Some examples of Normal approximations by Stein’s method

Amir Dembo* and Yosef Rinott†

May 27, 1994

Abstract

Stein’s method is applied to study the rate of convergence in the normal approximation for sums of non-linear functionals of correlated Gaussian random variables, for the exceedances of r-scans of i.i.d. random variables, and in a multinomial setting.

Keywords and phrases: Stein’s method, r-scans, exceedances of Gaussian processes, central limit theorems.

MSC 1991 Classification: 60F05, 60G15

*Department of Statistics and Department of Mathematics, Stanford, CA 94305. This work was supported in part by NSF grant DMS92-09712 and by a US-ITAL, RSF grant.

†Department of Mathematics, UCSD, La Jolla, CA 92037. This work was supported in part by NSF grant DMS 92-05759.
1 Introduction

The purpose of this paper is to demonstrate Stein’s Normal approximation method and its potential use, by applying a variant of it to three examples of random variable of the form $W = \sum_{i=1}^n \psi(U_i)$. In Section 2 the random variables $U_i$ are correlated and jointly Gaussian, while in Section 3 they have the multinomial distribution, and in both cases $\psi$ is typically a non-linear function satisfying appropriate growth conditions. Because of the global dependence among the summand variables in both examples, we find the normal approximation theorem of [8, Theorem 1.1] most suitable here. The bound on the normal approximation rate of a random variable $W$ appearing in this theorem is expressed in terms of another random variable $W^*$, whose marginal distribution is the $W$–size biased distribution. The contribution of this paper is in the explicit construction of the $W$–size biased variable $W^*$, which is required for the application of this theorem in each case, and in showing that the method yields a rather tight upper bounds on the rate of convergence (see Theorems 2.3 and 3.1 below). We shall compare our results to the existing literature on these problems. In Section 4 we analyze the number of exceedances of the $r$-scans generated by $n$ i.i.d. random variables, for the regime in which the Poisson approximation does not apply. In contrast with the previous two examples, here $U_i$ are $r$-dependent, hence the dependence structure is local, allowing for the direct applicability of Stein’s (1986) original normal approximation theorem. The analysis of the $r$-scan example, follows from a new version of Stein’s theorem which provides a useful bound on $|\Pr((W-EW)/\sqrt{\text{Var}W} \leq w) - \Phi(w)|$ in situations where $E[\psi(U)] \to 0$ as $n \to \infty$, i.e. on the “boundary” between the Normal and Poisson approximations (see Theorem 4.2 below). We also show how the rate of convergence in the normal approximation for exceedances of $r$-spacings on the sphere can then be deduced.

2 Additive functionals of correlated Gaussian random variables

In this section we provide an application of the following normal approximation theorem from [8, Theorem 1.1], which is an extension of an earlier result, [1, Lemma 2.4].

**Theorem 2.1** Let $W \geq 0$ be a random variable with distribution $dF(w)$ and let $W^*$ be defined on the same probability space as $W$, and having the marginal distribution $d\psi F(w)/\lambda$, where $\lambda = EW$, and $\sigma^2 = \text{Var} W$. Then for any twice piecewise continuously differentiable $h : R \to R$,

$$|Eh(W - \lambda)/\sigma - Nh| \leq 2||h|| \frac{\lambda}{\sigma^2} \sqrt{\text{Var}E(W^* - W | W)} + ||h'|| \frac{\lambda^3}{\sigma^3} E(W^* - W)^2,$$

where $Nh = \int h(x)\phi(x)dx$, $\phi(\cdot)$ denotes the standard normal probability density and $|| \cdot ||$ denotes throughout the supremum-norm.

Note that the distribution of $W^*$ in Theorem 2.1 is well known as the $W$–size biased (or length biased) distribution in other contexts. See [8] for a detailed discussion and references. For completeness, we provide here a short proof of this theorem.

**Proof:** Let $Z = \lambda W/\sigma^2$ and $\mu = EZ = \lambda^2/\sigma^2$. Note that $\text{Var}Z = \mu$ and that $Z^* = \lambda W^*/\sigma^2$ has the $Z$–size biased distribution. Let $Y = Z^* - Z$ and observe that for all $g : R \to R$

$$\mu E_{g}(Z + Y) = EZ_{g}(Z).$$

(2)

In particular, $EY = \text{Var}Z/EZ = 1$. Subtract $\mu E_{g}(Z)$ from both sides of (2) to obtain by a Taylor series expansion of $g(Z + Y) - g(Z)$, that

$$E\left(\frac{Z - \mu}{\mu} g(Z)\right) = E(g(Z + Y) - g(Z)) = E(g'(Z) + (Y - 1)g'(Z) + \frac{Y^2}{2} g''(\xi)),$$
for some $\xi$ between $Z$ and $Z + Y$. Consequently,

$$|E(g'(Z) - \frac{Z - \mu}{\mu}g(Z))| \leq \|g\| \|E[|Y - 1|Z]\| + \frac{1}{2}\|g''\|EY^2.$$  

For a given continuous and piecewise continuously differentiable $h : \mathbb{R} \to \mathbb{R}$, let $f$ denote the unique solution of the differential equation $f'(x) - xf(x) = h(x) - Nh$ with $\lim_{x \to \infty} f(x)e^{-x^2/2} = 0$, and set $g(z) = \sqrt{\mu}f(\frac{z - \mu}{\sqrt{\mu}})$. Let $\hat{Z} = (Z - \mu)/\sqrt{\mu} = (W - \lambda)/\sigma$ and note that

$$Eh(\hat{Z}) - Nh = E(f'(\hat{Z}) - \hat{Z}f(\hat{Z})) = E(g'(Z) - \frac{Z - \mu}{\mu}g(Z)).$$

Since $\|g\| = \|f\| \leq 2\|h\|$ ([1, (2.17)]) and $\|g''\| = \|f''\|/\sqrt{\mu} \leq 2\|h''\|/\sqrt{\mu}$ ([20, page 25]), it follows that

$$|Eh(\hat{Z}) - Nh| \leq 2\|h\|E[|Y - 1|Z]\| + \frac{\|h''\|}{\sqrt{\mu}}EY^2.$$  

Since $EY = 1$, by the Cauchy-Schwarz inequality $E[|Y - 1|Z]\| \leq \sqrt{\text{Var}E(Y|Z)}$, and (1) follows from (3).  

**Remark 2.1** Applying (2) to $g(z) = z^2$ it follows by simple manipulations and the Cauchy-Schwarz inequality that

$$\frac{EY^2}{\sqrt{\mu}} \leq \mu^{-3/2}|EZ^3 - \mu^3 - 3\mu^2| + 2\sqrt{\text{Var}E(Y|Z)}.$$  

It is easy to check that the first term on the r.h.s vanishes for $Z \sim N(\mu, \mu)$, and equals $1/\sqrt{\mu}$ for $Z \sim \text{Poisson}(\mu)$. In the latter case, $Y = 1$ almost surely, with (3) implying that $|Eh(\hat{Z}) - Nh| \leq \|h''\|/\sqrt{\mu}$. Indeed, insight to the nature of Theorem 2.1 may be gained by comparing (3) with the inequality

$$|Eh(Z) - Pois(h)| \leq 2(1 - e^{-\mu})\|h\|E[Y - 1]|$$  

which holds for every $Z \in \mathbb{Z}_+$ with $\mu = EZ$. The inequality (4) is the key to the “coupling approach” in the context of the Chen-Stein Poisson approximation method (see [3, Theorem 2.1]) or [20, page 93]. With $\|g\|_L = \sup_{z \in \mathbb{R}} |g(z + 1) - g(z)|$, a short derivation of (4) starts by subtracting $\mu E g(Z + 1)$ from both sides of (2), then using the bound $|E(g(Z + Y) - g(Z + 1))| \leq \|g\|_L E[Y - 1]$, and finally considering $g : \mathbb{N} \to \mathbb{R}$ which is the solution of $zg(z) - \mu g(z + 1) = h(z) - Pois(h)$ (for then $\|g\|_L \leq 2(1 - e^{-\mu})\|h\|/\mu$, see [3] or [20, page 89]).

In order to apply Theorem 2.1 we have to construct a random variable $W^*$ having the required marginal distribution wdf$(w)/\lambda$ (and with small $\text{Var}E(W^* - W|W)$). For a more general treatment and details of such constructions, see [8]. The following lemma from [8] and the subsequent discussion provide guidelines leading to useful constructions in the setup of this paper.

**Lemma 2.1** Let $U = (U_1, \ldots, U_n)$ be a random vector, and let $\psi_i$ be nonnegative functions such that $E\psi_i(U_i) < \infty$, $i = 1, \ldots, n$. Let $Y^{(i)} = (Y^{(i)}_1, \ldots, Y^{(i)}_n)$ satisfy $P(Y^{(i)} \in dy) = P(U \in dy)\psi_i(y_i)/E\psi_i(U_i)$. Let $I$ be a random variable taking values in $\{1, \ldots, n\}$, distributed independently of all the above variables, with $P(I = i) = E\psi_i(U_i)/\sum_{j=1}^n E\psi_j(U_j)$.

Let $W = \sum_{j=1}^n \psi_j(U_j)$ have the distribution $F$. Then $W^* = \sum_{j=1}^n \psi_j(Y^{(i)}_j)$ has the distribution wdf$(w)/\lambda$, where $\lambda = EW$.
Note that the distribution of $Y^{(i)}$ can be described as the distribution of $U$ conditioned on a value of $U_i$, where $U_i$ has a new marginal distribution, $\psi_i(\cdot)\text{d}F_{U_i}(\cdot)/E\psi_i(U_i)$.

Given $W = \sum_{j=1}^{n} \psi_j(U_j)$, this suggests the following construction of $W^*$: Let $V_i$ be independent random variables, $V_i \sim \psi_i(\cdot)\text{d}F_{U_i}(\cdot)/E\psi_i(U_i)$. If $V_i$ is assigned the value $u$, let $(Y_1^{(i)}, \ldots, Y_{i-1}^{(i)}), Y_i^{(i)}, Y_{i+1}^{(i)}, \ldots, Y_n^{(i)}$ have the conditional distribution of $(U_1, \ldots, U_{i-1}, U_{i+1}, \ldots, U_n)$ given $U_i = u$, and set $Y_i^{(i)} = u$. Now choose $I$ and set $W^* = \sum_{j=1}^{n} \psi_j(Y_j^{(I)})$ as in Lemma 2.1.

**Proof of Lemma 2.1** We prove that for any function $G$, $EG(W^*) = EWG(W)/\lambda$, where $\lambda = EW = \sum_{j=1}^{n} E\psi_j(U_j)$. Indeed, we have,

$$EG(W^*) = E \sum_{i=1}^{n} G(\sum_{j=1}^{n} \psi_j(Y_j^{(I)}))E\psi_i(U_i)/\lambda$$

$$= \int \sum_{i=1}^{n} G(\sum_{j=1}^{n} \psi_j(y_j))P(U = dy)\psi_i(y_i)/\lambda$$

$$= E \sum_{i=1}^{n} \psi_i(U_i)G(\sum_{j=1}^{n} \psi_j(U_j))/\lambda = EWG(W)/\lambda.$$ 

**Remark 2.2** (a) For $W = \int_T \psi_x(U_x)dm(x)$ with a finite, non-negative measure $m(\cdot)$ on $T$, the $W*$-size biased construction $W^* = \int_T \psi_x(Y_x^{(I)})dm(x)$ applies as in Lemma 2.1. Here, I has the law $\text{d}P_t(x) = E(\psi_x(U_x))dm(x)/\lambda$ while $P(Y(x) = dy) = P(U = dy)\psi_x(y_x)/E\psi_x(U_x)$.
(b) While Lemma 2.1 is stated and proved for real-valued $U_i$, it easily extends to $U_i \in \mathbb{R}^d$. In particular, at the cost of somewhat more cumbersome computations, one can extend Theorems 2.3 and 3.1 below, so as to handle variables of the form $W = \sum_{i=1}^{n} \psi_j(U_j^{(1,i)}, \ldots, U_j^{(m,i)})$, with fixed $m > 1$ and $j : \{1, \ldots, m\} \times \{1, \ldots, n\} \rightarrow \{1, \ldots, n\}$ of an appropriate structure.
(c) Note that the construction of Lemma 2.1 is restricted to nonnegative functions $\psi_i$. Functions $\psi$ which are bounded below (and not necessarily non-negative), are handled by translation prior to the construction of $W^*$ as in Lemma 2.1. The only effect of such a translation on the conclusion of Theorem 2.1 is in increasing the value of $\lambda$. A multivariate version of Theorem 2.1 is given in [8, Theorem 1.3]. From that result one can obtain a bivariate normal approximation for $(\sum_{j=1}^{n} \psi_j^{+}(U_j)), (\sum_{j=1}^{n} \psi_j^{-}(U_j))$, where $\psi^+, \psi^-$ denote positive and negative parts. This will lead to a normal approximation for $\sum_{j=1}^{n} \psi_j(U_j)$ without the assumption that $\psi_i$ are nonnegative. This approach will not be pursued here. However we comment that in some cases it leads to sharp approximation rates, while in other cases, the dependence structure of the positive and negative parts may be different from that of the original functions in a way that leads to unsatisfactory bounds.

The following (somewhat forgotten) fact appears in [15] (for further references see, e.g., [2]).

**Lemma 2.2** Let $X = (X_1, \ldots, X_k)$ and $Y = (Y_1, \ldots, Y_m)$ be random vectors such that the joint distribution of $(X, Y)$ is multivariate normal. Let $\rho = \rho(X; Y)$ denote the (first) canonical correlation between $X$ and $Y$, that is, $\rho = \max \text{Corr}(a^T X, b^T Y)$, where the maximum is taken over all vectors $a \in \mathbb{R}^k$ and $b \in \mathbb{R}^m$. Then $\text{Corr}(f(X), g(Y)) \leq \rho$ for any pair of real valued functions $f$ and $g$ such that the correlation is defined.

For completeness we provide the proof of the following easy extension of Lemma 2.2 which is a special case of [4, page 40, Theorem 6.2].

**Lemma 2.3** Let $(X, Y)$ be multivariate normal vectors, and $V_1, V_2$ independent random vectors which are also independent of $(X, Y)$. Then, when defined,

$$\text{Cov}(f(X, V_1), g(Y, V_2)) \leq \rho(X; Y)\sqrt{\text{Var} f(X, V_1) \text{Var} g(Y, V_2)}.$$  

(5)
Proof With \( f = f(X, V_1), \) and \( g = g(Y, V_2) \) we have
\[
\text{Cov}(f, g) = \text{Cov}(E[f \mid V_1, V_2], E[g \mid V_1, V_2]) + E[\text{Cov}(f, g \mid V_1, V_2)].
\] (6)

Since \( f \) is independent of \( V_2 \) and \( g \) is independent of \( V_1 \), the first term on the r.h.s. of (6) vanishes by the independence of \( V_1 \) and \( V_2 \). Invoking Lemma 2.2 for \( f \mid V_1, V_2 \) and \( g \mid V_1, V_2 \), by the independence of \( V_1 \)'s we have
\[
E[\text{Cov}(f, g \mid V_1, V_2)] \leq \rho(X; Y)E[\sqrt{\text{Var}(f \mid V_1)}]E[\sqrt{\text{Var}(g \mid V_2)}]
\leq \rho(X; Y)E[\sqrt{\text{Var}(f \mid V_1)}]E[\sqrt{\text{Var}(g \mid V_2)}].
\]

Ignoring the conditioning leads to (5).

We now state our first result, a normal approximation theorem suitable for smooth \( \psi \) (see Theorem 2.3 below for the more general case).

Theorem 2.2 Let \( U = (U_1, \ldots, U_n) \) have the multivariate normal distribution \( N(0, \Sigma) \), where the diagonal elements of \( \Sigma = \rho_{ii} \) are \( \rho_{ii} = 1 \). Set \( \rho_{ij}^* = \rho_{ij}1_{|\rho_{ij}|<1} \) and
\[
\tilde{\rho}_{ij} = \min\left\{ \max\left\{ \frac{\rho_{ii'}}{\rho_{jj'}}, \frac{|\rho_{ij}|}{|\rho_{ij'}|}, \frac{|\rho_{ij'}}{|\rho_{ij}|} \right\}\right\}^{1/2}, 1 \right\}, \]

Let \( W = \sum_{i=1}^n \psi(U_i) \) where \( \psi \geq 0 \) is scaled such that \( NW = 1 \) (hence \( EW = n \)). Denote \( \sigma^2 = \text{Var}W \) and \( \phi(p, \cdot, \cdot) \) the bivariate normal density with standard marginals and correlation \( \rho \). Define
\[
M(\rho) = \int \int \int [\psi(u_1 + \rho(v - w_2)) - \psi(u_1)]^2 \phi_p(u_1, u_2) \psi(v) \psi(v) du_1 du_2 dv.
\]

Then for any piecewise continuously differentiable \( h \),
\[
|Eh(W - n/\sigma) - Nh| \leq \frac{2\|h\|}{\sigma^2} \left\{ \sum_{i=1}^n \sum_{j=1}^n \sum_{i'=1}^n \sum_{j'=1}^n \tilde{\rho}_{ij}^{ij'} [M(\rho_{ij})M(\rho_{ij'})]^{1/2} \right\}^{1/2}
+ \frac{\|h'\|}{\sigma^3} \sum_{i=1}^n \left\{ M(\rho_{ij})^{1/2} \right\}^2. \tag{7}
\]

Remark 2.3 Note that there is no loss of generality in our assumption of \( NW = 1 \). One could easily extend Theorem 2.2 to the case of distinct functions \( \psi_i(U_i) \), at the expense of somewhat more cumbersome notation. Similarly, following part (a) of Remark 2.2, Theorem 2.2 extends to \( W = \int_T \psi(U_t) \text{d}m(t) \) with \( U_t \) a centered Gaussian process indexed by a parameter set \( T \) and scaled such that \( E(U_t^2) = 1 \). Here, \( NW = 1 \) results with \( n = m(T) \) and the summations in the r.h.s. of (7) become integrals with respect to \( m(\cdot) \). In particular, one can thus extend Corollary 2.1 below to the current setting (with \( \rho = 0 \) in (14)) , provided that \( \sup_T |\rho_{ts}|(1 - |\rho_{ts}|)^{-1/2} dm(s) = B < \infty \).

Proof We construct a random variable \( W^* \) having the \( W \)-size biased distribution as suggested by Lemma 2.1. Specifically, here
\[
Y^{(i)} = U_i + \rho_{ii}(V_i - U_i), \ldots, U_{i-1} + \rho_{i-1, i}(V_i - U_i), V_i, U_{i+1} + \rho_{i+1, i+1}(V_i - U_i), \ldots, U_n + \rho_{nn}(V_i - U_i),
\]
where \( V_i \sim \psi(u)P(U_i \in du) \) independently of all other variables. A direct calculation shows that \( Y^{(i)} \), given \( V_i = u \) is distributed like \( U \) given \( U_i = u \). (Ignoring the \( i \)th component, this conditional distribution is \( N(\rho_i u, \Sigma_i - \rho_i^2) \) where \( \Sigma_i \) denotes the principal submatrix.
obtained from \( \Sigma \) by deleting its \( i \)th row and column, and \( \rho_i^T = (\rho_{i1}, \ldots, \rho_{i \, i-1}, \rho_{i \, i+1}, \ldots, \rho_{i \, m}) \). Consequently, by Lemma 2.1, for \( I \) a random variable which is uniform on \{1, \ldots, n\} and independent of all other variables, \( W^* = \sum_{j=1}^{n} \psi(Y_j^{(I)}) \) and \( W \) are defined on a joint space, and \( W^* \) has the \( W \)-size biased distribution.

In order to apply Theorem 2.1, define \( A_{ij} = \psi(U_j + \rho_{ij}(V_i - U_i)) - \psi(U_j) \), and note that

\[
E[W^* - W|W] = \frac{1}{n} E[\sum_{i=1}^{n} \sum_{j=1}^{n} \psi(U_j + \rho_{ij}(V_i - U_i)) - \psi(U_j)|W] = \frac{1}{n} E[\sum_{i=1}^{n} \sum_{j=1}^{n} A_{ij}|W]
\]  

(8)

To compute the variance of this expression note that by ignoring the conditioning on \( W \)

\[
\text{Var}[\sum_{i=1}^{n} \sum_{j=1}^{n} A_{ij}|W] \leq \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{i'=1}^{n} \sum_{j'=1}^{n} \text{Cov}(A_{ij}, A_{i'j'})
\]  

(9)

To deal with the typical case of \( i \neq i' \neq j \neq j' \) we use Lemma 2.3 to exploit both the possible smallness of the correlations between the terms \( A_{ij} \) and \( A_{i'j'} \) and the possible smallness of the terms themselves. Indeed, by Lemma 2.3

\[
\text{Cov}(A_{ij}, A_{i'j'}) \leq \rho(U_i, U_j; U_{i'}, U_{j'}) \sqrt{\text{Var}A_{ij} \text{Var}A_{i'j'}}.
\]  

(10)

where \( \rho(U_i, U_j; U_{i'}, U_{j'}) \) denotes the canonical correlation. Note that \( \text{Var}(A_{ij}) \leq E(A_{ij}^2) \leq M(\rho_{ij}) \). A direct calculation shows that \( \rho(U_i, U_j; U_{i'}, U_{j'}) \leq \rho_{ij,ij'} \). (See [16] for a similar calculation.) Consequently, by (8), (9) and (10),

\[
\text{Var}[W^* - W|W] \leq \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{i'=1}^{n} \sum_{j'=1}^{n} \rho_{i,ij'} \rho_{i',jj'} \sqrt{\text{Var}A_{ij} \text{Var}A_{i'j'}}
\]  

(11)

Turning now to bound the term \( E(W^* - W)^2 \) note that

\[
E(W^* - W)^2 = \sum_{i=1}^{n} \frac{1}{n} E(\sum_{j=1}^{n} [\psi(U_j + \rho_{ij}(V_i - U_i)) - \psi(U_j)]^2 = \frac{1}{n} \sum_{i=1}^{n} E(\sum_{j=1}^{n} A_{ij}^2).
\]

By the Cauchy-Schwarz inequality

\[
E(\sum_{j=1}^{n} A_{ij}^2) = \sum_{j=1}^{n} \sum_{j'=1}^{n} E[A_{ij} A_{ij'}] \leq \sum_{j=1}^{n} \sum_{j'=1}^{n} E[A_{ij}^2]^{1/2} E[A_{ij'}^2]^{1/2},
\]

implying that

\[
E(W^* - W)^2 \leq \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} E[A_{ij}^2]^{1/2} E[A_{ij'}^2]^{1/2} \leq \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} M(\rho_{ij})^{1/2}^2,
\]

(12)

Since \( E(W) = n \), combining the bounds of (11) and (12) with Theorem 2.1 leads to (7).

In Corollary 2.1 below we use the following estimate on \( M(\rho) \) to make the bound of Theorem 2.2 more explicit for smooth \( \psi \).

**Proposition 2.1** If \( N \psi = 1 \) and for some \( q < 2 \) and \( K < \infty \)

\[
|\psi(y) - \psi(x)| \leq K|y - x| e^{K|x|^q + |y|^q} \quad \forall x, y \in \mathbb{R},
\]

(13)

then \( M(\rho) \leq C \rho^2 \) for some constant \( C(K, q) < \infty \) which depends only on \( K \) and \( q \).

**Remark 2.4** In particular, Proposition 2.1 applies when \( \psi \) has a bounded derivative, that is, \( |\psi'(u)| \leq K \) or more generally is uniformly Lipschitz continuous, i.e. \( |\psi(y) - \psi(x)| \leq K|y - x| \).
Proof Let \((U_1, U_2)\) denote the bivariate normal vector of unit variances and correlation \(\rho\) and \(U\) an additional standard normal variable independent of \((U_1, U_2)\). Then, by the assumption
\[
M(\rho) = E\{[\psi(U_1 + \rho(U - U_2)) - \psi(U_1)]^2 \psi(U)\} \leq K^2 \rho^2 E[|U - U_2|^2 e^{2K(||U_1||^2 + ||U - U_2||^2)}] \psi(U)\}.
\]
By the Cauchy-Schwarz inequality
\[
K^2 E[|U - U_2|^2 e^{2K(||U_1||^2 + ||U - U_2||^2)}] \psi(U)\} \leq K^2 E[e^{8K||U_1||^2} E[|U - U_2|^2 e^{8K||U - U_2||^2}]}^{1/4} E[\psi(U)]^{1/2}
\]
The r.h.s. is bounded by \(C(K, q) < \infty\) which is independent of \(\rho\) and \(\psi\), since
\[
E[\psi(U)^2] = 1 + E[(\psi(U) - \psi(U_1))^2] / 2 \leq 1 + K^2 E[|U - U_1|^2 e^{2K(||U||^2 + ||U_1||^2)}] / 2.
\]

Corollary 2.1 Under the conditions of Theorem 2.2, assume \(\max \sum_{j=1}^n |\rho_{ij}| \leq \beta < \infty \) and \(\max i \sum_{j=1}^n |\rho_{ij}| \leq \beta < \infty \). Let \(\psi \geq 0\) with \(N\psi = 1\) be such condition (13) holds and define \(D = \max \{||h||, ||h'||\}\). Then, with \(C - C(K, q) < \infty\) from Proposition 2.1,
\[
|Eh(W - \sigma n) - Nh| \leq \frac{4DB}{\sigma^2} \left\{ \frac{2nBC}{1 - r} \right\}^{1/2} + \frac{DCB^2n}{\sigma^3}.
\]

Proof This is a direct consequence of Theorem 2.2 and Proposition 2.1. Indeed, we use the simple bound
\[
\tilde{\rho}_{ij'}j' \leq \frac{2}{1 - r}(|\rho_{ii'}| + |\rho_{ij'}| + |\rho_{ji'}| + |\rho_{ij}|)
\]
which applies also when \(i = j\) or \(i' = j\)' (see the definition of \(\tilde{\rho}_{ij}^n\) in Theorem 2.2). Then, since \(M(\rho) \leq C\rho^2\) the r.h.s. of (7) is further bounded by
\[
\frac{2D}{\sigma^2} \left\{ \frac{2C}{1 - r} \sum_{i=1}^n \sum_{j=1}^n \sum_{i'=1}^n \sum_{j'=1}^n |\rho_{ij}| |\rho_{ij'}| (|\rho_{ii'}| + |\rho_{ij'}| + |\rho_{ji'}| + |\rho_{ij}|) \right\}^{1/2} \leq \frac{2DC}{\sigma^3} \left\{ \frac{2nBC}{1 - r} \right\}^{1/2} + \frac{DCB^2n}{\sigma^3}.
\]

For a triangular array or a stationary sequence \(\{U_i\}\), we set \(B = B_n\) and \(\sigma = \sigma_n\) in Corollary 2.1 and assume for simplicity that \(r\) and \(\psi\) do not depend on \(n\). Then, since \(\sigma^2 \leq nB^2 \text{Var}(\psi(U))\) (by Lemma 2.2), the second term on the r.h.s. of (14) dominates, leading to the rate of \(nB^2\sigma_n^2/\sigma_n^3\). The bound of (14) is then non-trivial only when \(n^{-1}B_n \to 0\) with asymptotic normality holding for \(W\) provided \(n^{-1/3}B_n^{-2/3}\sigma_n \to \infty\) (note that unlike \(B_n\), the value of \(\sigma_n\) depends on the function \(\psi\)).

Much is known about the asymptotic normality of \(W\) for a stationary sequence \(\{U_i\}\) and \(\psi\) such that \(N\psi < \infty\). In this context, for \(r_i = E(U_0U_i)\) which is both square summable and Fejér summable (i.e., \(\sum_{i=1}^n (1 - i/n)r_i\) has a finite limit), it is known that \(n^{-1/2}\sigma_n\) has a finite and typically positive limit and that \(W\) is then asymptotically normal (see [21, Theorem 1] and the references therein). In this case Corollary 2.1 supplies the rate of \(n^{-1/2}(\sum_{i=0}^n |r_i|^2)^{1/2}\) in the Normal approximation for \(W\). In particular, a rate of \(n^{-1/2}\) results when \(\{r_i\}\) is absolutely summable. When \(\sum r_i^2 = \infty\), in general the asymptotic normality of \(W\) fails if the Hermite rank of \(\psi - N\psi\) is not \(1\) (see [21, Example 1] and the references therein to the earlier works of [22] and [7]). For \(\psi - N\psi\) of Hermite rank \(1\) and \(r_i = i^{-\alpha}L(i)\), with \(\alpha \in (0, 1)\) and \(L(i)\) a slowly varying function, it is known (see [22]) that \(\sigma_n = O(n^{1-\alpha/2}L(n)^{1/2})\). With \(L(n)\) now denoting any slowly varying function we have \(B_n = O(n^{1-\alpha}L(n))\), and Corollary 2.1 supplies the rate of \(n^{-\alpha/2}L(n)^{1/2}\) for the Normal approximation of \(W\).

The following lemma is key for extending the Normal approximation of Corollary 2.1 to non-smooth \(\psi\) (e.g., \(\psi(u) = 1_{[u, \infty)}(u)\) corresponding to a count of large exceedances of the Gaussian process).
Lemma 2.4 Let \( f(x, z) = \psi(x_1)[\psi(x_3 + z(x_1 - x_2)) - \psi(x_3)] \), where \( x = (x_1, x_2, x_3) \), and \( \psi \geq 0 \) is scaled such that \( N\psi = 1 \) and satisfying \( \psi(u) \leq Ke^{K|u|^q} \) for some \( 0 < K < \infty \) and \( q < 2 \). Let \( X = (X_1, X_2, X_3) \) and \( Y = (Y_1, Y_2, Y_3) \) be normal vectors, having zero means and unit variances, such that \( X_1 \) and \( Y_1 \) are independent of each other and of \( X_2, X_3, Y_2, Y_3 \). Define \( r_x = \rho(X_2, X_3) \), \( r_y = \rho(Y_2, Y_3) \) and \( \rho = \max_{i,j} |\text{Corr}(X_i, Y_j)| \) and suppose that \( \max\{|r_x|, |r_y|\} + 2\rho \leq s < 1 \). Then, there exists a constant \( C = C(s, K, q) < \infty \) such that for any \( a, b \in [-1, 1] \)

\[
\text{Cov}(f(X, a), f(Y, b)) \leq C|ab|,
\]

and

\[
|Ef(X, a)f(Y, b)| \leq C|ab|.
\]

Proof Fixing the (marginal) distributions of the vectors \( X \) and \( Y \), let \( \rho_{i,j} = \text{Corr}(X_i, Y_j) \) and let \( \phi_\rho(x, y) \) denote the joint normal density of \( X \) and \( Y \), having marginals equal those corresponding to the laws of \( X \) and of \( Y \). Clearly,

\[
\text{Cov}(f(X, a), f(Y, b)) = \int f(x, a)f(y, b)[\phi_\rho(x, y) - \phi_0(x, y)]dx dy.
\]

Let

\[
g(x, z) = \int_{x_3}^{x_3 + z(x_1 - x_2)} \psi(x_1) \psi(t) dt,
\]

and observe that \( f(x, z) = \frac{\partial}{\partial x_3} g(x, z) \). Note that for any \( \xi \in [0, 1] \) the minimal eigenvalue of the covariance matrix for \( \phi_\rho \) is at least \( 1 - 2\rho - \max\{|r_x|, |r_y|\} \geq (1 - s) \). In particular, both \( \phi_\rho \) and \( \phi_0 \) are non-degenerate, and since

\[
|g(x, a)| \leq K^2|a||x_1 - x_2|e^{2K(|x_1| + |x_2| + |x_3|)}
\]

boundary terms vanish, and by integration by parts we obtain

\[
\text{Cov}(f(X, a), f(Y, b)) = \int g(x, a)g(y, b) \frac{\partial^2}{\partial x_3 \partial y_3} [\phi_\rho(x, y) - \phi_0(x, y)] dx dy.
\]

Noting that \( \frac{\partial^2}{\partial x_3 \partial y_3} \phi_\rho(x, y) = \frac{\partial}{\partial x_3} \phi_\rho(x, y) \) for any vector \( \rho \) (Heat equation) we have that for some \( \xi = \xi(x, y) \) such that \( 0 \leq \xi \leq 1 \)

\[
\frac{\partial^2}{\partial x_3 \partial y_3} [\phi_\rho(x, y) - \phi_0(x, y)] = \sum_{i,j} \rho_{i,j} \frac{\partial^2}{\partial x_i \partial y_j} \phi_\rho(x, y).
\]

Fixing \( \xi \) and \( \rho \), the ratio \( \frac{\partial^2}{\partial x_i \partial y_j} \phi_\rho(x, y) \) is a polynomial expression of order of at most 4 in \( (x, y) \) and of order at most 4 in the entries of the inverse of the covariance matrix for \( \phi_\rho \). As such it is bounded by \( c(1 - s)^{-1}(1 + |x|^4 + |y|^4) \) for some universal constant \( c \). It thus follows by (15) and (16) that for some \( c_1 = c_1(K, s) < \infty \),

\[
\text{Cov}(f(X, a), f(Y, b)) \leq c_1|a||b|\rho \int |x_1 - x_2||y_1 - y_2|(1 + |x|^4 + |y|^4)e^{2K(|x_1| + |x_2| + |x_3|) + |y_1| + |y_2| + |y_3|)} \phi_\rho(x, y) dx dy,
\]

and the stated bound on \( \text{Cov}(f(X, a), f(Y, b)) \) results since \( \phi_\rho \leq c_2e^{-|x|^2 + |y|^2}/(2(1-s)) \) for some \( c_2 = c_2(s) < \infty \). The bound on \( |Ef(X, a)f(Y, b)| \) results by similar (and somewhat simpler) considerations.

Theorem 2.3 Under the conditions of Theorem 2.2, assume \( \max_{i \neq j} |\rho_{ij}| \leq r < 1/3 \) and \( \max_{i} \sum_{j=1}^{n} |\rho_{ij}| \leq B < \infty \). Let \( \psi \geq 0 \) with \( N\psi = 1 \) be such that \( \psi(u) \leq Ke^{K|u|^q} \) for some \( 0 < K < \infty \) and \( q < 2 \), and define \( D = \max\{||h||, ||h'||\} \). Then, for some \( C(r, K, q) < \infty \),

\[
|Ef(\frac{W - n}{\sigma}) - Nh| \leq \frac{4DB(nBC)^{1/2}}{\sigma} + \frac{DCB^2n}{\sigma^3}.
\]

(17)
Remark 2.5 Comparing (17) with (14) we see that all the conclusions in the discussion below Corollary 2.1 apply also to non-smooth $\psi$.

Proof From the proofs of Theorem 2.2 and Corollary 2.1 it is clear that to obtain (17) suffices to show that
\[ \text{Cov}(A_{ij}, A_{i'j'}) \leq C(|\rho_{ii'}| + |\rho_{ij'}| + |\rho_{j'i'}|)|\rho_{ij}||\rho_{i'j'}| \]  
(18)
and
\[ |E(A_{ij}A_{i'j'})| \leq C|\rho_{ij}||\rho_{i'j'}|. \]  
(19)
For $i \neq i' \neq j \neq j'$ we obtain these bounds with $C = C(3r, K, q) < \infty$ from Lemma 2.4, by applying this lemma for $a = \rho_{ij}$ and $b = \rho_{i'j'}$. While the cases in which not all four indices $(i, j, i', j')$ are distinct require special considerations, the bounds of (18) and (19), perhaps with a different value for $C$, can be proved by properly adapting the proof of Lemma 2.4. For example, one may assume that $i \neq j$ and $i' \neq j'$ in (18) by following the same argument as in (22) and (23). We omit the details here.

Remark The assumption that $r < 1/3$ has been made in order to simplify the proof of Theorem 2.3 and may be somewhat relaxed. For example, in case of a stationary sequence $\{U_i\}$ with $B = B_n$ such that $n^{-1}B_n \to 0$, clearly $|r_i| \to 0$. Hence, for each $i$ there are no more than $L < \infty$ values of $j$ with $|\rho_{ij}| = |r_{i-j}| \geq 1/3$. By considering the corresponding $A_{ij}$ separately, one can deduce a bound similar to (17) even when $L > 0$.

3 An example concerning the multinomial distribution

Our next example deals with the multinomial distribution where the numbers of balls $n$ and the number of cells $N$ are of the same rate. For simplicity, we shall state a result for $n = N$ and equally likely cells, though the method readily lends itself to generalizations. In this case the number of balls in each cell is asymptotically Poisson$(1)$, and there is negative dependence (e.g., negative association), between the cell counts. However, if we denote the cell counts by $U_1, \ldots, U_N$ then negative association of $\psi(U_1), \ldots, \psi(U_N)$ is guaranteed only for a monotone $\psi$. Even in the monotone case, Newman’s CLT ([18]) does not apply in general for $W = \sum_{i=1}^N \psi(U_i)$. The case where $W$ counts the number of empty cells was extensively studied in [14]. The present approach can easily be extended to the case of non-equal but commensurate cell probabilities, and for $W = \sum_{i=1}^n \psi_i(U_i)$, with different $\psi_i(·)$. Limit theorems for such a $W$, without rates, were studied for example in [17], and in a more general setting in [9] where the first step is to represent the multinomial distribution by independent Poisson variables conditioned on their sum. More recently, this approach and characteristic function methods, were taken in [10] and [11] to assess the convergence rate. See also [12] for statistical applications and further references. We comment briefly that for $W = \sum_{i=1}^n \psi_i(U_i)$, where $(U_1, \ldots, U_n)$ are jointly distributed as independent variables conditioned on their sum, a $W$–size biased coupling variable can be constructed naturally, extending the ideas of the discussion below.

Theorem 3.1 Let $U = (U_1, \ldots, U_n)$ have the multinomial distribution $P(U_1 = k_1, \ldots, U_n = k_n) = \binom{n}{k_1, \ldots, k_n}(\frac{1}{n})^n$. Let $W = \sum_{i=1}^n \psi(U_i)$, where $\psi \geq 0$ is scaled such that $E\psi(U) = 1$ (hence $\lambda = EW = n$), and is such that $\psi(u) \leq Ke^{Ku}$ for some $K < \infty$ and all $u \geq 0$. Let $\sigma^2 = \text{Var} W$ and $D = \max\{\|h\|, \|h'\|\}$. Then there exists a constant $C = C(K) < \infty$ such that
\[ |Eh\left(\frac{W - \lambda}{\sigma}\right) - Nh| \leq \frac{CD}{\sigma} \max\{1, \frac{n}{\sigma^2}\} \]  
(20)
Remark 3.1 The fact that \( U_1 \) given \( U_2 = k_2 \) is Binomial \((n - k_2, 1/(n - 1))\) leads to the expansion 
\[
P(U_1 = k_1, U_2 = k_2)/P(U_1 = k_1)P(U_2 = k_2) = 1 - (k_1 k_2 - k_1 - k_2 + 1)/n + o(1/n)
\] as \( n \to \infty \). Applying this to 
\[
\text{Cov}(\psi(U_1), \psi(U_2)) = \sum \sum \psi(k_1)\psi(k_2)P(U_1 = k_1, U_2 = k_2)/P(U_1 = k_1)P(U_2 = k_2) - 1\] \( P(U_1 = k_1)P(U_2 = k_2) \) 
and \( n^{-1}\sigma^2 = \text{Var}\psi(U) + (n - 1)\text{Cov}(\psi(U_1), \psi(U_2)) \), we conclude that for fixed \( \psi \) and large \( n \)
\[
n^{-1}\sigma^2 = \text{Var}\psi(U) - \text{Cov}(\psi(U), U)^2/\text{Var}U + o(1) .
\]
Hence, \( n^{-1}\sigma^2 \) is bounded away from zero for any non-linear \( \psi \), and the rate obtained in (20) is of order \( n^{-1/2} \) as expected.

By translation, Theorem 3.1 applies also to \( \psi \) which is possibly negative but bounded below. Since we know that \( P(\max_i U_i > \log n) \leq c/n \) (see the discussion below (21)), we may accommodate \( \psi \) such that \( \psi(u) \geq -u^k \) for some \( k \) by the \( n \)-dependent translation \( \psi + (\log n)^k \), at the cost of a somewhat slower rate of convergence.

Proof Let \( V_i \sim \psi(u)P(U_i < du) \), be i.i.d. random variables independent of all other variables. Let \( Z^{(i)} = (Z_1^{(i)}, \ldots, Z_n^{(i)}) \) be independent conditional on \( U \), with \( Z_j^{(i)} \) denoting the net change in the number of balls in the \( j \)-th cell when removing with equal probabilities \( m \) balls from the balls in all but the \( i \)-th cell, and adding them to the \( i \)-th cell if \( V_i - U_i = m \geq 0 \), while for \( \psi(u) = -m < 0 \) removing \( m \) balls from cell \( i \), and distributing them with equal probabilities into the remaining cells. To construct the random variable \( W^* \) needed for applying Theorem 2.1, choose a random index \( I \) uniformly in \( \{1, \ldots, n\} \) and let \( Y_j^{(I)} = U_j + Z_j^{(I)} \). Invoking Lemma 2.1, the random variable \( W^* = \sum_{j=1}^{n} \psi(Y_j^{(I)}) \) has the \( W \)-size biased distribution.

We turn now to bound the r.h.s. of (1). Our assumptions on \( \psi(\cdot) \) imply that for all \( k \),
\[
E(e^{kW^*}) \leq KE(e^{(K+k)U}) \leq e^{c(k,K)}
\] (21)
where \( c = c(k,K) < \infty \) is independent of \( n \). By Markov's inequality \( P(V_i > \log n) \leq c/n^k \) and also \( P(U_j > \log n) \leq c/n^k \). Since \( h \) is bounded, we may and shall assume that \( V_i \leq \log n \) for all \( i \), and \( U_j \leq \log n \) for all \( j \), thereby neglecting events whose probability is of smaller order than our approximation rate. Let now \( A_{ij} = \psi(U_j + Z_j^{(i)}) - \psi(U_j) \) and note that
\[
E[W^* - W|W] = \frac{1}{n}E[\sum_{i=1}^{n} A_{ii}|W] + \frac{1}{n}E[\sum_{i=1 \neq j}^{n} A_{ij}|W] ,
\]
hence, using \( \text{Var}(A + B) \leq 2\text{Var}A + 2\text{Var}B \) and ignoring the conditioning on \( W \) in the second term,
\[
\text{Var}E[W^* - W|W] \leq \frac{2}{n^2} \text{Var}E[\sum_{i=1}^{n} A_{ii}|W] + \frac{2}{n^2} \text{Var} \sum_{i=1 \neq j}^{n} A_{ij} .
\] (22)
Since \( A_{ii} = \psi(V_i) - \psi(U_i) \) and \( \sum_{i=1}^{n} \psi(V_i) \) is independent of \( W \) we see that
\[
\text{Var}E[\sum_{i=1}^{n} A_{ii}|W] = \text{Var}W = \sigma^2 .
\] (23)
By the symmetry in the problem,
\[
\frac{1}{n^2} \text{Var} \sum_{i=1 \neq j}^{n} A_{ij} = \frac{1}{n^2} \sum_{i=1 \neq j}^{n} \sum_{j' \neq i}^{n} \sum_{j'' \neq j',j''}^{n} \text{Cov}(A_{ij}, A_{j'j''}) = \frac{n-1}{n} \sum_{i=1 \neq j}^{n} \sum_{j' \neq i}^{n} \text{Cov}(A_{ij}, A_{j'j''}) \]
\[
\leq 2E(A_{ij}^2) + 4nE(A_{12})^2 + nE(A_{ij}A_{ij}) + nE(A_{ij}A_{32}) + 2nE(A_{12}A_{23}) + n^2|\text{Cov}(A_{12}, A_{34})|.
\] (24)
Define \( L_{ij} = 2(U_j + 1)|V_i - U_i| \), and \( T_{ij} = K e^{K(U_j + |V_i - U_i|)} \). Our assumptions on \( \psi(\cdot) \) imply that \(|A_{ij}| \leq T_{ij}\). Moreover, for all \( j \neq i \) we have

\[
P(A_{ij} \neq 0|U, V) \leq P(|Z^{(i)}_j| \geq 1|U, V) \leq E[|Z^{(i)}_j||U, V] \leq \frac{L_{ij}}{n},
\]

where for the last inequality to hold we assumed that \( V_i < n/2 \). Consequently,

\[
E(A_{12}^2) = E[E(A_{12}^2|U, V)] \leq E[T_{12}^2 P(A_{12} \neq 0|U, V)] \leq \frac{1}{n} E[T_{12}^2 L_{12}] \leq \frac{c_1(K)}{n},
\]

where \( c_1(K) \propto \infty \) is independent of \( n \) (the last inequality is obtained by a calculation similar to that of (21)). In much the same way we see that \( E[|A_{12}| \leq c_2(K)/n \). Conditionally on \( U, V \), the random variable \( A_{12} \) is independent of \( A_{i'j'} \), for all \( i' \neq 1 \). Therefore,

\[
E[|A_{12} A_{32}|] = E[E[|A_{12}| |U, V] E[|A_{32}| |U, V]] \leq \frac{1}{n^2} E[T_{12} L_{12} T_{32} L_{32}] \leq \frac{c_3(K)}{n^2},
\]

and similarly also \( E[|A_{12} A_{23}|] \leq c_4(K)/n^2 \). Assuming with no loss of generality that \( V_2 < n/2 \) and \( V_3 < n/2 \), it is easy to check that

\[
P(A_{12} \neq 0 \text{ and } A_{13} \neq 0|U, V) \leq E[|Z^{(1)}_2 Z^{(1)}_3| |U, V] \leq \frac{1}{n^2} L_{12} L_{13},
\]

and thus to obtain the bound \( E[|A_{12} A_{13}|] \leq c_5(K)/n^2 \). Let \( P^* \) be the product of the laws of the triplets \((V_1, U_1, U_2)\) and \((V_3, U_3, U_4)\). Denote by \( f_n \) the probability ratio \( dP/dP^* \), and by \( E^*[\cdot] \) expectations according to the distribution \( P^* \), so that \( E[\cdot] = E^*[\cdot f_n] \). Define \( B_{ij} = E[A_{ij}|V_i, U_i, U_j] \), observing that \( E[A_{12} E[A_{34}] = E^*[B_{12} B_{34}] \) and by the conditional independence alluded to above, also \( E[A_{12} A_{34}] = E[B_{12} B_{34}] \). Thus,

\[
Cov(A_{12}, A_{34}) = E[A_{12} A_{34}] - E[A_{12}] E[A_{34}] = E^*[f_n - 1] B_{12} B_{34}.
\]

As we have seen, \( n|B_{ij}| \leq T_{ij} L_{ij} \) implying that

\[
n^2 |Cov(A_{12}, A_{34})| \leq n^{-1} E^*[f_n - 1|T_{12} L_{12} T_{32} L_{34}].
\]

(25)

(25)

Clearly, \( f_n \) depends only on the value of \( U_1 + U_2 \) and that of \( U_3 + U_4 \), with

\[
f_n(x, y) = \frac{P(U_1 = k_1, U_2 = x - k_1, U_3 = k_3, U_4 = y - k_3)}{P(U_1 = k_1, U_2 = x - k_1, U_3 = k_3, U_4 = y - k_3)} \quad \text{for all } k_1, k_3.
\]

By straightforward calculations, for all \( n \geq \max(6, x + y) \),

\[
-2(x + y) \leq n(f_n(x, y) - 1) \leq 3(x + y) 2(x + y).
\]

(26)

Thus, with \( U_j < n/4 \), we easily deduce from (25) and (26) that \( |Cov(A_{12}, A_{34})| \leq c_6(K)/n^3 \). Combining all these bounds with (22)-(24), we see that for some \( C_1 = C_1(K) < \infty \)

\[
Var E[W^* - W|W] \leq \frac{2\sigma^2}{n^2} + \frac{C_1}{n}.
\]

(27)

(27)

Turning now to bound the term \( E(W^* - W)^2 \), by the symmetry of the problem,

\[
E[(W^* - W)^2] = \frac{1}{n} \sum_{j=1}^{n} E[(\sum_{j=1}^{n} A_{ij})^2] = E[(\sum_{j=1}^{n} A_{ij})^2] = \sum_{j=1}^{n} \sum_{j'=1}^{n} E[A_{ij} A_{ij'}]
\]

\[
\leq E[A_{12}^2] + 2nE[|A_{11} A_{12}|] + nE[A_{12}^2] + n^2 E[|A_{12} A_{13}|].
\]

Since \( E[A_{12}^2] \leq E[T_{12}^2] \leq c_7(K) \) and \( nE[|A_{11} A_{12}|] \leq E[T_{11} T_{12} L_{12}] \leq c_8(K) \), it now follows that \( E[(W^* - W)^2] \leq C_2 \) for some \( C_2 = C_2(K) < \infty \). Combining the latter with (1) and (27), one can easily verify that (20) holds for \( C = 2\sqrt{2 + C_1} + C_2 \).
4 Exceedances of the \( r \)-scans process and the \( r \)-spacings on the sphere

Let \( X_1, X_2, \ldots, X_{n+r-1} \) be i.i.d. random variables with \( R_i = \sum_{k=0}^{i-1} X_{i+k}, \) \( i = 1, 2, \ldots, n \) denoting their \( r \)-scans process. We are interested in counting the exceedances of \( \{R_i\} \), that is, \( W = \sum_{i=1}^{n} J_i \) where \( J_i = 1_{R_i \leq a} \). Let \( p = P(R_1 \leq a) \). Clearly, \( \lambda = EW = np \) and (for \( n \geq r \)),

\[
\sigma^2 = \text{Var}W = \lambda[1 - P(R_1 \leq a)] + 2 \sum_{\Delta=1}^{r-1} (1 - \frac{\Delta}{n})\psi(\Delta)
\]

(28)

where \( \psi(\Delta) = P(R_{\Delta+1} \leq a | R_1 \leq a) - P(R_1 \leq a) \geq 0 \). In particular, for non-negative \( X \)'s whose distribution function is strictly positive on \((0, \infty)\) and continuous at zero we have \( \sigma^2/\lambda \to 1 \) when \( a \to 0 \) suggesting that in this regime \( W \) is approximately Poisson(\( \lambda \)). Applying the Chen-Stein Poisson approximation method, this result is proved in [6, Theorem 1] which also supplies the total-variation error bound

\[
\|\mathcal{L}(W) - \text{Poisson}(\lambda)\|_{\text{var}} \leq 4 \sum_{\Delta=1}^{\infty} P\left(\sum_{k=1}^{\Delta} X_i \leq a\right).
\]

The study in [6] of the asymptotics of \( W \) when \( n \to \infty \) is motivated by its applicability to the evaluation of the significance of observed inhomogeneities in the distribution of markers (e.g. restriction sites) along the length of long DNA sequences (see also [13] for more detailed biomolecular applications). Note that for non-negative \( X \)'s the Poisson approximation is not valid if \( \lambda/n \to 0 \) and \( a/r \) is bounded away from 0, for then \( \sigma^2/\lambda - 1 \) is also bounded away from 0.

In the following theorem, we remove the restriction of \( X \)'s being non-negative and approximate the law of \( W \) by \( N(\lambda, \sigma^2) \) in the Kolmogorov-Smirnov distance. While the Poisson approximation for the count \( \tilde{W} = \sum_{i=1}^{n}(1 - J_i) \) of large exceedances \( 1_{R_i > a} \) might fail depending on the tails of \( X \) (see [6, Theorem 2]), the \( N(\lambda, \sigma^2) \) approximation to \( W \) presented below is equivalent to the \( N(n - \lambda, \sigma^2) \) approximation to \( \tilde{W} \).

**Theorem 4.1** For all \( w \),

\[
\left| P\left( \frac{\sum_{i=1}^{n} J_i - \lambda}{\sigma} \leq w \right) - \Phi(w) \right| \leq 14 \frac{(2r-1)^2}{\sigma} \leq 14 \frac{(2r-1)^2}{\sqrt{np(1-p)}},
\]

(29)

where \( \Phi \) denotes the standard normal distribution function.

For expectations of smooth functions of \( W \), applying [20, page 110] directly with calculations similar to those appearing around (33), leads to

\[
|E[h((W - \lambda)/\sigma)] - Nh| \leq 2 \frac{(2r-1)^2}{\sigma} (\|h\| + \frac{4}{\sqrt{2r-1}} \|h\|).
\]

(30)

The Normal approximations in (29) and (30) have an error rate of order \( r^2/\sqrt{n \min(p, 1-p)} \). Consequently, if \( p = P(\sum_{i=1}^{n} X_i \leq a) \) is held bounded away from 0 and 1 (e.g., by taking \( a = a(r) \) such that \( a(r)/r \to E(X) \) at the appropriate rate), then asymptotic normality follows provided \( r(n)n^{-1/4} \to 0 \). Also, if \( c = \limsup, a(r)/r < E(X) \) is such that \( P(X < c) > 0 \) then \( p \to 0 \) but \( p e^{I(c)} \) is bounded below for some \( I(c) < \infty \) and we have asymptotic normality when \( \limsup_n r(n)/\log n < 1/I(c) \). For \( X \)'s non-negative of positive density at 0 and \( a \) fixed, \( p r e^{K(a)} \) is bounded below for some \( K(a) < \infty \) and hence \( \limsup_n r(n)/\log n/\log n < 1 \) implies asymptotic normality.

Theorem 4.1 is a corollary of the next theorem.
Theorem 4.2 Let $Y_1, \ldots, Y_n$ be random variables satisfying $|Y_i - E(Y_i)| \leq B$ a.s., $i = 1, \ldots, n$, $E \sum_{i=1}^n Y_i = \lambda$, $\text{Var} \sum_{i=1}^n Y_i = \sigma^2 > 0$ and $\frac{1}{n} E \sum_{i=1}^n |Y_i - E(Y_i)| = \mu$. Let $S_i \subset \{1, \ldots, n\}$ be such that $j \in S_i$ if and only if $i \in S_j$ and $(Y_i, Y_j)$ is independent of $\{Y_k\}_{k \notin S_i \cup S_j}$ for $i, j = 1, \ldots, n$. Then, for $D = \max_{1 \leq i \leq n} |S_i|,$

$$
P \left( \frac{\sum_{i=1}^n Y_i - \lambda}{\sigma} \leq w \right) - \Phi(w) \leq \frac{7 \mu}{\sigma^3} (DB)^2$$

(31)

Remark In triangular arrays for which $\mu \ll B$, the above is an improvement of [19, Theorem 2.2] (which in turn is based on [20, page 110]). By combining Stein’s method with the study of characteristic functions, the Normal approximation with error term comparable to that of [19, Theorem 2.2] is provided in [23, Theorem 5] for $Y_i$ any strictly stationary $D$-dependent sequence of random variables with finite third moment.

Proof of Theorem 4.1 The result follows from (31) by setting $S_i = \{i-r+1, \ldots, i+r-1\}$ with $D = 2r - 1$ and $B = 1$ and noting that $\sigma^2 \geq \lambda(1-p) = n\mu/2$ (see (28)).

Proof of Theorem 4.2: Without loss of generality assume that $E(Y_i) = 0$, hence $|Y_i| \leq B$ a.s., $i = 1, \ldots, n$ and $\mu = E(|Y_i|)$, where the random index $I$ is uniformly distributed over $\{1, \ldots, n\}$ independent of the $Y$’s. Set $A = DB/\sigma$, $X_i = Y_i/\sigma$, $i = 1, \ldots, n$, and $W = \sum_{i=1}^n X_i$.

The proof of (31) parallels that of [19, Theorem 2.1], departing from the latter by using the conditional (upon $I = i$) independence of $X_I$ and $W^* = W - \sum_{j \in S_I} X_j$ which is not available in the more general framework considered there.

For completeness we review the basic steps of the proof, leaving out some of the technical calculations.

For a given function $h : R \to [0,1]$, continuous and piecewise continuously differentiable, let $f$ denote the unique solution of the differential equation $f'(w) - w f(w) = h(w) - Nh$ with $\lim_{w \to \infty} f(w)e^{-w^2/2} = 0$. Since $E[nX_I f(W^*)] = 0$, simple manipulations and a Taylor series expansion of $f(W) - f(W^*)$ (with integral remainder) yield

$$Eh(W) - Nh = E[f'(W)[1 - nX_I(W-W^*)]] + E \int_{W^*}^W nX_I(t-W^*)df'(t).$$

With the bound $\|f'\| \leq 1$ (see [1, (2.17)]), we obtain

$$|E[f'(W)[1 - nX_I(W-W^*)]]| \leq \sqrt{E \left\{ \sum_{i=1}^n \sum_{j \in S_i} (X_i X_j - EX_i X_j) \right\}^2}$$

We now apply the above to the function

$$h(x) = \begin{cases} 1 & \text{if } x \leq w \\ 1 - \frac{1}{\epsilon}(x-w) & \text{if } w \leq x \leq w + \epsilon \\ 0 & \text{otherwise} \end{cases}$$

Since $Nh \leq \Phi(w) + \frac{\epsilon}{2\sqrt{2\pi}}$, and $P(W \leq w) \leq Eh(W)$, we obtain

$$P(W \leq w) - \Phi(w) \leq \frac{\epsilon}{2\sqrt{2\pi}} + \sqrt{E \left\{ \sum_{i=1}^n \sum_{j \in S_i} (X_i X_j - EX_i X_j) \right\}^2} + E \int_{W^*}^W nX_I(t-W^*)df'(t).$$

(32)

Let $U_{i,j} = X_i X_j - EX_i X_j$, so that

$$E \left\{ \sum_{i=1}^n \sum_{j \in S_i} (X_i X_j - EX_i X_j) \right\}^2 = \sum_{i=1}^n \sum_{k=1}^n \sum_{\ell \in S_k} E U_{i,j} U_{k,\ell}. \quad (33)$$
Observe that per fixed $i$ there are at most $4D^3$ non-zero terms in the r.h.s. of (33), each of which is bounded by $2(B/\sigma)^3 E|X_i|$. Consequently, the r.h.s. of (33) is bounded by $8n\mu A^3/\sigma$.

Let $\psi$ denote the indicator function of the interval $[w, w + \epsilon]$. Then

$$E \int_{W^*} nX_i(t - W^*)dt f'(t) dt = E \int_{W^*} nX_i(t - W^*) \{ f(t) + tf'(t) - \frac{1}{\epsilon} \psi(t) \} dt. \quad (34)$$

The first term in (34) is readily bounded by $\frac{1}{\epsilon} \| f \| \| En X_i (W - W^*) \|^2 \leq \frac{1}{\epsilon} \sqrt{\pi/2} n\mu A^2/\sigma$, using the estimate $\| f \| \leq \sqrt{\pi/2} \| h - Nh \| \leq \sqrt{\pi/2} \| h \|$ (20, page 25), and noting that $|W - W^*| \leq A$, and $E[X_i] = \frac{b_i}{\sigma}$.

For the second term in (34), simply replacing $t$ by $t = W^* + W - W^*$ and integrating, lead to the bound $\frac{1}{\epsilon} \| f \| \| En X(t)(W - W^*) \|^2 + \frac{1}{\epsilon} \| f \| \| nE [X(t)(W - W^*)]^2 W^* \|$. From $\| f \| \leq 1$ (see [1, (2.17)]), we obtain $\frac{1}{\epsilon} \| f \| \| nE [X(t)(W - W^*)]^2 \| \leq \frac{1}{\epsilon} n\mu A^2/\sigma$. Next we also use $E[W] \leq \sqrt{EW^2} = 1$, and the conditional independence of $X_I$ and $W^*$ given $I$, to conclude that

$$\frac{1}{\epsilon} \| f \| \| nE [X(t)(W - W^*)]^2 W^* \| \leq \frac{n}{\epsilon} A^2 \| E[X(t) W^*] = \frac{A^2}{\epsilon} \sum_{i=1}^n E[X_i E[|W^*| | I = i]$$

$$\leq \frac{A^2}{\epsilon} \sum_{i=1}^n E[|X_i| E[|W| + A | I = i] = \frac{nA^2}{\epsilon} E[X(t)(E[|W| + A)] \leq n\mu A^2/\epsilon (1 + A)/\sigma.$$

Returning to (34) we now study the term $\frac{1}{\epsilon} E \int_{W^*} nX_i(t - W^*) \psi(t) dt$. It can be written as $\frac{1}{\epsilon} E \int_{W^*} nX_i(t - W^*) \psi(t) dt Q$, where $Q = \{(W \land W^*, W \lor W^*) \cap (w', w + \epsilon) \neq \emptyset \}$ and $I_Q$ denotes the indicator of the event $Q$. Since

$$Q = \{(W \land W^*, W \lor W^*) \cap (w, w + \epsilon) \neq \emptyset \} \subseteq \{|W^* - w| \leq A + \epsilon\} = R,$$

we have

$$\frac{1}{\epsilon} E \int_{W^*} nX_i(t - W^*) \psi(t) dt \leq \frac{n}{\epsilon} E[X(t)(W - W^*)^2 I_R] \leq \frac{nA^2}{\epsilon} E[X_I I_R]$$

$$= \frac{A^2}{\epsilon} \sum_{i=1}^n E[X_i I_R | I = i] = \frac{A^2}{\epsilon} \sum_{i=1}^n E[X_i | P(I_R | I = i] \leq \frac{nA^2}{\epsilon} E[X_i | P(I_R | I = i] \leq \frac{2A + \epsilon}{\epsilon}.$$

As in [19], setting $\Delta = \sup_w |P(W \leq w) - \Phi(w)|$ it follows that

$$P(|W - w| \leq 2A + \epsilon) \leq \Phi(w + 2A + \epsilon) - \Phi(w - 2A - \epsilon) + 2\Delta \leq 2(2A + \epsilon)/\sqrt{2\pi} + 2\Delta.$$

From (32) and the above bounds we obtain:

$$P(W \leq w) - \Phi(w) \leq \frac{\epsilon}{2\sqrt{2\pi}} + \sqrt{8n\mu A^3/\sigma} + \frac{1}{\epsilon} \sqrt{\pi/2} n\mu A^2/\sigma + \frac{1}{3} n\mu A^3/\sigma$$

$$+ \frac{n}{\epsilon} \mu A^2(1 + A)/\sigma + \frac{n\mu A^2}{\sigma \epsilon} \left\{(2A + \epsilon)/\sqrt{2\pi} + \Delta \right\}.$$

A corresponding lower bound is computed similarly. Note that

$$\sigma^2 = \sum_{i=1}^n \sum_{j \in S_i} E(Y_i Y_j) \leq \sum_{i=1}^n \sum_{j \in S_i} B E|Y_i| \leq BDn\mu,$$

and therefore $An\mu/\sigma \geq 1$. The latter fact is useful in the final calculations. In particular it implies that the bound in (31) is trivial unless $A = (DB/\sigma) \leq 1/7$. Choosing $\epsilon = 7n\mu A^3/\sigma$, the bound (31) follows by straightforward calculations.
Next, we briefly study the problem of normal approximation for counts of exceedances of
r-spacings, in which the dependence structure is more involved, and is not restricted to "local" dependence.
Let $U_0, \ldots, U_{n-1}$ be the adjacent arc lengths on the unit sphere, corresponding to its
partition by independent random points, distributed uniformly. Let $W = \sum_{i=0}^{n-1} I_{T_i \leq a/n}$ where
$T_i = \sum_{j=0}^{r-1} U_{(i+j) \bmod n}$.
Taking $R_i = \sum_{j=0}^{r-1} X_{(i+j) \bmod n}$ to be the r-scans of i.i.d. Exponential(1) random variables
$\{X_j\}$ and $S_n = \sum_{i=0}^{n-1} X_i$, it is well known that one can generate the sequence $\{T_i\}$ by
the formula $T_i = \frac{R_i}{S_n}$. Hence, one approach to the approximation of $W$ is to use the $N(\lambda, \sigma^2)$
approximation already obtained for $W = \sum_{i=0}^{n-1} 1_{R_i \leq a}$ (when $r^2/\sigma \to 0$) and use the above
mentioned coupling to show that $W$ and $W$ are not far apart. Indeed, for every $h : R \to R$
bounded, continuous and piecewise continuously differentiable function

$$|E[h(\frac{W - \lambda}{\sigma})] - N h| \leq |E[h(\frac{W - \lambda}{\sigma})] - E[h(\frac{W - \lambda}{\sigma})]| + |E[h(\frac{W - \lambda}{\sigma})]| - N h|$$

$$\leq \frac{1}{\sigma} E\left[\left|\frac{h'}{\sigma} W - \frac{h'}{\sigma} \tilde{W}\right| + 2\frac{(2r - 1)^2}{\sigma} \left(\|h\|^2 + \frac{4}{\sqrt{2r - 1}}\right)\right].$$

The first term on the r.h.s. is further bounded by

$$\frac{\|h\|}{\sigma} E \left|\sum_{i=1}^{n-r+1} \left(1_{R_i \leq a/s_n} - 1_{R_i \leq a/s_n}\right)\right| \leq \frac{\|h\|}{\sigma} E[|1_{R_1 \leq a/s_n} - 1_{R_1 \leq a/s_n}|] = \Delta_n.$$  

Some calculations based on the fact that $S_n$ is a Gamma($n, 1$) random variable, suggest that
$\Delta_n$ is of order at most $\max(r, a/\sqrt{n})^2$.  

An alternative approach to the Normal approximation of $W$ is by applying Theorem 2.1.
This approach is expected to have a better approximation rate for choices of $a, r, n$ for which
the difference between the first two moments of $W$ and of $W$ is non-negligible. To construct
the $W$-size biased random variable $W^*$, choose independently $I \in \{0, \ldots, n-1\}$ uniformly
and $V \sim P(T \in dq|T \leq a/n)$ both independent of all other variables. For $I = i$, if $T_i > a/n$
multiply each of the $r$ arc lengths contributing to $T_i$ by $V/T_i$ and each of the remaining $(n-r)$
arc lengths by $(1-V)/(1-T_i)$. Otherwise, leave the arc lengths unchanged. In both cases let
$Z_j$ be the $r$-spacings corresponding to the newly obtained arc lengths and $W^* = \sum_{j=0}^{n-1} 1_{Z_j \leq a/n}$.
We shall not pursue this approach further.

References

local maxima of a random function on a graph, In Probability, Statistics and


dependent random variables with arbitrary index set. Preprint.


