Bernoulli **19**(4), 2013, 1268–1293
DOI: 10.3150/12-BEJSP02

Normal approximation and smoothness for sums of means of lattice-valued random variables

GEOFFREY DECROUEZ[1] and PETER HALL[2]

[1]*Department of Mathematics and Statistics, The University of Melbourne, VIC 3010, Australia.*
E-mail: dgg@unimelb.edu.au
[2]*Department of Mathematics and Statistics, The University of Melbourne, VIC 3010, Australia and Department of Statistics, University of California Davis, Davis, CA 95616, USA.*
E-mail: halpstat@ms.unimelb.edu.au

Motivated by a problem arising when analysing data from quarantine searches, we explore properties of distributions of sums of independent means of independent lattice-valued random variables. The aim is to determine the extent to which approximations to those sums require continuity corrections. We show that, in cases where there are only two different means, the main effects of distribution smoothness can be understood in terms of the ratio $\rho_{12} = (e_2 n_1)/(e_1 n_2)$, where e_1 and e_2 are the respective maximal lattice edge widths of the two populations, and n_1 and n_2 are the respective sample sizes used to compute the means. If ρ_{12} converges to an irrational number, or converges sufficiently slowly to a rational number; and in a number of other cases too, for example those where ρ_{12} does not converge; the effects of the discontinuity of lattice distributions are of smaller order than the effects of skewness. However, in other instances, for example where ρ_{12} converges relatively quickly to a rational number, the effects of discontinuity and skewness are of the same size. We also treat higher-order properties, arguing that cases where ρ_{12} converges to an algebraic irrational number can be less prone to suffer the effects of discontinuity than cases where the limiting irrational is transcendental. These results are extended to the case of three or more different means, and also to problems where distributions are estimated using the bootstrap. The results have practical interpretation in terms of the accuracy of inference for, among other quantities, the sum or difference of binomial proportions.

Keywords: algebraic irrational number; bootstrap; confidence interval; continuity correction; difference of binomial proportions; discontinuity; irrational number; sum of binomial proportions; transcendental number

1. Introduction

1.1. Background: The case of a single sample mean

Let $\hat{\theta}$ denote a statistical estimator of an unknown quantity θ, and assume that $\hat{\theta} - \theta$ is asymptotically normally distributed with zero mean and variance $n^{-1}\sigma^2$, where n is a measure of the sample size from which $\hat{\theta}$ was computed. In particular, the statistic $T = n^{1/2}(\hat{\theta} - \theta)/\sigma$ is asymptotically normal N$(0, 1)$. Under additional assumptions an Edgeworth expansion of the

1350-7265

distribution of T can generally be formulated, having the form

$$P(T \le x) = \Phi(x) + n^{-1/2} P(x)\phi(x) + o\bigl(n^{-1/2}\bigr), \tag{1.1}$$

uniformly in x, where Φ and ϕ are the standard normal distribution and density functions, respectively, and P is an even, quadratic polynomial.

For example, if $\hat\theta$ denotes the mean of a sample of size n from a population with mean θ, and if data from the population have finite third moment and a nonlattice distribution, then (1.1) holds with $P(x) = \frac{1}{6}\beta(1 - x^2)$, where β is the standardised skewness of the population. On the other hand, still in the case of the mean of a population with finite third moment, if the population is lattice then an extra, discontinuous term has to be added to (1.1).

This extra term reflects the discrete continuity correction that statisticians are often obliged to introduce when approximating a lattice distribution, for example the binomial distribution or the distribution of a sum of Poisson variates, using the smooth normal distribution:

$$P(T \le x) = \Phi(x) + n^{-1/2} P(x)\phi(x) + n^{-1/2} d_n(x)\phi(x) + o\bigl(n^{-1/2}\bigr), \tag{1.2}$$

where $d_n(x) = (e_0/\sigma)\psi_n(x)$ denotes the discontinuous term in the Edgeworth expansion, e_0 is the maximal span of the lattice, σ^2 is the population variance,

$$\psi_n(x) = \psi\bigl\{(x - \xi_n)\sigma n^{1/2}/e_0\bigr\}, \qquad \xi_n = \bigl(e_0/\sigma n^{1/2}\bigr)\bigl\{\tfrac{1}{2} - \psi(n x_0/e_0)\bigr\},$$

$\psi(x) = \lfloor x \rfloor - x + \frac{1}{2}$, $\lfloor x \rfloor$ is the largest integer not strictly exceeding x, and it is assumed that all points of support in the distribution (of which θ is the mean) have the form $x_0 + \nu e_0$ for an integer ν.

1.2. Contributions of this paper

We show that in multi-sample problems, where $\hat\theta$ is a sum of several independent means, the discontinuous term can be ignored if sample sizes are chosen judiciously. For example, if there are just two sample sizes (as in the case of a sum or difference of two binomial proportions), and if the lattice edge widths are identical (this simplifies our discussion here, but is not essential), then it is sufficient that the ratio of the two sample sizes converge to an irrational number, or converge sufficiently slowly to a rational number. These results are corollaries of Theorem 1 in Section 2.1, and they and other properties are discussed in Section 2.2.

We also show that the discontinuous term can be replaced by $O(n^{\delta-1})$, for all $\delta > 0$, provided that four moments are finite and the ratio of the two sample sizes converges sufficiently quickly to an irrational number of "type" 1. (See Section 2.3 for a definition of the type of an irrational number.) More generally, we explore the effect that type has on the size of the discontinuous term. Theorem 2 also gives an explicit formula for the discontinuous term, up to a remainder of order n^{-1}. Sections 2.5 and 2.6 show how to bound the discontinuous remainder term, for two different approaches to defining that quantity, and show how the effects of irrationals of different types can be teased from the remainder. Applications to the bootstrap are straightforward, and in fact Section 2.7 outlines a bootstrap version of Theorem 2 and discusses its implications.

We do not treat in any detail cases where the differences between two lattice distributions arise mainly in terms of their centres, rather than their lattice edge widths. For example, if two independent sample means $\bar{X}_j$, for $j = 1, 2$, are respectively averages of n_j independent variables and are distributed on lattices $x_j + \nu n_j^{-1} e_j$; and if the difference $x_1 - x_2$ between the lattice translations equals an irrational multiple of the ratio $\rho_{12} = (e_2 n_1)/(e_1 n_2)$; then the distribution of $\bar{X}_1 + \bar{X}_2$ is non-lattice. While this problem and its implications are of mathematical interest, they do not enjoy the practical motivation of problems where, say, $x_1 = x_2$ and ρ_{12} can be almost arbitrary. For example, $x_1 = x_2$ in the problem of constructing confidence intervals for the sum or difference of two binomial probabilities, based on samples of unequal size. Therefore, we address cases where the focus of attention is ρ_{12} rather than $x_1 - x_2$. Differences between lattice centres are permitted by our regularity conditions, but their role is not treated in detail.

1.3. Practical motivation

The extra term in (1.2), relative to (1.1), is of significant interest to a practitioner, since it causes significant inaccuracy when the central limit theorem is used to approximate the distribution of T. The presence of this extra term motivates the continuity correction, and also the fiducial approach taken by Clopper and Pearson [6] and Sterne [21] to estimating a binomial proportion, as well as a large, more recent literature discussing methodology for solving problems such as constructing confidence intervals for the difference or sum of two binomial proportions. See, for example, Hall [11], Duffy and Santer [7], Lee et al. [14], Agresti and Caffo [1], Brown et al. [4,5], Zhou et al. [23], Price and Bonnett [16], Brown and Li [3], Borkowf [2], Roths and Tebbs [19], Wang [22] and Zieliński [24].

The practical motivation for the work described in this paper came from data acquired during quarantine searches, where the construction of confidence intervals for the sum, rather than difference, of two binomial proportions was of interest. In detail, shipping containers arriving at a frontier contained a certain number, N say, of consignments. Some of the consignments might be clean, but others could contain pests which needed to be detected and removed to prevent their introduction to the environment. To reduce the costs associated with inspection, quarantine services usually inspect only $n_1 < N$ consignments. Consignments are assumed to be contaminated with probability p_1, and the number, $n_1 \bar{X}_1$ say, of contaminated consignments found after routine (but incomplete) inspection of the items in each of n_1 consignments is assumed to follow a binomial distribution. Contaminated consignments are then "cleaned," and the members of a subsample of n_2 of them are reinspected. (The data gathered in this way comprise a "leakage survey.") The number of items, $n_2 \bar{X}_2$, still found contaminated (for example, contaminated with a different kind of pest) are assumed to follow a binomial distribution with parameters n_2 and p_2, and typically it is argued that $\bar{X}_1$ and $\bar{X}_2$ are independent. An estimator of the proportion of items that pass through this inspection process, and are still contaminated, is given by

$$\bar{X}_1\left(1 - \frac{n_1}{N}\right) + \bar{X}_2\left(1 - \frac{n_2}{N}\right), \tag{1.3}$$

which can be viewed as a sum of means of lattice-valued random variables where the lattice edge lengths are $e_j = 1 - N^{-1} n_j$ for $j = 1, 2$.

The quarantine inspection service aims to develop a strategy for choosing consignments, and items, to inspect. This reduces the associated costs, and minimises, to at least some extent, the number of contaminated items that cross the border. The performance of such a strategy is assessed, by the quarantine service, using a variety of statistics based on sums of binomials; (1.3) is just one example. Quarantine services are usually interested in providing confidence intervals as well as point estimators, and hence there is significant interest in estimating the distributions of statistics such as that at (1.3).

2. Main results

2.1. Edgeworth expansions with remainder equal to $o(n^{-1/2})$

Let X_{ji}, for $1 \le i \le n_j$ and $j = 1, \ldots, k$, denote independent random variables. Assume that each X_{ji} has a nondegenerate lattice distribution, depending on j but not on i and with maximal lattice edge width e_j and finite third moment. Suppose too that $k \ge 2$. Put $\bar{X}_j = n_j^{-1} \sum_i X_{ji}$, $\mu_j = E(X_{ji})$, $\sigma_j^2 = \mathrm{var}(X_{ji})$ and

$$S = \sum_{j=1}^{k} \bar{X}_j. \tag{2.1}$$

The model (2.1) includes cases of apparently greater generality, for example where signed weights are incorporated in the series, since the absolute values of the weights can be incorporated into (2.1) by modifying the lattice edge widths, and negative signs can be addressed by reflecting the summand distributions.

Since third moments are finite then, if the distributions of $X_{11}, \ldots, X_{k1}$ were to satisfy a smoothness condition, such as that of Cramér, we could express the distribution of S in a one-term Edgeworth expansion:

$$P\left\{\frac{S - E(S)}{(\mathrm{var}\, S)^{1/2}} \le x\right\} = \Phi(x) + n^{-1/2}\frac{1}{6}\beta\bigl(1 - x^2\bigr)\phi(x) + o\bigl(n^{-1/2}\bigr), \tag{2.2}$$

where we take $n = n_1 + \cdots + n_k$ to be the asymptotic parameter, and

$$\beta = \beta(n) = \frac{n^{1/2} E(S - ES)^3}{(\mathrm{var}\, S)^{3/2}} = \frac{n^{1/2} \sum_j n_j^{-2} E(X_{j1} - EX_{j1})^3}{(\sum_j n_j^{-1}\, \mathrm{var}\, X_{j1})^{3/2}} \tag{2.3}$$

is a measure of standardised skewness and, under our assumptions, is bounded as $n \to \infty$. Result (2.2) is a version of (1.1) in a particular case.

However, in general (2.2) does not hold in the lattice-valued case that we are considering. For example, if $k \ge 1$ and the X_{ji}s, for all i and j, have a common lattice distribution, then, as was made clear by Esseen [9], any expansion of the distribution of X has to include a discontinuous term of size $n^{-1/2}$ (specifically, the term $n^{-1/2} d_n(x)\phi(x)$ in (1.2)) that reflects the "continuity correction" needed to approximate the discontinuous distribution of $T = \{S - E(S)\}/(\mathrm{var}\, S)^{1/2}$ by a continuous normal distribution.

When exploring this problem, we suppose that the sequence of values of n is strictly increasing. Further, we assume that

$$\min_{1 \le j \le k} \liminf_{n \to \infty} (n_j/n) > 0. \tag{2.4}$$

In Theorem 1, below, we fix both k and the distributions of X_{j1}, for $1 \le j \le k$. This means that $e_1, \dots, e_k$ are fixed too. However, for each n we consider there to be a potentially new sequence of values $n_1, \dots, n_k$. In particular, the ratios n_{j_1}/n_{j_2} can change considerably from one choice of n to another, although, in view of (2.4), n_{j_1}/n_{j_2} is bounded away from zero and infinity as $n \to \infty$.

In the first part of Theorem 1, below, we also impose the following condition on at least one of the ratios $\rho_{j_1 j_2} = (e_{j_2} n_{j_1})/(e_{j_1} n_{j_2})$:

$$\text{for each integer } \ell \ge 1, \qquad n^{1/2} |\sin(\ell \rho_{j_1 j_2} \pi)| \to \infty \tag{2.5}$$

as $n \to \infty$.

Theorem 1. *Assume that $E|X_{j1}|^3 < \infty$ for $j = 1, \dots, k$; that X_{j1} is distributed on a lattice $x_j + \nu e_j$, for integers ν, where e_j is the maximal lattice edge width; and that (2.4) holds.* (i) *If, for some pair j_1, j_2 with $1 \le j_1 < j_2 \le k$, $\rho_{j_1 j_2}$ satisfies (2.5), then the one-term Edgeworth expansion at (2.2) holds uniformly in x.* (ii) *However, if $\rho_{j_1 j_2}$ equals a fixed rational number (not depending on n) for each pair j_1, j_2, and if the points x_j can all be taken equal, then the expansion at (2.2) fails to hold because it does not include an appropriate discontinuous term of size $n^{-1/2}$.*

2.2. Circumstances where (2.5) holds

If ρ_0 is irrational, then $|\sin(\ell \rho_0 \pi)| > 0$ for all integers ℓ. Therefore, (2.5) holds if $\rho_{j_1 j_2}$ converges to an irrational number as $n \to \infty$.

However, in many cases (2.5) holds without the sequence $\rho_{j_1 j_2}$ converging. For example, assume for simplicity that the lattice edge widths e_j are all identical, let ρ_1 and ρ_2 be two distinct irrational numbers, and let the sequence of values of the ratio n_{j_1}/n_{j_2} be a sequence of convergents of ρ_1 and ρ_2, chosen so that an infinite number of convergents come from each ρ_j. (For a definition of convergents of irrational numbers, see, e.g., Leveque [15], p. 70.) Then (2.5) holds, although the sequence $\rho_{j_1 j_2}$ does not converge.

Importantly, (2.5) also holds in many cases where each $\rho_{j_1 j_2}$ is close to a rational number, indeed where each $\rho_{j_1 j_2}$ converges to a rational number. For example, we claim that (2.5) obtains if $\rho_{j_1 j_2} = 1 + \varepsilon_{j_1 j_2}$, where $\varepsilon_{j_1 j_2} = \varepsilon_{j_1 j_2}(n)$, which can be either positive or negative, converges to zero strictly more slowly than $n^{1/2}$:

$$\varepsilon_{j_1 j_2} \to 0, \qquad n^{1/2} |\varepsilon_{j_1 j_2}| \to \infty. \tag{2.6}$$

In this case, for each fixed, positive integer ℓ,

$$\sin(\ell\rho_{j_1 j_2}\pi) = \sin(\ell\pi) + \ell\pi\varepsilon_{j_1 j_2}\cos(\ell\pi) + O\big(\varepsilon_{j_1 j_2}^2\big),$$

from which it follows that

$$\big|\sin(\ell\rho_{j_1 j_2}\pi)\big| \sim \begin{cases} \ell\pi|\varepsilon_{j_1 j_2}| & \text{if } \ell \text{ is an even integer} \\ 1 & \text{if } \ell \text{ is an odd integer,} \end{cases} \tag{2.7}$$

where $a_n \sim b_n$ means that the ratio a_n/b_n converges to 1. Assumption (2.5) follows from (2.6) and (2.7).

A similar argument can be used to prove that if $\rho_{j_1 j_2} = \rho_0 + \varepsilon_{j_1 j_2}$, where ρ_0 is a fixed rational number and $\varepsilon_{j_1 j_2} = \varepsilon_{j_1 j_2}(n)$ satisfies (2.6), then (2.5) is true. (The case $\rho_0 = 0$ is excluded by (2.4).) These examples make it clear that there is not a great deal of latitude in the assumption, imposed in part (ii) of Theorem 1, that each $\rho_{j_1 j_2}$ should equal a fixed rational number. In particular, for (2.5) to fail it is not sufficient that each $\rho_{j_1 j_2}$ converge to a rational.

2.3. Refinement of bound on remainder term in Edgeworth expansions

In Section 2.1, we showed that, if (2.5) holds, the discontinuous term of size $n^{-1/2}$, in expansions such as (1.2), is actually of smaller order than $n^{-1/2}$. To obtain a more concise bound on the discontinuous term, we shall investigate in detail cases where one or more of the ratios $\rho_{j_1 j_2}$ converge to an irrational number as n diverges. However, this treatment requires a definition of the "type" of an irrational, and we give that next.

If x is a real number, let $\langle x\rangle$ denote the distance from x to the nearest integer. (In particular, if $\lfloor x\rfloor$ is the integer part function, $\langle x\rangle = \min\{x - \lfloor x\rfloor, 1 - (x - \lfloor x\rfloor)\}$.) We say that the irrational number ρ is of type η if η equals the supremum of all ζ such that $\liminf_{p\to\infty} p^{\zeta}\langle p\rho\rangle = 0$, where $p \to \infty$ through integer values. Properties of convergents of irrational numbers (specifically, Dirichlet's Theorem) can be used to prove that the type of any given irrational number always satisfies $\eta \ge 1$. It follows from Roth's Theorem (Roth [18]) that all algebraic irrationals (that is, all irrational numbers that are roots of polynomials with rational coefficients) are of minimal type, i.e., $\eta = 1$, which is one of the cases we consider below.

More generally, if a number is chosen randomly, for example as the value of a random variable having a continuous distribution on the real line, then with probability 1 it is an irrational of type 1. Irrationals that are not algebraic are said to be transcendental, and can have type strictly greater than 1. (However, the transcendental number e is of type 1.) Known upper bounds to the types of π, π^2 and $\log 2$ are 6.61, 4.45 and 2.58, respectively. Liouville numbers have type $\eta = \infty$. The type of an irrational number is one less than its irrationality measure (or equivalently, one less than its approximation exponent or Liouville-Roth constant). We refer the reader to Ribenboim [17] for more information about types of irrational numbers.

Next, we introduce notation which helps us to define an approximation to the discontinuous term, an analogue of $d_n(x)$ in (1.2), when $k = 2$. (Here, k is as in (2.1).) Assuming that the lattice, on which the distribution of X_{ji} is supported, consists of points $x_j + \nu e_j$ for integers ν, define

$\xi_{jn} = e_j(\sigma_j n_j^{1/2})^{-1}\{(n_j x_j/e_j) - \lfloor n_j x_j/e_j \rfloor\}$ and

$$\xi_n(x) = \{x - (c_1\xi_{1n} + c_2\xi_{2n})\}\frac{\sigma_1 n_1^{1/2}}{c_1 e_1}, \tag{2.8}$$

where, recalling that $\sigma_j^2 = \mathrm{var}(X_{ji})$, we define c_j for $j = 1$ and 2 by

$$c_j = \left(\frac{n_j^{-1}\sigma_j^2}{n_1^{-1}\sigma_1^2 + n_2^{-1}\sigma_2^2}\right)^{1/2}. \tag{2.9}$$

Let $\alpha \in (0, \frac{1}{2})$ and partition the set of all integers into adjacent blocks each comprised of $2\lfloor n^\alpha \rfloor + 1$ consecutive integers. Write $\bar{\nu}_\ell$ for the central integer in the ℓth block, which we denote by $\mathcal{N}_\ell$ where $-\infty < \ell < \infty$ and $\mathcal{N}_{\ell+1}$ is immediately to the right of $\mathcal{N}_\ell$ on the number line. Given $\nu \in \mathcal{N}_\ell$, put $\nu_\ell = \nu - \bar{\nu}_\ell$.

Let $c_3 = e_2 n_1/\sigma_1 n_2$ and $c_4 = (e_1/\sigma_2)(n_1/n_2)^{1/2}$, and note that $c_1, \ldots, c_4$ are strictly positive functions of n and are bounded away from zero and infinity as n diverges. Put $\gamma = \prod_{j=1,2}(e_j/\sigma_j)$, and, given an integer $r_0 \ge 1$, define

$$\phi(u, x) = \phi\{(x/c_1) - c_3 u\}\phi(c_4 u), \qquad \phi_r(u, x) = (\partial/\partial u)^r \phi(u, x),$$
$$K_n(x) = \gamma \sum_{r=0}^{r_0} \sum_{-\infty<\ell<\infty} \frac{\phi_r(\bar{\nu}_\ell/n_1^{1/2}, x)}{r!\, n_1^{r/2}} \sum_{\nu\in\mathcal{N}_\ell} \nu_\ell^r \psi\left\{\xi_n(x) - \frac{e_2 n_1}{e_1 n_2}\nu\right\}, \tag{2.10}$$

where, as in Section 1.1, $\psi(x) = \lfloor x \rfloor - x + \frac{1}{2}$.

We claim that the infinite series in the definition of $K_n(x)$ is absolutely convergent, uniformly in x. To appreciate why, note that

$$\sup_{-\infty<x<\infty} \left|\phi_r(\bar{\nu}_\ell/n_1^{1/2}, x)\right| \le C_1(r)\phi(c_4\bar{\nu}_\ell/n_1^{1/2}), \tag{2.11}$$

where, here and below, the notation $C_j(r)$ will denote a constant depending on r but not on n. Using (2.4) and (2.11), we deduce that

$$\sum_{-\infty<\ell<\infty} \left\{\sup_{-\infty<x<\infty} \left|\phi_r(\bar{\nu}_\ell/n_1^{1/2}, x)\right|\right\} \le C_2(r) n^{(1/2)-\alpha}. \tag{2.12}$$

(In more detail, without loss of generality the block $\mathcal{N}_0$ is centred at 0, in which case, when bounding the series on the left-hand side of (2.12), $\bar{\nu}_\ell$ can be interpreted as $2\ell n^\alpha$. Consequently the left-hand side of (2.12) is bounded by a constant multiple of $C_1(r)\int \phi(2un^\alpha/n_1^{1/2})\,du$, and (2.12) follows.)

More simply, since (a) $|\nu_\ell| \le n^\alpha$, (b) $|\mathcal{N}_\ell| \le (2n^\alpha + 1)$ (where we define $|\mathcal{N}_\ell| = \#\mathcal{N}_\ell$), and (c) $|\psi| \le \frac{1}{2}$, then

$$\sup_{-\infty<x<\infty} \left|\sum_{\nu\in\mathcal{N}_\ell} \nu_\ell^r \psi\left\{\xi_n(x) - \frac{e_2 n_1}{e_1 n_2}\nu\right\}\right| \le C_3(r) n^{(r+1)\alpha}. \tag{2.13}$$

Combining (2.10), (2.12) and (2.13), and replacing each summand on the right-hand side of (2.10) by its absolute value, we obtain the bound: $n^{-1/2}|K_n(x)| \le C_4(r_0)$, uniformly in x. This inequality demonstrates the claimed absolute convergence of the series in (2.10).

Recall the definition of S at (2.1), and that $\rho_{j_1 j_2} = (e_{j_2} n_{j_1})/(e_{j_1} n_{j_2})$. Part (i) of Theorem 2, below, captures the analogue of the discontinuous term, $d_n(x)$, in a multisample version of (1.2), and part (ii) gives conditions under which the net contribution of that term equals $O(n^{\delta-(1/2)-(1/2\eta)})$, for all $\delta > 0$ when some $\rho_{j_1 j_2}$ is sufficiently close to an irrational number of type η.

Theorem 2. *Assume that* $E|X_{j1}|^4 < \infty$ *for* $j = 1, \ldots, k$; *that* X_{j1} *is distributed on a lattice* $x_j + \nu e_j$, *for integers* ν, *where* e_j *is the maximal lattice edge width*; *and that* (2.4) *holds. Choose* $r_0 \ge 4\alpha/(1-2\alpha)$ *in* (2.10). (i) *If* $k = 2$ *and* K_n *is as defined at* (2.10), *then*

$$P\left\{\frac{S - E(S)}{(\operatorname{var} S)^{1/2}} \le x\right\} = \Phi(x) + n^{-1/2}\frac{1}{6}\beta\bigl(1 - x^2\bigr)\phi(x) + (n_1 n_2)^{-1/2} K_n(x) + O\bigl(n^{-1}\bigr), \quad (2.14)$$

uniformly in x. (ii) *If, for some pair* j_1, j_2 *with* $1 \le j_1 < j_2 \le k$, *the ratio* $\rho_{j_1 j_2} = (e_{j_2} n_{j_1})/(e_{j_1} n_{j_2})$ *satisfies*

$$|\rho_{j_1 j_2} - \rho_0| = O\bigl(n^{-(1/2)\{1+(1/\eta)+\delta\}}\bigr) \quad (2.15)$$

for some $\delta > 0$, *where* ρ_0 *is an irrational number of type* η, *then, for each* $\delta > 0$,

$$P\left\{\frac{S - E(S)}{(\operatorname{var} S)^{1/2}} \le x\right\} = \Phi(x) + n^{-1/2}\frac{1}{6}\beta\bigl(1 - x^2\bigr)\phi(x) + O\bigl(n^{\delta-(1/2)-(1/2\eta)}\bigr), \quad (2.16)$$

uniformly in x.

Result (2.16) is of particular interest in the case $\eta = 1$, which encompasses almost all irrational numbers (with respect to Lebesgue measure), including all the algebraic irrationals and some transcendental numbers. When $\eta = 1$,

$$P\left\{\frac{S - E(S)}{(\operatorname{var} S)^{1/2}} \le x\right\} = \Phi(x) + n^{-1/2}\frac{1}{6}\beta\bigl(1 - x^2\bigr)\phi(x) + O\bigl(n^{\delta-1}\bigr), \quad (2.17)$$

uniformly in x for each $\delta > 0$. Result (2.17) implies that the lattice nature of the distribution of X_{ji} can be ignored, almost up to terms of second order in Edgeworth expansions, when considering the impact of latticeness on the accuracy of normal approximations.

2.4. Practical choice of n_1 and n_2

In practice it is not difficult to choose n_1 and n_2 so that (2.15) holds. To see how, assume for simplicity that the lattice edge widths e_1 and e_2 are identical, as they would be if (for example) S were equal to a sum or difference of estimators of binomial proportions. If ρ_0 is an irrational number then the convergents m_1/m_2 of ρ_0 satisfy

$$\bigl|(m_1/m_2) - \rho_0\bigr| \le m_2^{-2}. \quad (2.18)$$

(See e.g. Leveque [15], equation (29), p. 180.) Therefore, if n_1 and n_2 are relatively prime and n_1/n_2 is a convergent of ρ_0, then (2.15), for each $\delta \in (0, 3-(1/\eta)]$, follows from (2.18). The most difficult case, as far as (2.15) is concerned, is the one where the convergence rate in (2.15) is fastest, and arises when $\eta = 1$. There we need to ensure that

$$|\rho_{j_1 j_2} - \rho_0| = O\big(n^{-1-\delta}\big) \tag{2.19}$$

for some $\delta > 0$. Now, (2.19) holds whenever n_1/n_2 is a convergent of ρ_0, and the Khinchin-Lévy Theorem (see, e.g., pp. 82–83 of Einsiedler and Ward [8]) implies that the convergents are reasonably closely spaced; the numerators and denominators generally increase by factors of only $\pi^2/(12\log 2) \approx 1.87$. Moreover, there are many ratios n_1/n_2 on either side of convergents for which (2.19) holds.

The pair (n_1, n_2) can be chosen from tables of, or formulae for, convergents for commonly arising irrationals of type 1. See, for example, Griffiths [10] and references therein, and note that e and any algebraic irrational is of type 1.

2.5. Alternative formula for K_n, and derivation of (2.16) from (2.14) when $\eta = 1$

Part (i) of Theorem 2 can be stated for a version of $K_n(x)$ simpler than that at (2.10):

$$K_n(x) = \gamma \sum_\nu \phi\bigg(\frac{x}{c_1} - \frac{e_2 n_1^{1/2}}{\sigma_1 n_2}\nu\bigg)\phi\big\{e_2\big(\sigma_2 n_2^{1/2}\big)^{-1}\nu\big\}\psi\bigg\{\xi_n(x) - \frac{e_2 n_1}{e_1 n_2}\nu\bigg\}. \tag{2.20}$$

Indeed, the $K_n(x)$ at (2.20) is just $\gamma I_4(x)$, where $I_4(x)$ is as defined at (4.17) in the proof of Theorem 1, and in fact that formula provides a convenient point of access to a proof of (2.14) with $K_n(x)$ as at (2.20). However, in the case $\eta > 1$ it is not straightforward to pass from (2.20) to (2.16), and that is why we used the definition of $K_n(x)$ at (2.10).

To appreciate that (2.16) follows from (2.20) when $\eta = 1$, note that the definition of $K_n(x)$ at (2.20) is equivalent to:

$$\gamma^{-1} K_n(x) = \sum_\nu \Psi(x,\nu)\psi\bigg\{\xi_n(x) - \frac{e_2 n_1}{e_1 n_2}\nu\bigg\}, \tag{2.21}$$

where

$$\Psi(x,\nu) = \phi\bigg(\frac{x}{c_1} - \frac{e_2 n_1^{1/2}}{\sigma_1 n_2}\nu\bigg)\phi\big\{e_2\big(\sigma_2 n_2^{1/2}\big)^{-1}\nu\big\}.$$

If (2.15) holds with $\eta = 1$ then a standard argument for bounding discrepancies of sequences (see p. 123 of Kuipers and Niederreiter [13]) can be used to prove that for all $\delta > 0$,

$$\sup_{-\infty<z<\infty}\bigg|\sum_{\nu=1}^{N}\psi\bigg(z - \frac{e_2 n_1}{e_1 n_2}\nu\bigg)\bigg| = O\big(N^\delta\big). \tag{2.22}$$

Note too that

$$\sup_{\nu\geq 1}\sup_{-\infty<x<\infty}\left|\Psi(x,\nu+1)-\Psi(x,\nu)\right|\leq Cn^{-1/2}. \tag{2.23}$$

Taking $a_\nu=\Psi(x,\nu)$ and $b_\nu=\psi\{\xi_n(x)-(e_2n_1/e_1n_2)\nu\}$, and employing Abel's method of summation, we can write:

$$\sum_{\nu=1}^{N}a_\nu b_\nu=a_N\sum_{\nu=1}^{N}b_\nu-\sum_{\nu=1}^{N-1}(a_{\nu+1}-a_\nu)\sum_{j=1}^{\nu}b_j,$$

which in company with (2.22) and (2.23) allows us to prove that, provided $N=O(n^C)$ for some $C>0$,

$$\sup_{-\infty<x<\infty}\left|\sum_{\nu=1}^{N}\Psi(x,\nu)\psi\left\{\xi_n(x)-\frac{e_2n_1}{e_1n_2}\nu\right\}\right|=O\left(N^\delta\right) \tag{2.24}$$

for all $\delta>0$. More simply, if $N\geq n^2$ then

$$\sup_{-\infty<x<\infty}\left|\sum_{\nu=N+1}^{\infty}\Psi(x,\nu)\psi\left\{\xi_n(x)-\frac{e_2n_1}{e_1n_2}\nu\right\}\right|$$
$$\leq\frac{1}{2}\sup_{-\infty<x<\infty}\sum_{\nu=N+1}^{\infty}\Psi(x,\nu)=O(1). \tag{2.25}$$

Combining (2.24) and (2.25), using a similar argument to treat series where $\nu\leq 0$, and noting the definition of $K_n(x)$ at (2.21), we deduce that $\sup_x|K_n(x)|=O(n^\delta)$ for all $\delta>0$. In the case $k=2$, and for $\eta=1$, this gives (2.16) as a corollary of (2.14).

2.6. Derivation of (2.16) from (2.14) when $\eta\geq 1$

Our proof of (2.16), in Section 4.2, will proceed by deriving implicitly a version of (2.14) in the case $k\geq 2$, and showing that, if (2.15) holds, then that version of (2.14) entails (2.16). The relative complexity of a form of (2.14) for general k discouraged us from including it in Theorem 2, but it is nevertheless instructive to show how, when $k=2$, one can obtain (2.16) from (2.14). We outline the proof below, highlighting the properties of irrational numbers, particularly the differences between the case of irrationals of type $\eta=1$ and the case of those of larger type, that determine the bound for the remainder in (2.16).

Note that if q is a polynomial function then, applying Koksma's inequality (see, e.g., Theorem 5.1, p. 143 of [13]) and the Erdős-Turán inequality (see, e.g., formula (2.42), p. 114 of [13]), it can be shown that

$$\chi(N,q,\tau)\equiv\sup_{-\infty<z<\infty}\left|\sum_{i=1}^{N}q(i/N)\psi(z-\tau i)\right|\leq C_1(q)\left\{\frac{N}{m}+\sum_{\ell=1}^{m}\frac{1}{\ell|\sin(\ell\tau\pi)|}\right\}, \tag{2.26}$$

for all integers $m \geq 1$. Here $\tau > 0$ is permitted to vary with N, and the constant $C_1(q)$ depends on the degree and the coefficients of q but not on the positive integer N or on m, z or τ.

We shall take $\tau = \rho_{12}$, a function of n, in which case, since

$$\big||\sin(\ell\tau\pi)| - |\sin(\ell\rho_0\pi)|\big| \leq \ell\pi|\rho_{12} - \rho_0|,$$

we have:

$$|\sin(\ell\tau\pi)| + \ell\pi|\rho_{12} - \rho_0| \geq |\sin(\ell\rho_0\pi)| = \sin\big(\pi\langle\ell\rho_0\rangle\big) \geq 2\langle\ell\rho_0\rangle \geq 2C_2(\delta)\ell^{-\eta-\delta} \tag{2.27}$$

for any given $\delta > 0$ and all $\ell \geq 1$, where $C_2(\delta) > 0$ depends on δ but not on ℓ, the last inequality in (2.27) follows from the assumption that ρ_0 is of type η, and the second-last inequality comes from the fact that $0 \leq \langle x\rangle \leq \frac{1}{2}$ for all real numbers x, and $\sin(\pi x) \geq 2x$ whenever $0 \leq x \leq \frac{1}{2}$. If

$$|\rho_{12} - \rho_0| \leq C_2(\delta)\pi^{-1}m^{-(1+\eta+\delta)} \tag{2.28}$$

then it follows from (2.27) that $|\sin(\ell\tau\pi)| \geq \langle\ell\rho_0\rangle$ for $1 \leq \ell \leq m$, and so

$$\sum_{\ell=1}^{m} \frac{1}{\ell|\sin(\ell\tau\pi)|} \leq \sum_{\ell=1}^{m} \frac{1}{\ell\langle\ell\rho_0\rangle}. \tag{2.29}$$

A standard argument for bounding the discrepancy of a sequence (see, e.g., p. 123 of [13]) can be used to show that, since ρ_0 is an irrational number of type η,

$$\sum_{\ell=1}^{m} \frac{1}{\ell\langle\ell\rho_0\rangle} = O\big(m^{\eta-1+\delta}\big) \tag{2.30}$$

for all $\delta > 0$. Therefore, provided that (2.28) holds, we can deduce from (2.26) and (2.29) that

$$\chi(N, q, \rho_{12}) \leq C_3(q)m^{-1}\big(N + m^{\eta+\delta}\big) = O\big(N^{1-(1/\eta)+\delta_1}\big), \tag{2.31}$$

where $\delta_1 > 0$ and the inequality holds for all m and the identity is true if $m/N^{1/\eta}$ is bounded away from zero and infinity as $N \to \infty$. When m has the latter property, (2.28) is satisfied, for all sufficiently large N, if

$$|\rho_{12} - \rho_0| = O\big(N^{-\{1+(1/\eta)+\delta_2\}}\big) \tag{2.32}$$

for some $\delta_2 > \delta_1$.

Note that it is at (2.30) that the type, η, of the irrational number ρ_0 enters into consideration. In the case $\eta = 1$ the exponent δ in (2.30) could not be removed or reduced, perhaps by replacing the implicit factor m^δ in (2.30) by $(\log m)^C$ for some $C > 0$, without an analogous strengthening of Roth's Theorem. Formula (2.30) also marks the step at which it becomes apparent that a poorer bound will be obtained for an irrational number of type 1, relative to one of type $\eta > 1$.

Applying the bound (2.31), for several versions of the polynomial q, in the case $N = 2\lfloor n^\alpha\rfloor + 1$, we deduce that

$$n^{-r/2} \sup_{-\infty<z<\infty} \left|\sum_{\nu\in\mathcal{N}_\ell} \nu_\ell^r \psi\left(z - \frac{e_2 n_1}{e_1 n_2}\nu\right)\right| = O\big(n^{r\{\alpha-(1/2)\}+\alpha\{1-(1/\eta)+\delta_1\}}\big), \tag{2.33}$$

provided that (2.32) holds, i.e., $|\rho_{12}-\rho_0| = O(n^{-\alpha\{1+(1/\eta)+\delta_2\}})$. Now, the only constraint on α is $0<\alpha<\frac{1}{2}$, and so we can choose α as close to $\frac{1}{2}$, but less than $\frac{1}{2}$, as we desire. In particular, if $\delta_3>0$ is given, and we choose $\alpha=\frac{1}{2}-\delta_4$ where $\delta_4>0$ is sufficiently small, then by taking δ_1 in (2.33) to be small we obtain:

$$\max_{1\le r\le r_0} n^{-r/2} \sup_{-\infty<z<\infty}\left|\sum_{\nu\in\mathcal{N}_\ell} \nu_\ell^r \psi\left(z-\frac{e_2n_1}{e_1n_2}\nu\right)\right| = O\left(n^{(1/2)\{1-(1/\eta)\}+\delta_3}\right), \tag{2.34}$$

provided that

$$|\rho_{12}-\rho_0| = O\left(n^{-(1/2)\{1+(1/\eta)+\delta_5\}}\right), \tag{2.35}$$

where $\delta_5>0$ can be made as small as we like simply by choosing δ_4 small. Now, (2.35) follows from (2.15). It therefore follows from (2.34), and the definition of $K_n(x)$ at (2.10), that if (2.15) holds for some $\delta>0$ then

$$\sup_{-\infty<x<\infty}(n_1n_2)^{-1/2}\left|K_n(x)\right| = O\left(n^{\delta-(1/2)\{1+(1/\eta)\}}\right) \tag{2.36}$$

for all $\delta>0$. Results (2.14) and (2.36) imply (2.16), as had to be shown.

2.7. Expansions relating to the bootstrap

In this section we show that, despite the potential for problems arising from discreteness, the bootstrap (including the double bootstrap) applied to inference based on the distribution of $\{S-E(S)\}/(\operatorname{var} S)^{1/2}$, generally (when (2.15) holds and ρ_0 is of type 1) produces confidence regions and hypothesis tests with the same orders of magnitude of coverage or level accuracy, up to terms of size $n^{\delta-1}$ for all $\delta>0$, as it would in the case of smooth sampling distributions. This result is of practical importance, since standard percentile bootstrap methods applied to lattice distributions are frustrated by the effects of discontinuities; see, e.g., Singh [20] and Hall [12].

For brevity, when establishing this property we treat only the context of Theorem 2. We begin by stating an analogue of (2.14) there, valid when $k=2$. The arguments used to prove part (i) of Theorem 2 can be employed to show that

$$P\left[\frac{S^*-E(S^*\mid\mathcal{X})}{\{\operatorname{var}(S\mid\mathcal{X})\}^{1/2}}\le x \,\middle|\, \mathcal{X}\right] = \Phi(x)+n^{-1/2}\frac{1}{6}\hat\beta\left(1-x^2\right)\phi(x) \\ + (n_1n_2)^{-1/2}\widehat{K}_n(x)+n^{-1}\Delta_1(x), \tag{2.37}$$

where, analogously to the definitions in Section 2.1, $S^*=\sum_j \bar X_j^*$; $\bar X_j^* = n_j^{-1}\sum_i X_{ji}^*$ and $X_{j1}^*,\dots,X_{jn_j}^*$ are drawn by sampling randomly, with replacement, from $\mathcal{X}_j=(X_{j1},\dots,X_{jn_j})$; $\mathcal{X}=(\mathcal{X}_1,\dots,\mathcal{X}_k)$;

$$\hat\beta = \frac{n^{1/2}E[\{S^*-E(S^*\mid\mathcal{X})\}^3\mid\mathcal{X}]}{\{\operatorname{var}(S^*\mid\mathcal{X})\}^{3/2}};$$

using (2.10) or (2.20), respectively, as the model for $K_n(x)$,

$$\widehat{K}_n(x) = \hat{\gamma} \sum_{r=0}^{r_0} \sum_{-\infty<\ell<\infty} \frac{\hat{\phi}_r(\bar{\nu}_\ell/n_1^{1/2}, x)}{r! n_1^{r/2}} \sum_{\nu\in\mathcal{N}_\ell} \nu_\ell^r \psi\left\{\hat{\xi}_n(x) - \frac{e_2 n_1}{e_1 n_2}\nu\right\},$$

$$\widehat{K}_n(x) = \hat{\gamma} \sum_{\nu} \phi\left(\frac{x}{\hat{c}_1} - \frac{e_2 n_1^{1/2}}{\hat{\sigma}_1 n_2}\nu\right) \phi\left\{e_2\left(\hat{\sigma}_2 n_2^{1/2}\right)^{-1}\nu\right\} \psi\left\{\hat{\xi}_n(x) - \frac{e_2 n_1}{e_1 n_2}\nu\right\},$$

where $\hat{\gamma} = \prod_{j=1,2}(e_j/\hat{\sigma}_j)$, $\hat{\phi}_r(u,x) = (\partial/\partial u)^r \hat{\phi}(u,x)$, $\hat{\phi}(u,x) = \phi\{(x/\hat{c}_1) - \hat{c}_3 u\} \times \phi(\hat{c}_4 u)$,

$$\hat{c}_j = \left(\frac{n_j^{-1}\hat{\sigma}_j^2}{n_1^{-1}\hat{\sigma}_1^2 + n_2^{-1}\hat{\sigma}_2^2}\right)^{1/2}$$

for $j = 1$ and 2, $\hat{c}_3 = e_2 n_1/\hat{\sigma}_1 n_2$, $\hat{c}_4 = (e_1/\hat{\sigma}_2)(n_1/n_2)^{1/2}$, and $\hat{\xi}_n(x)$ is defined using the empirical analogue of (2.8); and, for $C_1 > 0$ sufficiently large and for some $C_2 > 0$,

$$P\left\{\sup_{-\infty<x<\infty} \left|\Delta_1(x)\right| > C_1 n^{-1}\right\} = O\left(n^{-C_2}\right). \tag{2.38}$$

The assumptions needed for (2.37) are those imposed for part (i) of Theorem 2. The size of C_2 in (2.38) depends to some extent on the distributions of X_{1i} and X_{2i} (recall that at this point we are assuming that $k = 2$), but for distributions such as the Bernoulli or Poisson, which have all moments finite, C_2 can be taken arbitrarily large if C_1 is sufficiently large. The connection to moments here arises because the $O(n^{-C_2})$ bound in (2.38) is derived using a method related to Markov's inequality, which can be applied at a higher order if more moments are finite.

The methods used in Sections 2.5 and 2.6 to derive uniform bounds to K_n can also be employed to bound $\widehat{K}_n$, giving

$$P\left\{\sup_{-\infty<x<\infty} \left|\widehat{K}_n(x)\right| > n^{\delta+(1/2)\{1-(1/\eta)\}}\right\} = O\left(n^{-C_3}\right) \tag{2.39}$$

for all $\delta > 0$ and some $C_3 > 0$, provided that (2.15) holds. In (2.39), η denotes the type of the irrational number ρ_0 appearing in (2.15), and for sampling distributions such as the Bernoulli or Poisson (with all moments finite), C_3 can be taken arbitrarily large. Therefore, treating the case of irrationals of type 1, we deduce from (2.37)–(2.39) that

$$P\left[\frac{S^* - E(S^* \mid \mathcal{X})}{\{\mathrm{var}(S^* \mid \mathcal{X})\}^{1/2}} \le x \,\middle|\, \mathcal{X}\right] = \Phi(x) + n^{-1/2}\frac{1}{6}\hat{\beta}\left(1 - x^2\right)\phi(x) + n^{-1}\Delta_2(x), \tag{2.40}$$

where

$$P\left\{\sup_{-\infty<x<\infty} \left|\Delta_2(x)\right| > n^{\delta}\right\} = O\left(n^{-C_2}\right) \qquad \text{for all } C_2, \delta > 0.$$

A similar argument, employing the methods introduced in Section 2.6, can be used to prove that (2.40) continues to hold if $k \ge 2$, provided that the assumptions imposed in part (ii) of Theorem 2 hold. Therefore, the properties stated in the first paragraph of this section hold.

3. Numerical properties

Throughout this section, we take $k = 2$ and let X_{ji} be a Bernoulli random variable satisfying $P(X_{ji} = 0) = 1 - P(X_{ji} = 1) = p_j$ for $j = 1, 2$, where $p_1 = 0.4$ and $p_2 = 0.6$. Thus, $\rho_{12} = e_2 n_1/(e_1 n_2) = n_1/n_2$, where n_1 and n_2 are the two sample sizes. We take n_2 to be the integer nearest to $\rho_0 n_1$, and vary n_1 between 10 and 80; n_1 is plotted on the horizontal axes of each of our graphs. The probability

$$P(x) = P\big[\big\{S - E(S)\big\}/(\operatorname{var} S)^{1/2} \le x\big] \tag{3.1}$$

was approximated by averaging over the results of 10^5 Monte Carlo simulations.

To illustrate the influence of ρ_{12} on the oscillatory behaviour of $P(x)$, and in particular on the accuracy of the normal approximation, each panel in Figure 1 plots $P(x)$ against n_1 for $x = \Phi^{-1}(\alpha) = z_\alpha$ and $\alpha = 0.95$, 0.85 and 0.75. The top left panel of Figure 1 shows results for $\rho_0 = 1$ (indicated by the lines with circles) and $\rho_0 = 2$ (lines with dots), and it is clear that in both cases there is significant oscillatory behaviour, arising principally from the term in $K_n(x)$ in (2.14). The top right panel of Figure 1 shows that these oscillations decline markedly, and the accuracy of the normal approximation improves considerably, if $\rho_0 = 2^{1/2}$. This property reflects the results reported in Section 2.

Of course, $\rho_0 = 2^{1/2}$ is an algebraic irrational. The bottom left panel of Figure 1 shows that broadly similar values of $P(x)$, although with somewhat more oscillation (reflecting the relatively low upper bounds given in Theorem 1), are obtained for $\rho_0 = \pi/2$, a transcendental irrational whose type is bounded above by 6.61. The bottom right panel of Figure 1 addresses one of the results reported in Section 2.2, specifically that there may be less oscillatory behaviour when ρ_{12} converges slowly to a rational number than when it converges quickly. We consider the cases $n_2 = n_1 + [n_1^{1/5}]$ and $n_2 = n_1 + [n_1^{3/5}]$, where $[x]$ denotes the integer nearest to x. In the first case, ρ_{12} converges relatively quickly to 1, and in the second case the convergence is relatively slow. Figure 1 demonstrates that, as anticipated, the oscillatory behaviour is less pronounced, and the normal approximation better, in the "slow" case.

Finally, Figure 2 shows that broadly similar results are obtained for coverage probabilities of percentile bootstrap confidence intervals for $E(S)$. Let s_α denote the α-level quantile of the distribution of $S - E(S)$, and let $\hat{s}_\alpha$, the parametric bootstrap estimator of s_α, be the α-level quantile of the distribution of $S^* - S$ given $\mathcal{X}$, i.e. $\hat{s}_\alpha = \inf\{s : P(S^* - S \le s \mid \mathcal{X}) \ge \alpha\}$. A naive α-level one-sided percentile-bootstrap confidence interval for $E(S)$, with nominal coverage probability $1 - \alpha$, is given by

$$\mathcal{I}_\alpha = (-\infty, S - \hat{s}_\alpha]. \tag{3.2}$$

In the figure, we give plots of estimates of the coverage probability $P\{E(S) \in \mathcal{I}_\alpha\}$ of $\mathcal{I}_\alpha$ against n_1, estimated using 10^5 Monte-Carlo simulations, for $\alpha = 0.95$. We used $B = 9999$ simulations in each bootstrap step. Each panel depicts the case $\rho_0 = 1$, and successive panels also give results when $\rho_0 = 3^{1/2}$, $5^{1/2}$, e and $\phi = (1 + 5^{1/2})/2$, respectively. Each of these values of ρ_0 is an irrational of type 1, and in each instance the oscillations are markedly less, and the normal approximation markedly improved, relative to the case $\rho_0 = 1$.

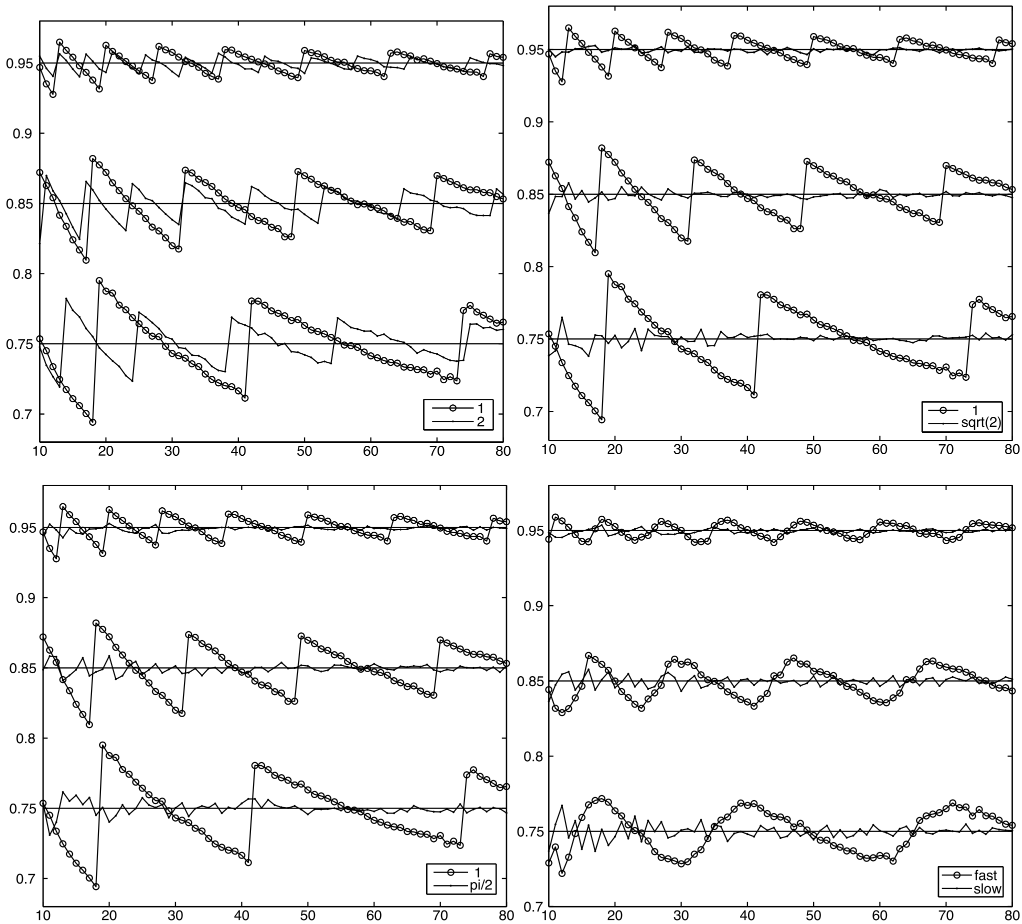

Figure 1. Plots of $P(x)$ against n_1. Plots are given for $x = \Phi^{-1}(\alpha) = z_\alpha$ and $\alpha = 0.95$, 0.85, and 0.75, and for n_2 equal to the nearest integer to $\rho_0 n_1$, with $\rho_0 = 1$ or 2 (top left), ρ_0= 1 or $2^{1/2}$ (top right), ρ_0= 1 or $\pi/2$ (bottom left) and ρ_0 converges to 1 rapidly or slowly (bottom right; see text for details).

4. Proofs

4.1. Proof of Theorem 1

4.1.1. *Proof of part* (*i*) *of Theorem* 1

Here we show that if (2.5) holds for some $\rho_{j_1 j_2}$, where $j_1 \neq j_2$, then (2.2) obtains. Some of the asymptotic expansions in our argument are taken a little further than is necessary for (2.2); the extra detail will be used in the proof of Theorem 2.

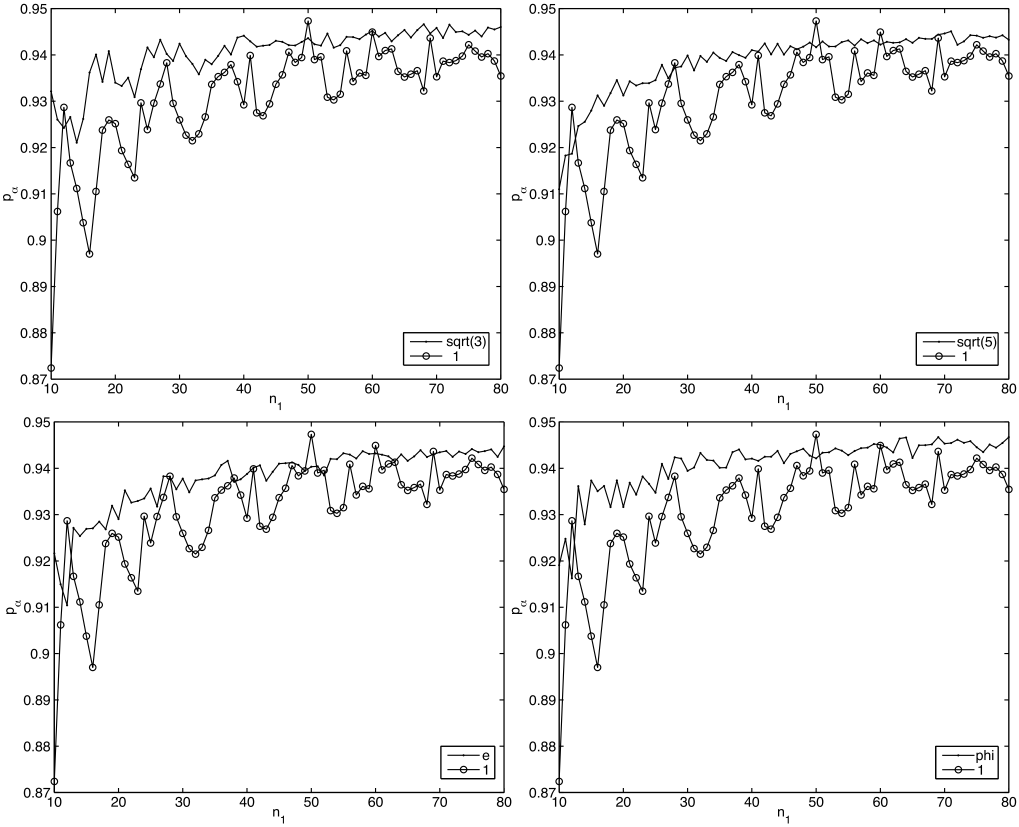

Figure 2. Plots of estimates of $P\{E(S) \in \mathcal{I}_\alpha\}$, against n_1; see text for details. Each panel shows the case ρ_0= 1 and also, in respective panels, the cases $\rho_0 = 3^{1/2}$, $\rho_0 = 5^{1/2}$, $\rho_0 = e$ and $\rho_0 = (1 + 5^{1/2})/2$. Throughout, $x = \Phi^{-1}(\alpha)$ where $\alpha = 0.95$.

Step 1: Proof that it is sufficient to consider the case $k = 2$. Without loss of generality, (2.5) holds for ρ_{12}, and in this case we write $S - E(S) = S_1 + S_2$, where $S_1 = (1 - E)(\bar{X}_1 + \bar{X}_2)$ and $S_2 = (1 - E)(\bar{X}_3 + \cdots + \bar{X}_k)$, where E denotes the expectation operator. Recall that S_2 is independent of $\bar{X}_1$ and $\bar{X}_2$. Suppose we can prove that, analogously to (2.2),

$$P\left\{\frac{S_1}{(\operatorname{var} S_1)^{1/2}} \le x\right\} = \Phi(x) + n^{-1/2}\frac{1}{6}\beta_1\bigl(1 - x^2\bigr)\phi(x) + o\bigl(n^{-1/2}\bigr), \tag{4.1}$$

uniformly in x, where, reflecting (2.3),

$$\beta_1 = \beta_1(n) = \frac{n^{1/2}E(S_1^3)}{(\operatorname{var} S_1)^{3/2}}.$$

If we prove that (2.2), in the case of general k, follows from (4.1), we shall have shown that it is sufficient to derive Theorem 1 the case $k = 2$.

Since $P(S \le x) = E\{P(S_1 \le x - S_2 \mid S_2)\}$ then we can deduce from (4.1) that

$$P(S \le x) = E\left(\Phi\left\{\frac{x - S_2}{(\operatorname{var} S_1)^{1/2}}\right\} + \frac{\beta_1}{6n^{1/2}}\left[1 - \left\{\frac{x - S_2}{(\operatorname{var} S_1)^{1/2}}\right\}^2\right] \times \phi\left\{\frac{x - S_2}{(\operatorname{var} S_1)^{1/2}}\right\}\right) + o\bigl(n^{-1/2}\bigr), \tag{4.2}$$

uniformly in x. Let $R = S_2/(\operatorname{var} S_1)^{1/2}$, and put $\tau_1^2 = \operatorname{var}(R)$, which is bounded away from zero and infinity as $n \to \infty$. It is straightforward to prove that, if N denotes a normally distributed random variable with the same mean (i.e., zero mean) and variance as S_2, then

$$E\left(\left[1 - \left\{\frac{x - S_2}{(\operatorname{var} S_1)^{1/2}}\right\}^2\right]\phi\left\{\frac{x - S_2}{(\operatorname{var} S_1)^{1/2}}\right\}\right)$$

$$= \int\left[1 - \left\{\frac{x}{(\operatorname{var} S_1)^{1/2}} - t\right\}^2\right]\phi\left\{\frac{x}{(\operatorname{var} S_1)^{1/2}} - t\right\} dP(R \le t) \tag{4.3}$$

$$= \int\left[1 - \left\{\frac{x}{(\operatorname{var} S_1)^{1/2}} - t\right\}^2\right]\phi\left\{\frac{x}{(\operatorname{var} S_1)^{1/2}} - t\right\}\frac{1}{\tau_1}\phi\left(\frac{t}{\tau_1}\right) dt + O\bigl(n^{-1/2}\bigr) \tag{4.4}$$

$$= E\left(\left[1 - \left\{\frac{x - N}{(\operatorname{var} S_1)^{1/2}}\right\}^2\right]\phi\left\{\frac{x - N}{(\operatorname{var} S_1)^{1/2}}\right\}\right) + O\bigl(n^{-1/2}\bigr), \tag{4.5}$$

uniformly in x. The passage from (4.3) to (4.4) can be accomplished by integrating by parts in (4.3), then using an Edgeworth expansion of the distribution of R, then separating out the term in $n^{-1/2}$ in that expansion, and finally, undoing the integration by parts as it applies to the leading term in the Edgeworth expansion.

Let $\tau_2^2 = \operatorname{var} S_2$ and $\beta_2 = \beta_2(n) = n^{1/2} E(S_2^3)/\tau_2^3$. If

$$\Phi(r/\tau_2) + n^{-1/2}\tfrac{1}{6}\beta_2\{1 - (r/\tau_2)^2\}\phi(r/\tau_2)$$

represents the two-term Edgeworth approximation to $P(S_2 \le r)$ that would be employed if the distribution of S_2 were continuous, then it can be proved that, uniformly in x,

$$E\left[\Phi\left\{\frac{x - S_2}{(\operatorname{var} S_1)^{1/2}}\right\}\right] = \int \Phi\left\{\frac{x - r}{(\operatorname{var} S_1)^{1/2}}\right\} d_r\left\{\Phi(x/\tau_2) + n^{-1/2}\frac{1}{6}\beta_2\{1 - (r/\tau_2)^2\}\phi(r/\tau_2)\right\} + \begin{cases} o\bigl(n^{-1/2}\bigr) & \text{if } \max_j E|X_{j1}|^3 < \infty \\ O\bigl(n^{-1}\bigr) & \text{if } \max_j E|X_{j1}|^4 < \infty. \end{cases} \tag{4.6}$$

To derive (4.6), first integrate by parts on the left-hand side, writing it as

$$\frac{1}{(\operatorname{var} S_1)^{1/2}} \int \phi\left\{\frac{x-r}{(\operatorname{var} S_1)^{1/2}}\right\} P(S_2 \le r)\, dr = \int \phi\left\{\frac{x}{(\operatorname{var} S_1)^{1/2}} - t\right\} P(R \le t)\, dt. \tag{4.7}$$

Next, write down an Edgeworth expansion, (E) say, for the joint distribution of $\bar{X}_3, \dots, \bar{X}_k$, up to terms of $o(n^{-1/2})$ when $\max_j E|X_{j1}|^3 < \infty$ and $O(n^{-1})$ when $\max_j E|X_{j1}|^4 < \infty$. The expansion will include the conventional discontinuous terms of size $n^{-1/2}$. Use (E) to the derive discontinuous term $n^{-1/2}D$, say, up to a remainder of smaller order $n^{-1/2}$, in an Edgeworth expansion of the distribution of R. Since the function ϕ is smooth, the impact of $n^{-1/2}D$ on the right-hand side of (4.7) equals $O(n^{-1})$, this being obtained by multiplying together the factor $n^{-1/2}$ and another term of order $n^{-1/2}$ that results from integrating D against a smooth function. Therefore, (4.6) holds.

Combining (4.2), (4.5) and (4.6), we deduce that

$$P(S \le x) = \int \left(\Phi\left\{\frac{x-r}{(\operatorname{var} S_1)^{1/2}}\right\} + \frac{\beta_1}{6n^{1/2}}\left[1 - \left\{\frac{x-r}{(\operatorname{var} S_1)^{1/2}}\right\}^2\right]\phi\left\{\frac{x-r}{(\operatorname{var} S_1)^{1/2}}\right\}\right) \times d_r\left\{\Phi(x/\tau_2) + n^{-1/2}\frac{1}{6}\beta_2\{1-(r/\tau_2)^2\}\phi(r/\tau_2)\right\} + o(n^{-1/2}), \tag{4.8}$$

uniformly in x. Result (4.8) is equivalent to (2.2), and so (4.2), representing (2.2) in the case $k = 2$, implies (2.2) for general $k \ge 2$, as had to be shown.

Step 2: Proof of (2.2) when $k = 2$. In this section, we shall show that, if $k = 2$ and (2.5) holds for $\rho_{12} = e_2 n_1/(e_1 n_2)$, then (2.2) holds.

To this end, define

$$T = (S - ES)/(\operatorname{var} S)^{1/2} = \frac{\bar{X}_1 + \bar{X}_2 - \mu_1 - \mu_2}{(n_1^{-1}\sigma_1^2 + n_2^{-1}\sigma_2^2)^{1/2}} = c_1 T_1 + c_2 T_2,$$

where $T_j = (\bar{X}_j - \mu_j)/(n_j^{-1}\sigma_j^2)^{1/2}$ and c_1 and c_2 are defined as at (2.9). In this notation,

$$\begin{aligned} P(T \le x) &= P(c_1T_1 + c_2T_2 \le x) = E\{P(c_1T_1 \le x - c_2T_2 \mid T_2)\} \\ &= E\left\{\Phi\left(\frac{x - c_2T_2}{c_1}\right) + n_1^{-1/2}A_1\left(\frac{x - c_2T_2}{c_1}\right) + n_1^{-1/2}D_1\left(\frac{x - c_2T_2}{c_1}\right)\right\} \\ &\quad + \begin{cases} o(n^{-1/2}) & \text{if } \max_j E|X_{j1}|^3 < \infty \\ O(n^{-1}) & \text{if } \max_j E|X_{j1}|^4 < \infty, \end{cases} \end{aligned} \tag{4.9}$$

where A_j and D_j will refer to the smooth and discontinuous terms, respectively, in the $n_j^{-1/2}$ component of an Edgeworth expansion of the distribution of T_j for $j = 1, 2$. In particular,

$n_j^{-1/2} A_j$ and $n_j^{-1/2} D_j$ are the counterparts of the second and third terms, respectively, on the right-hand side of formula (35) p. 56 of Esseen [9].

Writing B for either Φ or A_1, appearing on the right-hand side of (4.9), we have:

$$E\left\{B\left(\frac{x - c_2 T_2}{c_1}\right)\right\} = \int B\left(\frac{x - c_2 u}{c_1}\right) dP(T_2 \le u)$$
$$= \frac{c_2}{c_1} \int B'\left(\frac{x - c_2 u}{c_1}\right) P(T_2 \le u)\, du.$$

As in the argument leading to (4.6) it can be shown that the discontinuous term $n_2^{-1/2} D_2$, in the Edgeworth expansion of $P(T_2 \le x)$, contributes only $O(n^{-1})$. Therefore, if we write $\mathcal{E}_2(u)$ for the Edgeworth approximation to $P(T_2 \le u)$ that includes the leading Gaussian term, plus the continuous part of the component of order $n_2^{-1/2}$, and neglects everything else, we deduce from (4.9) that

$$P(T \le x) = \int \left\{\Phi\left(\frac{x - c_2 u}{c_1}\right) + n_1^{-1/2} A_1\left(\frac{x - c_2 u}{c_1}\right)\right\} d_u \mathcal{E}_2(u)$$
$$+ n_1^{-1/2} E\left\{D_1\left(\frac{x - c_2 T_2}{c_1}\right)\right\} + \begin{cases} o\left(n^{-1/2}\right) & \text{if } \max_j E|X_{j1}|^3 < \infty \\ O\left(n^{-1}\right) & \text{if } \max_j E|X_{j1}|^4 < \infty. \end{cases} \tag{4.10}$$

Now we turn our attention to:

$$E\left\{D_1\left(\frac{x - c_2 T_2}{c_1}\right)\right\} = \int D_1\left(\frac{x - c_2 u}{c_1}\right) dP(T_2 \le u)$$
$$= -\int P(T_2 \le u)\, d_u D_1\left(\frac{x - c_2 u}{c_1}\right) \tag{4.11}$$
$$= I_1(x) + n_2^{-1/2} I_2(x) + \begin{cases} o(1) & \text{if } \max_j E|X_{j1}|^3 < \infty \\ O\left(n^{-1/2}\right) & \text{if } \max_j E|X_{j1}|^4 < \infty, \end{cases}$$
$$= n_2^{-1/2} I_2(x) + \begin{cases} o(1) & \text{if } \max_j E|X_{j1}|^3 < \infty \\ O\left(n^{-1/2}\right) & \text{if } \max_j E|X_{j1}|^4 < \infty, \end{cases} \tag{4.12}$$

where

$$I_1(x) = \int D_1\left(\frac{x - c_2 u}{c_1}\right) \phi(u)\, du, \qquad I_2(x) = \int D_1\left(\frac{x - c_2 u}{c_1}\right) dD_2(u). \tag{4.13}$$

To obtain the third identity in the string of formulae leading to (4.12), we used the integration by parts step at (4.11), a short Taylor expansion of $P(T_2 \le u)$ with a remainder of $o(n^{-1/2})$ if $\max_j E|X_{j1}|^3 < \infty$ and $O(n^{-1})$ if $\max_j E|X_{j1}|^4$, and the fact that $\int |dD_1| = O(n^{1/2})$ uniformly in x. (This can be deduced either directly or by making use of (4.14) below.) Finally, it can be shown, arguing as in the proof of (4.6), that $I_1(x) = O(n^{-1/2})$, from which (4.12) follows.

Note too that, with σ_j defined as immediately above (2.1),

$$D_j(x) = \frac{e_j}{\sigma_j}\psi\left\{\frac{(x-\xi_{1n})\sigma_j n_j^{1/2}}{e_j}\right\}\phi(x)$$

$$= \frac{e_j}{\sigma_j}\psi\left[\frac{\sigma_j n_j^{1/2} x}{e_j} - \left\{(n_j x_j/e_j) - \lfloor n_j x_j/e_j \rfloor\right\}\right]\phi(x), \tag{4.14}$$

where, as in Sections 1 and 2, $\psi(x) = \lfloor x \rfloor - x + \frac{1}{2}$, $\lfloor x \rfloor$ is the largest integer not strictly exceeding x, and $\xi_{jn} = e_j(\sigma_j n_j^{1/2})^{-1}\{(n_j x_j/e_j) - \lfloor n_j x_j/e_j \rfloor\}$ if the lattice is located at points $x_j + \nu e_j$ for integers ν, see Esseen [9], (29), (31) and (35) pp. 55/56. Defining $\gamma = (e_1 e_2/\sigma_1\sigma_2)$, as in Section 2.3; putting

$$\psi_j(x) = \psi\left\{\frac{\sigma_j n_j^{1/2}}{e_j}(x - \xi_{jn})\right\};$$

and noting that, by (4.14), $D_j(x) = (e_j/\sigma_j)\psi_j(x)\phi(x)$; we deduce that

$$I_2(x)/\gamma = \frac{1}{\gamma}\int D_1\left(\frac{x-c_2u}{c_1}\right) dD_2(u) = \int (\psi_1\phi)\left(\frac{x-c_2u}{c_1}\right) d\left\{\psi_2(u)\phi(u)\right\}$$

$$= \int (\psi_1\phi)\left(\frac{x-c_2u}{c_1}\right)\left\{\phi(u)\, d\psi_2(u) + \psi_2(u)\, d\phi(u)\right\}$$

$$= \int (\psi_1\phi)\left(\frac{x-c_2u}{c_1}\right)\phi(u)\, d\psi_2(u) + \frac{1}{2}\int (\psi_1\phi)\left(\frac{x-c_2u}{c_1}\right)\psi_2(u)\, d\phi(u).$$

The last-written integral equals $O(1)$, uniformly in x, and so, with I_2 as at (4.13),

$$I_2(x) = \gamma I_3(x) + O(1), \tag{4.15}$$

uniformly in x, where

$$I_3(x) = \int (\psi_1\phi)\left(\frac{x-c_2u}{c_1}\right)\phi(u)\, d\psi_2(u).$$

Since ψ_2 has jumps of size $+1$ at points u where $(u-\xi_{2n})\sigma_2 n_2^{1/2}/e_2$ is an integer, i.e. $u = u_\nu \equiv \xi_{2n} + e_2(\sigma_2 n_2^{1/2})^{-1}\nu$ for an integer ν, then

$$I_3(x) = \sum_\nu (\psi_1\phi)\left(\frac{x-c_2u_\nu}{c_1}\right)\phi(u_\nu)$$

$$= \sum_\nu \phi\left(\frac{x}{c_1} - \frac{c_2}{c_1}\left\{\xi_{2n} + e_2\left(\sigma_2 n_2^{1/2}\right)^{-1}\nu\right\}\right)\phi\left\{\xi_{2n} + e_2\left(\sigma_2 n_2^{1/2}\right)^{-1}\nu\right\}$$

$$\times\ \psi\left\{\xi_n(x) - \frac{e_2 n_1}{e_1 n_2}\nu\right\}, \tag{4.16}$$

where ξ_n is as at (2.8) and we have used the fact that

$$\psi_1\left(\frac{x-c_2u_\nu}{c_1}\right)=\psi\left[\frac{\sigma_1 n_1^{1/2}}{e_1}\left\{(x-c_2u_\nu)c_1^{-1}-\xi_{1n}\right\}\right]$$
$$=\psi\left\{\frac{(x-c_2u_\nu-c_1\xi_{1n})\sigma_1 n_1^{1/2}}{c_1e_1}\right\}$$
$$=\psi\left[\frac{\{x-(c_1\xi_{1n}+c_2\xi_{2n})-c_2e_2(\sigma_2 n_2^{1/2})^{-1}\nu\}\sigma_1 n_1^{1/2}}{c_1e_1}\right]$$
$$=\psi\left\{\xi_n(x)-\frac{e_2n_1}{e_1n_2}\nu\right\},$$

with

$$\xi_n(x)=\left\{x-(c_1\xi_{1n}+c_2\xi_{2n})\right\}\frac{\sigma_1 n_1^{1/2}}{c_1e_1}.$$

Recall that $\xi_{jn}=e_j(\sigma_j n_j^{1/2})^{-1}\{(n_jx_j/e_j)-\lfloor n_jx_j/e_j\rfloor\}$ if the lattice is located at points $x_j+\nu e_j$ for integers ν. In particular, $\xi_{jn}=O(n^{-1/2})$ for $j=1,2$. Therefore, Taylor expanding the arguments of the functions ϕ at (4.16), and defining

$$I_4(x)=\sum_\nu \phi\left(\frac{x}{c_1}-\frac{e_2n_1^{1/2}}{\sigma_1n_2}\nu\right)\phi\left\{e_2\left(\sigma_2n_2^{1/2}\right)^{-1}\nu\right\}\psi\left\{\xi_n(x)-\frac{e_2n_1}{e_1n_2}\nu\right\}, \tag{4.17}$$

we deduce that

$$I_3(x)=I_4(x)+O(1), \tag{4.18}$$

uniformly in x. Combining (4.10), (4.12), (4.15) and (4.18), we deduce that

$$P(T\le x)=\int\left\{\Phi\left(\frac{x-c_2u}{c_1}\right)+n_1^{-1/2}A_1\left(\frac{x-c_2u}{c_1}\right)\right\}d_u\mathcal{E}_2(u)$$
$$+(n_1n_2)^{-1/2}\gamma I_4(x)+\begin{cases}o\left(n^{-1/2}\right) & \text{if } \max_j E|X_{j1}|^3<\infty\\ O\left(n^{-1}\right) & \text{if } \max_j E|X_{j1}|^4<\infty.\end{cases} \tag{4.19}$$

If we can show that

$$\sup_{-\infty<x<\infty}\left|I_4(x)\right|=o\left(n^{1/2}\right) \tag{4.20}$$

then it will follow from (4.19), in cases where $\max_j E|X_{j1}|^3<\infty$, that

$$P(T\le x)=\int\left\{\Phi\left(\frac{x-c_2u}{c_1}\right)+n_1^{-1/2}A_1\left(\frac{x-c_2u}{c_1}\right)\right\}d_u\mathcal{E}_2(u)+o\left(n^{-1/2}\right). \tag{4.21}$$

The right-hand side here is Edgeworth expansion we would expect the distribution of T to enjoy if we were able to ignore the latticeness of the distributions of X_{j1} for $j = 1, 2$. That is, (4.21) is just (2.2) in the particular case $k = 2$. Therefore, provided (4.20) holds then we shall have shown that (2.2) holds whenever $k = 2$. It remains to derive (4.20).

Step 3: Proof of (4.20). Given $\varepsilon > 0$, partition the set of all integers into adjacent blocks $\mathcal{N}_\ell$, for $-\infty < \ell < \infty$, where each block consists of just $2\lfloor n^{1/2}\varepsilon\rfloor + 1$ consecutive integers, and the central integer is denoted by $\bar{\nu}_\ell$. Recalling the definition of $I_4(x)$ at (4.17), we deduce that

$$I_4 = \sum_{-\infty<\ell<\infty} J_{1,\ell}, \tag{4.22}$$

where

$$J_{1,\ell}(x) = \sum_{\nu\in\mathcal{N}_\ell} \phi\left(\frac{x}{c_1} - \frac{e_2 n_1^{1/2}}{\sigma_1 n_2}\nu\right)\phi\left\{e_2\left(\sigma_2 n_2^{1/2}\right)^{-1}\nu\right\}\psi\left\{\xi_n(x) - \frac{e_2 n_1}{e_1 n_2}\nu\right\}. \tag{4.23}$$

Now,

$$J_{1,\ell} = J_{2,\ell} + R_\ell, \tag{4.24}$$

where

$$J_{2,\ell}(x) = \phi\left(\frac{x}{c_1} - \frac{e_2 n_1^{1/2}}{\sigma_1 n_2}\bar{\nu}_\ell\right)\phi\left\{e_2\left(\sigma_2 n_2^{1/2}\right)^{-1}\bar{\nu}_\ell\right\}\sum_{\nu\in\mathcal{N}_\ell}\psi\left\{\xi_n(x) - \frac{e_2 n_1}{e_1 n_2}\nu\right\} \tag{4.25}$$

and R_ℓ is defined naively by (4.24). Given an integer r, let $\ell(r)$ denote the unique value of ℓ such that $r \in \mathcal{N}_\ell$. Then, since $|\psi| \le 1$,

$$\begin{aligned}\left|\sum_{-\infty<\ell<\infty} R_\ell\right| &\le \sum_r \left|\phi\left(\frac{x}{c_1} - \frac{e_2 n_1^{1/2}}{\sigma_1 n_2}r\right)\phi\left\{e_2\left(\sigma_2 n_2^{1/2}\right)^{-1}r\right\}\right. \\ &\qquad \left. - \phi\left(\frac{x}{c_1} - \frac{e_2 n_1^{1/2}}{\sigma_1 n_2}\bar{\nu}_{\ell(r)}\right)\phi\left\{e_2\left(\sigma_2 n_2^{1/2}\right)^{-1}\bar{\nu}_{\ell(r)}\right\}\right| \\ &\le C_1\varepsilon n^{1/2},\end{aligned} \tag{4.26}$$

where the constant C_1 does not depend on ε or n.

Let $\rho = e_2 n_1/(e_1 n_2)$, and define

$$\chi_{\mathcal{N}}(z, \rho) \equiv \sum_{\nu\in\mathcal{N}} \psi(z - \rho\nu).$$

In this notation,

$$J_{2,\ell} = \phi\left(\frac{x}{c_1} - \frac{e_2 n_1^{1/2}}{\sigma_1 n_2}\bar{\nu}_\ell\right)\phi\left\{e_2\left(\sigma_2 n_2^{1/2}\right)^{-1}\bar{\nu}_\ell\right\}\chi_{\mathcal{N}_\ell}\left\{\xi_n(x), \rho\right\}. \tag{4.27}$$

If we can prove that, whenever the set $\mathcal{N}$ consists of $|\mathcal{N}|$ consecutive integers and $C_2 < C_3$ are positive constants,

$$\sup_{C_2 n^{1/2} \le |\mathcal{N}| \le C_3 n^{1/2}} \sup_{-\infty < z < \infty} \left|\chi_{\mathcal{N}}(z, \rho)\right| = o\left(n^{1/2}\right) \tag{4.28}$$

as $|\mathcal{N}| \to \infty$, then it will follow from (4.27) that

$$\begin{aligned} \left|\sum_{-\infty<\ell<\infty} J_{2,\ell}\right| &= o\left[n^{1/2} \sum_{-\infty<\ell<\infty} \phi\left(\frac{x}{c_1} - \frac{e_2 n_1^{1/2}}{\sigma_1 n_2}\bar{\nu}_\ell\right)\phi\left\{e_2\left(\sigma_2 n_2^{1/2}\right)^{-1}\bar{\nu}_\ell\right\}\right] \\ &= o\left(n^{1/2}\right), \end{aligned} \tag{4.29}$$

for each $\varepsilon > 0$, since the series on the first right-hand side of (4.29) is bounded uniformly in n. (To appreciate why, observe that $\bar{\nu}_\ell$ is approximately an integer multiple of $n^{1/2}$, plus a constant.) Note that, since the left-hand side of (4.28) involves the supremum over z, then that quantity does not depend on the location of the set $\mathcal{N}$ on the line, only on the number of consecutive integers it contains.

The desired result (4.20) follows from (4.22), (4.24), the fact that (4.26) holds for each $\varepsilon > 0$, and (4.29). To complete the proof of (4.20), we shall derive (4.28). Specifically, we shall prove that, in cases where (2.5) is satisfied for $\rho_{12} = \rho = e_2 n_1/(e_1 n_2)$, (4.28) obtains.

Assume that $\mathcal{N}$ consists of p consecutive integers, where $C_2 n^{1/2} \le p \le C_3 n^{1/2}$. Koksma's inequality (see, e.g., Theorems 1.3 and 5.1, pp. 91 and 143 of [13]), and the Erdős-Turán inequality (see, e.g., formula (2.42), p. 114 of [13]), can be combined to prove that, for all integers $m \ge 1$,

$$\begin{aligned} \sup_z\left|\chi_{\mathcal{N}}(z, \rho_{12})\right| &\le C_4\left\{\frac{p}{m} + \sum_{\ell=1}^{m}\frac{1}{\ell}\sup_z\left|\sum_{r=1}^{p}\exp(2\pi i \ell r \rho_{12})\right|\right\} \\ &\le C_4\left\{\frac{p}{m} + \sum_{\ell=1}^{m}\frac{1}{\ell|\sin(\ell\rho_{12}\pi)|}\right\}, \end{aligned} \tag{4.30}$$

where C_4 is an absolute constant. Since (2.5) is assumed to hold with $(j_1, j_2) = (1, 2)$ then, for each fixed m,

$$\max_{1\le\ell\le m}\left|\sin(\ell\rho_{12}\pi)\right|^{-1} = o\left(n^{1/2}\right).$$

Hence, by (4.30),

$$\sup_z\left|\chi_{\mathcal{N}}(z, \rho_{12})\right| \le \frac{C_3 C_4 n^{1/2}}{m} + o\left(n^{1/2}\right), \tag{4.31}$$

where the $o(n^{1/2})$ term is of that order uniformly in $\mathcal{N}$ such that $C_2 n^{1/2} \le |\mathcal{N}| \le C_3 n^{1/2}$. However, m can be taken arbitrarily large, and none of C_2, C_3 and C_4 depends on m or n. Therefore, (4.31) implies (4.28).

4.1.2. *Proof of part* (*ii*) *of Theorem* 1

We can write

$$\bar{X}_1 + \cdots + \bar{X}_k = \frac{e_1}{n_1}(Y_1 + \cdots + Y_k) + \mu,$$

where μ is deterministic and, for each j, Y_j is the sum of n_j random variables $Y_{j1}, \ldots, Y_{jn_j}$, each having a lattice distribution (not depending on n) supported on the set of points $\rho_{1j}\ell$ for $\ell \in \mathbb{Z}$, and with the Y_{ji}s being totally independent. Of course, $\rho_{11} = 1$. Since each $\rho_{j_1 j_2}$ equals a rational number, not depending on n, then the set $\bigcup_j \{\rho_{1j}\ell, \ell \in \mathbb{Z}\}$ can itself be represented as a maximal lattice, $\mathcal{L}$ say, not depending on n. The distribution of

$$Y_1 + \cdots + Y_k = \sum_{j=1}^{k} \sum_{i=1}^{n_j} Y_{ji}$$

can be viewed as the distribution of the sum of $n = n_1 + \cdots + n_k$ independent and identically distributed random variables each having a mixture distribution, D_n say, with support confined to $\mathcal{L}$. Although D_n depends on n, since it is always supported on the same lattice, standard methods can be used to derive an Edgeworth expansion of the distribution of $Y_1 + \cdots + Y_k$, from which it can be seen that there is a nonvanishing discontinuous term, not present in (2.2).

4.2. Proof of Theorem 2

Step 1: Proof that it is sufficient to consider the case $k = 2$. We give the argument only in outline, since it parallels that in step 1 of the derivation of Theorem 1. Suppose it is possible to derive the version of (4.1) where the remainder $o(n^{-1/2})$ is replaced by $O(n^{\xi-1})$, for all $\xi > 0$. Then, as in the earlier proof, we have (4.2) where the remainder term is $O(n^{\xi-1})$, for all $\xi > 0$, instead of $o(n^{-1/2})$. The string of arguments leading to (4.5) holds without change, as too does (4.6). Combining the revised (4.2) with the old (4.5) and (4.6) we deduce the following version of (4.8):

$$\begin{aligned} P(S \le x) = \int \Biggl(\Phi\biggl\{ \frac{x - r}{(\operatorname{var} S_1)^{1/2}} \biggr\} + \frac{\beta_1}{6n^{1/2}} \biggl[1 - \biggl\{ \frac{x - r}{(\operatorname{var} S_1)^{1/2}} \biggr\}^2 \biggr] \phi\biggl\{ \frac{x - r}{(\operatorname{var} S_1)^{1/2}} \biggr\} \Biggr) \\ \times d\biggl\{ \Phi(x/\tau_2) + n^{-1/2} \frac{1}{6} \beta_2 \bigl\{ 1 - (r/\tau_2)^2 \bigr\} \phi(r/\tau_2) \biggr\} + O\bigl(n^{\xi-1}\bigr), \end{aligned}$$

uniformly in x and for all $\xi > 0$. This formula is equivalent to (2.2), with the remainder there replaced by $O(n^{\xi-1})$, and so we have shown that it suffices to consider $k = 2$.

Step 2: Completion of proof of Theorem 2. Combining (4.10) and (4.12) in the case $\max_j E|X_{j1}|^4 < \infty$, and noting (4.15) and (4.18), we deduce the version of (4.19) when $\max_j E|X_{j1}|^4 < \infty$.

Next we reintroduce the notation noted below (2.8), where $\alpha \in (0, \frac{1}{2})$, $\mathcal{N}_\ell$ (for $-\infty < \ell < \infty$) is a partition of the set of all integers into adjacent blocks each containing $2\lfloor n^\alpha \rfloor + 1$ consecutive integers, $\bar{\nu}_\ell$ is the central integer in $\mathcal{N}_\ell$, and $\nu_\ell = \nu - \bar{\nu}_\ell$ for $\nu \in \mathcal{N}_\ell$. Property (4.22) continues to hold, with $J_{1,\ell}$ still given by (4.23). Again we define R_ℓ and $J_{2,\ell}$ by (4.24) and (4.25). However, this time we give an expansion for, rather than an upper bound to, R_ℓ. As a first step, note that

$$\begin{aligned} R_\ell(x) &= J_{1,\ell}(x) - J_{2,\ell}(x) \\ &= \sum_{\nu \in \mathcal{N}_\ell} \left[\phi\left\{ \frac{x}{c_1} - \frac{e_2 n_1^{1/2}}{\sigma_1 n_2} (\bar{\nu}_\ell + \nu_\ell) \right\} \phi\left\{ e_2 \left(\sigma_2 n_2^{1/2}\right)^{-1} (\bar{\nu}_\ell + \nu_\ell) \right\} \right. \\ &\quad \left. - \phi\left(\frac{x}{c_1} - \frac{e_2 n_1^{1/2}}{\sigma_1 n_2} \bar{\nu}_\ell \right) \phi\left\{ e_2 \left(\sigma_2 n_2^{1/2}\right)^{-1} \bar{\nu}_\ell \right\} \right] \psi\left\{ \xi_n(x) - \frac{e_2 n_1}{e_1 n_2} \nu \right\}. \end{aligned}$$

Taylor-expanding, and using the argument in the paragraph immediately below that containing (2.10), we deduce that

$$\begin{aligned} \sum_{-\infty<\ell<\infty} R_\ell(x) = \sum_{r=1}^{r_0} \sum_{-\infty<\ell<\infty} \frac{\phi_r(\bar{\nu}_\ell / n_1^{1/2}, x)}{r! n_1^{r/2}} \sum_{\nu \in \mathcal{N}_\ell} \nu_\ell^r \psi\left\{ \xi_n(x) - \frac{e_2 n_1}{e_1 n_2} \nu \right\} \\ + O\left(n^\alpha \cdot n^{1/2} \cdot n^{(r_0+1)\{\alpha-(1/2)\}}\right), \end{aligned} \tag{4.32}$$

uniformly in x. Adding $\sum_\ell J_{2,\ell}$ to either side of (4.32) has the effect, on the right-hand side, of changing the range of summation of the first series to $0 \le r \le r_0$. Therefore,

$$\sum_{-\infty<\ell<\infty} \left\{ J_{2,\ell}(x) + R_\ell(x) \right\} = \gamma^{-1} K_n(x) + O\left(n^\alpha \cdot n^{1/2} \cdot n^{(r_0+1)\{\alpha-(1/2)\}}\right), \tag{4.33}$$

uniformly in x, where $\gamma = \prod_{j=1,2}(e_j/\sigma_j)$ and K_n is at (2.10). If $r_0 \ge 4\alpha/(1-2\alpha)$, as stipulated in Theorem 2, then the "O" remainder in (4.33) is just $O(1)$. In this case,

$$\gamma I_4(x) = \gamma \sum_{-\infty<\ell<\infty} J_{1,\ell}(x) = \gamma \sum_{-\infty<\ell<\infty} \left\{ J_{2,\ell}(x) + R_\ell(x) \right\} = K_n(x) + O(1), \tag{4.34}$$

uniformly in x. Part (i) of Theorem 2, which addresses only the case $k=2$, follows from (4.19) and (4.34). Part (ii) of Theorem 2, in the case $k=2$, follows from (4.34) and (2.36). In view of Part 1 of the proof of Theorem 2, this is sufficient to complete the proof of the theorem.

Acknowledgements

We are grateful to Professor Roger Heath-Brown for helpful discussion. The research was supported by the Australian Research Council and the National Science Foundation.

References

[1] Agresti, A. and Caffo, B. (2000). Simple and effective confidence intervals for proportions and differences of proportions result from adding two successes and two failures. *Amer. Statist.* **54** 280–288. MR1814845

[2] Borkowf, C.B. (2006). Constructing binomial confidence intervals and near nominal coverage by adding a single imaginary failure or success. *Stat. Med.* **25** 3679–3695. MR2252419
[3] Brown, L. and Li, X. (2005). Confidence intervals for two sample binomial distribution. *J. Statist. Plann. Inference* **130** 359–375. MR2128014
[4] Brown, L.D., Cai, T.T. and DasGupta, A. (2001). Interval estimation for a binomial proportion. *Statist. Sci.* **16** 101–133. With comments and a rejoinder by the authors. MR1861069
[5] Brown, L.D., Cai, T.T. and DasGupta, A. (2002). Confidence intervals for a binomial proportion and asymptotic expansions. *Ann. Statist.* **30** 160–201. MR1892660
[6] Clopper, C.J. and Pearson, E.S. (1934). The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika* **26** 404–413.
[7] Duffy, D. and Santer, T.J. (1987). Confidence intervals for a binomial parameter based on multistage tests. *Biometrics* **43** 81–94.
[8] Einsiedler, M. and Ward, T. (2011). *Ergodic Theory with a View Towards Number Theory*. *Graduate Texts in Mathematics* **259**. London: Springer London Ltd. MR2723325
[9] Esseen, C.G. (1945). Fourier analysis of distribution functions. A mathematical study of the Laplace-Gaussian law. *Acta Math.* **77** 1–125. MR0014626
[10] Griffiths, M. (2004). Formulae for the convergents to some irrationals. *Math. Gazette* **88** 28–38.
[11] Hall, P. (1982). Improving the normal approximation when constructing one-sided confidence intervals for binomial or Poisson parameters. *Biometrika* **69** 647–652. MR0695211
[12] Hall, P. (1987). On the bootstrap and continuity correction. *J. Roy. Statist. Soc. Ser. B* **49** 82–89. MR0893339
[13] Kuipers, L. and Niederreiter, H. (1974). *Uniform Distribution of Sequences*. New York: Wiley-Interscience [John Wiley & Sons]. MR0419394
[14] Lee, J.J., Serachitopol, D.M. and Brown, B.W. (1997). Likelihood-weighted confidence intervals for the difference of two binomial proportions. *Biometrical J.* **39** 387–407. MR1484813
[15] LeVeque, W.J. (1956). *Topics in Number Theory. Vol.* 1. Reading, MA: Addison-Wesley. MR0080682
[16] Price, R.M. and Bonett, D.G. (2004). An improved confidence interval for a linear function of binomial proportions. *Comput. Statist. Data Anal.* **45** 449–456. MR2050249
[17] Ribenboim, P. (2000). *My Numbers, My Friends: Popular Lectures on Number Theory*. New York: Springer. MR1761897
[18] Roth, K.F. (1955). Rational approximations to algebraic numbers. *Mathematika* **2** 1–20. Corrigendum 168. MR0072182
[19] Roths, S.A. and Tebbs, J.M. (2006). Revisiting Beal's confidence intervals for the difference of two binomial proportions. *Comm. Statist. Theory Methods* **35** 1593–1609. MR2328507
[20] Singh, K. (1981). On the asymptotic accuracy of Efron's bootstrap. *Ann. Statist.* **9** 1187–1195. MR0630102
[21] Sterne, T.E. (1954). Some remarks on confidence or fiducial limits. *Biometrika* **41** 275–278. MR0062387
[22] Wang, H. (2010). The monotone boundary property and the full coverage property of confidence intervals for a binomial proportion. *J. Statist. Plann. Inference* **140** 495–501. MR2558380
[23] Zhou, X.H., Li, C.M. and Yang, Z. (2001). Improving interval estimation of binomial proportions. *Phil. Trans. Roy. Soc. Ser. A* **366** 2405–2418.
[24] Zieliński, W. (2010). The shortest Clopper-Pearson confidence interval for binomial probability. *Comm. Statist. Simulation Comput.* **39** 188–193. MR2784561