

where for the first term inside the parentheses in (57), we have used the facts that $h^{\pm}_{A+t|y|} \in \mathcal{H}$ and $\int |\phi'(y)|dy \le 1$. For the second term in the parentheses, we have used (7) and the fact that $\int |y||\phi'(y)|dy = 1$. ■

In Sections 4.1 and 4.2, h_t is given by (54) and f is the bounded solution of the Stein equation (2) with $\mu = 0, \sigma^2 = 1$, and test function h_t. With $||\cdot||$ the sup norm, Lemma 3 of [28] gives

$$||f|| \le \sqrt{2\pi} \le 2.6 \quad \text{and} \quad ||f'|| \le 4. \tag{58}$$

4.1 Proof of Theorem 1.1 (zero biasing)

Lemma 4.2 *Let Y be a mean-zero random variable with variance σ^2, and let Y^* be defined on the same space as Y, with the Y-zero biased distribution, satisfying $|Y^* - Y|/\sigma \le A$ for some A. Then*

$$\delta_t \le (6.6 + a)A + 2A^2 + \frac{1}{t}\left(2\delta A + aA^2\right) \quad \textit{for all } t \in (0, 1).$$

Proof: Let $W = Y/\sigma$, whence $W^* = Y^*/\sigma$ and $|W^* - W| \le A$. By differentiation in (2) and (54) respectively, we have

$$f''(x) = f(x) + xf'(x) + h_t'(x), \quad \text{with} \quad h_t'(x) = -\frac{1}{t}\int h(x+ty)\phi'(y)dy. \tag{59}$$

By (5) and (59), with $N_t h = \mathbb{E}\, h_t(Z)$ for Z a standard normal variable, we also have

$$\begin{aligned} |\mathbb{E}\, h_t(W) - Nh_t| &= |\mathbb{E}\,[f'(W^*) - f'(W)]| = |\mathbb{E}\int_W^{W^*} f''(x)dx| \\ &= |\mathbb{E}\int_W^{W^*} \left(f(x) + xf'(x) + h_t'(x)\right)dx|. \end{aligned} \tag{60}$$

Let $V = W^* - W$. Applying the triangle inequality in (60) and using (58), for the first term we find that

$$|\mathbb{E}\int_W^{W^*} f(x)dx| \le 2.6\mathbb{E}\,|V| \le 2.6A \tag{61}$$

and for the second term, again using (58), and, now, $\mathbb{E}\,|W| \le (\mathbb{E}\,W^2)^{1/2} = 1$, we find that

$$\begin{aligned} &\left|\mathbb{E}\int_W^{W^*} xf'(x)dx\right| \le 4\mathbb{E}\left|\int_W^{W+V} |x|dx\right| = 2\mathbb{E}\;|(W+V)|W+V| - W|W|| \\ \le\; &\mathbb{E}\,\left(4|WV| + 2V^2\right) \le 4A\mathbb{E}\,|W| + 2A^2 \le 4A + 2A^2. \end{aligned} \tag{62}$$

For the final term in (60), with $U \sim \mathcal{U}[0,1]$ independent of W and V, we write

$$|\mathbb{E}\int_W^{W^*} h_t'(x)dx| = |\mathbb{E}\,V\int_0^1 h_t'(W+uV)du| = |\mathbb{E}\,Vh_t'(W+UV)|.$$

Then, using (59), $\int \phi'(y)dy = 0$, and Lemma 4.1, we have

$$\begin{aligned} &|\mathbb{E}\,Vh_t'(W+UV)| = \frac{1}{t}|\mathbb{E}\,V\int h(W+UV+ty)\phi'(y)dy| \\ = \;& \frac{1}{t}|\mathbb{E}\,V\int [h(W+UV+ty) - h(W+UV)]\phi'(y)dy| \\ \le \;& \frac{1}{t}\mathbb{E}\left(|V|\int [h^+_{|V|+t|y|}(W) - h^-_{|V|+t|y|}(W)]\,|\phi'(y)|dy\right) \le \frac{1}{t}A\mathbb{E}\left(\int \tilde{h}_{A+t|y|}(W)|\phi'(y)|dy\right) \\ \le \;& \frac{1}{t}A\left(2\delta + a(A+t)\right) = \frac{1}{t}(2\delta A + aA^2) + aA. \end{aligned} \tag{63}$$

By combining the bounds (61), (62), and (63) we complete the proof. ■

Proof of Theorem 1.1: Letting $t = \alpha A$ in Lemma 4.2, we have

$$\delta_t \;\le\; (6.6+a)A + 2A^2 + \frac{1}{\alpha A}\left(2\delta A + aA^2\right) = (6.6 + a + \frac{a}{\alpha})A + 2A^2 + \frac{2\delta}{\alpha}. \tag{64}$$

Substituting (64) into the bound for δ given by Lemma 4.1, we have

$$\begin{aligned}\delta &\le 2.8((6.6+a+\frac{a}{\alpha})A + 2A^2 + \frac{2\delta}{\alpha}) + 4.7a\alpha A \\ &\le 18.5A + 2.8aA + 2.8\frac{aA}{\alpha} + 5.6A^2 + 5.6\frac{\delta}{\alpha} + 4.7a\alpha A,\end{aligned}$$

meaning that

$$\delta \le A\left(\frac{18.5 + 5.6A + 2.8a + 2.8a/\alpha + 4.7a\alpha}{1 - 5.6/\alpha}\right). \tag{65}$$

Setting $\alpha = 2\times 5.6$, for which $t<1$ since $A \le 1/12$, we obtain (9) and, hence, the theorem. ■

4.2 Proof of Theorem 1.2 (size biasing)

Lemma 4.3 *Let $Y \ge 0$ be a random variable with mean μ and variance σ^2, and let Y^s be defined on the same space as Y, with the Y-size-biased distribution, satisfying $|Y^s - Y|/\sigma \le A$ for some A. Then for all $t \in (0,1)$,*

$$\delta_t \le \frac{\mu}{\sigma}\left(\frac{4\Delta}{\sigma} + (3.3 + \frac{1}{2}a)A^2 + \frac{2}{3}A^3 + \frac{1}{2t}(2\delta A^2 + aA^3)\right), \tag{66}$$

with Δ as in (13).

Proof: With $W = (Y-\mu)/\sigma$, let $W^s = (Y^s - \mu)/\sigma$ (which is a slight abuse of notation). Then, $|W^s - W| \le A$. Note that

$$\mathbb{E}Wf(W) = \frac{\mu}{\sigma}(f(W^s) - f(W)), \tag{67}$$

and, so, with $V = W^s - W$, we have

$$\begin{aligned}\mathbb{E}h_t(W) - Nh_t = \mathbb{E}\left(f'(W) - Wf(W)\right) &= \mathbb{E}\left(f'(W) - \frac{\mu}{\sigma}(f(W^s) - f(W)\right) \\ = \mathbb{E}\left(f'(W) - \frac{\mu}{\sigma}\int_W^{W^s} f'(x)dx\right) &= \mathbb{E}\left(f'(W) - \frac{\mu}{\sigma}V\int_0^1 f'(W+uV)du\right) \\ = \mathbb{E}\left(f'(W) - \frac{\mu}{\sigma}Vf'(W)\right) &+ \mathbb{E}\left(\frac{\mu}{\sigma}Vf'(W) - \frac{\mu}{\sigma}V\int_0^1 f'(W+uV)du\right).\end{aligned} \tag{68}$$

Since $\mathbb{E}(\mu V/\sigma) = \mu\mathbb{E}(W^s - W)/\sigma = \mathbb{E}(\mu Y^s - \mu Y)/\sigma^2 = 1$, for the first expectation in (68) we have

$$\mathbb{E}\left\{f'(W)\mathbb{E}\left(1 - \frac{\mu}{\sigma}V|W\right)\right\} \le 4\frac{\mu}{\sigma}\sqrt{\mathrm{Var}(\mathbb{E}(W^s - W|W))} = 4\frac{\mu}{\sigma^2}\Delta, \tag{69}$$

using (58) and (13). Now, using (59), we write the second expectation in (68) as

$$\frac{\mu}{\sigma}V\left\{f'(W)-\int_0^1 f'(W+uV)du\right\} = \frac{\mu}{\sigma}V\int_0^1 (f'(W)-f'(W+uV))du$$

$$= -\frac{\mu}{\sigma}V\int_0^1\int_W^{W+uV} f''(v)dvdu = -\frac{\mu}{\sigma}V\int_0^1\int_W^{W+uV}(f(v)+vf'(v)+h_t'(v))dvdu. \quad (70)$$

We apply the triangle inequality and bound the three resulting terms separately. For the expectation arising from the first term on the right-hand side of (70), by (58) we have

$$|\mathbb{E}\{\frac{\mu}{\sigma}V\int_0^1\int_W^{W+uV} f(v)dvdu\}| \le 2.6\frac{\mu}{\sigma}\mathbb{E}\{|V|\int_0^1 u|V|du\} \le 1.3\frac{\mu}{\sigma}A^2 \quad (71)$$

and, for the second term, arguing as in (62) we have

$$|\mathbb{E}\frac{\mu}{\sigma}V\int_0^1\int_W^{W+uV} vf'(v)dvdu| \le 2\frac{\mu}{\sigma}\mathbb{E}|V|\int_0^1\left|\int_W^{W+uV}2|v|dv\right|du$$

$$\le 2\frac{\mu}{\sigma}\mathbb{E}|V|\int_0^1(2u|WV|+u^2V^2)du \le 2\frac{\mu}{\sigma}A\int_0^1(2Au\mathbb{E}|W|+u^2A^2)du$$

$$\le 2\frac{\mu}{\sigma}A(A+A^2/3). \quad (72)$$

For the last term in (70), the computation is more involved than, yet similar to, that for zero biasing. Beginning with the inner integral, we have

$$\int_W^{W+uV} h_t'(v)dv = uV\int_0^1 h_t'(W+xuV)dx$$

and using (59),

$$\int \phi'(y)dy = 0,$$

and Lemma 4.1, for the last term in (70) we have

$$|\frac{\mu}{\sigma}\mathbb{E}\int_0^1\int_0^1 uV^2h_t'(W+xuV)dxdu|$$

$$= \frac{\mu}{\sigma t}|\mathbb{E}V^2\int_0^1\int_0^1\int uh(W+xuV+ty)\phi'(y)dydxdu|$$

$$= \frac{\mu}{\sigma t}|\mathbb{E}V^2\int_0^1\int_0^1\int u[h(W+xuV+ty)-h(W+xuV)]\phi'(y)dydxdu|$$

$$\le \frac{\mu}{\sigma t}\mathbb{E}\left(V^2\int\int_0^1 u[h^+_{|V|+t|y|}(W)-h^-_{|V|+t|y|}(W)]\,|\phi'(y)|dudy\right)$$

$$= \frac{\mu}{2\sigma t}\mathbb{E}\left(V^2\int[h^+_{|V|+t|y|}(W)-h^-_{|V|+t|y|}(W)]\,|\phi'(y)|dy\right)$$

$$\le \frac{\mu}{2\sigma t}A^2\mathbb{E}\left(\int\tilde{h}_{A+t|y|}(W)|\phi'(y)|dy\right)$$

$$\le \frac{\mu}{2\sigma t}A^2\left(2\delta+a(A+t)\right)$$

$$= \frac{\mu}{2\sigma t}(2\delta A^2+aA^3)+\frac{\mu}{2\sigma}aA^2. \quad (73)$$

By combining (69), (71), (72), and (73) we complete the proof. ∎

Proof of Theorem 1.2 Applying Lemma 4.1 using the bound (66) on δ_t, we have

$$\delta \leq 2.8\frac{\mu}{\sigma}\left(\frac{4\Delta}{\sigma} + (3.3 + \frac{1}{2}a)A^2 + \frac{2}{3}A^3 + \frac{1}{2t}(2\delta A^2 + aA^3)\right) + 4.7at,$$

or,

$$\delta \leq \frac{2.8(\mu/\sigma)\left(4\Delta/\sigma + (3.3 + \frac{1}{2}a)A^2 + \frac{2}{3}A^3 + aA^3/2t\right) + 4.7at}{1 - 2.8\mu A^2/(\sigma t)}. \tag{74}$$

Setting $t = 2 \times 2.8\mu A^2/\sigma$, such that $t < 1$ since $A \leq (\sigma/(6\mu))^{1/2}$, (12) now follows from

$$\begin{aligned}\delta &\leq 5.6\frac{\mu}{\sigma}\left(\frac{4\Delta}{\sigma} + (3.3 + \frac{1}{2}a)A^2 + \frac{2}{3}A^3 + \frac{\sigma}{2(5.6\mu)}aA\right) + 2(4.7)a(5.6\frac{\mu A^2}{\sigma})\\ &\leq \frac{aA}{2} + \frac{\mu}{\sigma}\left((19 + 56a)A^2 + 4A^3\right) + 23\frac{\mu\Delta}{\sigma^2}. \quad ∎\end{aligned}$$

There are compromises in the choice of smoothing parameter; if we take $\alpha = 4 \times 5.6$ in (65) for $B \leq \sigma/48$, and $t = 4 \times 2.8\mu A^2/\sigma$ in (74) for $B \leq \sigma^{3/2}/(12\mu)^{1/2}$, bounds (9) and (12) become

$$\delta \leq A(145a + 7.5A + 25) \tag{75}$$

and

$$\delta \leq \frac{aA}{6} + \frac{\mu}{\sigma}\left((13 + 73a)A^2 + 2.5A^3\right) + 15\frac{\mu\Delta}{\sigma^2}, \tag{76}$$

respectively.

5 Remarks

The zero- and size-bias coupling both conform well to Stein's characterizing equation, and their use produces bounds on the distance of a random variable Y to the normal in many instances. The couplings are adaptable to the situation; in particular, the size-biased coupling, previously used in [17] for global dependence, is applied here to handle cases of local dependence.

The applications in Section 2 illustrate how bounds on the distance δ from Y to the normal can be generated using only a zero-bias coupling and a bound on $|Y^* - Y|$; in particular, the bounds do not depend on the often-difficult calculation of variances of conditional expectations of the form $\text{Var}\{\mathbb{E}(\tilde{Y} - Y|Y)\}$, which appear in the exchangeable-pair and size-biased versions of Stein's method when coupling Y to some $\tilde{Y}$. It is hoped that this feature of the zero-bias method will motivate a better understanding of the construction of couplings of Y^* to Y in greater generality than those that depend on the existence of the exchangeable pair of Proposition 2.1. In particular, the applications in Section 3 show an evidently wider scope of applicability of the size bias coupling over the zero bias one, as it is presently understood.

Acknowledgments

The author would like to thank Martin Raic and Qi-Man Shao for their insightful comments on an earlier version of this manuscript.

Bibliography

1. Baldi, P., Rinott, Y., and Stein, C. (1989) A normal approximation for the number of local maxima of a random function on a graph. Probability, statistics, and mathematics, 59-81, Academic Press, Boston, MA.

2. Baldi, P. and Rinott, Y. (1989) Asymptotic normality of some graph-related statistics. *J. Appl. Probab.* **26**, 171-175.

3. Barbour, A. D., Janson, S., Karoński, M., and Ruciński, A. (1990) Small cliques in random graphs. *Random Structures and Algorithms*, **1**, 403-434.

4. Bayer, D. and Diaconis, P. (1992) Trailing the Dovetail Shuffle to its Lair. *The Annals of Applied Probability*, **2**, 294-313.

5. Bhattacharya, R.N. and Ranga Rao R. (1986). *Normal approximation and asymptotic expansion*, Krieger, Melbourne, Fla.

6. Biggs, N. (1993). *Algebraic Graph Theory*, Cambridge University Press.

7. Bolthausen, E. (1984) An estimate of the reminder in a combinatorial central limit theorem. *Z. Wahrsch. Verw. Gebiete.*, **66**, 379-386.

8. Bolthausen, E. and Götze, F. (1993). The rate of convergence for multivariate sampling statistics. *Ann. Statist.* **21**, 1692-1710.

9. Brouwer, A.E., Cohen, A. M., and Neumaier, A. (1989). *Distance-Regular Graphs*, Springer-Verlag, Berlin.

10. Chen, Louis H.Y., and Shao, Q.M. (2004) Normal approximation under local dependence. *Ann Prob.* **32**, 1985-2028.

11. Darling, R. W. R., and Waterman, M.S. (1986) Extreme value distribution for the largest cube in a random lattice. *SIAM J. Appl. Math.* **46**, 118-132.

12. Darling, R. W. R., and Waterman, M.S. (1985) Matching rectangles in d dimensions: algorithms and laws of large numbers. *Adv. in Math.* **55**, 1-12.

13. Glaz, J., Naus, J., and Wallenstein, S. (2001) *Scan statistics.* Springer Series in Statistics. Springer-Verlag, New York.

14. Goldstein, L. (2004) Normal Approximation for Hierarchical Sequences *Annals of Applied Probability*, **14**, pp. 1950-1969.

15. Goldstein, L. and Reinert, G. (1997) Stein's Method and the Zero Bias Transformation with Application to Simple Random Sampling *Annals of Applied Probability*, **7**, 935-952.

16. Goldstein, L. and Reinert, G. (2004) Distributional transformations, orthogonal polynomials, and Stein characterizations. To appear, *Journal of Theoretical Probability.*

17. Goldstein, L. and Rinott, Y. (1996). On multivariate normal approximations by Stein's method and size bias couplings. *J. Appl. Prob.* **33**, 1-17.

18. Goldstein, L. and Rinott, Y. (2004). A permutation test for matching and its asymptotic distribution. *Metron*, **61** (2003), pp. 375-388.

19. Götze, F. (1991). On the rate of convergence in the multivariate CLT. *Annals of Probability*, **19**, 724-739.

20. Ho, S. T. and Chen, Louis H. Y. (1978) An L_p bound for the remainder in a combinatorial central limit theorem. *Ann. Probab.* **6**, 231-249.

21. Huang, H. (2002) Error bounds on multivariate normal approximations for word count statistics. *Adv. in Appl. Probab.* **34**, 559-586.

22. Kolchin, V.F., and Chistyakov, V.P. (1973) On a combinatorial limit theorem. *Theory Probability Appl.*, **18**, 728-739.

23. Naus, J. I. (1982) Approximations for distributions of scan statistics. *J. Amer. Statist. Assoc.* **77**, 177-183.

24. Rinott, Y. and Rotar, V. (1996). A multivariate CLT for local dependence with $n^{-1/2}\log n$ rate and applications to multivariate graph related statistics. *J. Multivariate Anal.*, **56**, 333-350.

25. Rinott, Y. and Rotar, V. (1997). On coupling constructions and rates in the CLT for dependent summands with applications to the antivoter model and weighted U-statistics. *Ann. Appl. Probab.*, **7**, 1080-1105.

26. Stein, C. (1972). A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. Proc. Sixth Berkeley Symp. Math. Statist. Probab. **2**, 583-602, Univ. California Press, Berkeley.

27. Stein, C. (1972). Estimation of the mean of a multivariate normal distribution. *Ann. Probab.*, **9**, 1135-1151.

28. Stein, C. (1986). *Approximate Computation of Expectations.* IMS, Hayward, CA.

29. von Bahr, B. (1976) Remainder term estimate in a combinatorial limit theorem. *Z. Wahrsch. Verw. Gebiete.*, **35**, 131-139.