Combinatorics, Probability and Computing (2009) 18, 979–1017. © Cambridge University Press 2009
doi:10.1017/S096354830999023X First published online 10 August 2009

Step Size in Stein's Method of Exchangeable Pairs

NATHAN ROSS

Department of Mathematics, University of Southern California,
3620 South Vermont Avenue, KAP 108, Los Angeles, California 90089-2532, USA
(e-mail: nathanfr@usc.edu)

Received 12 August 2008; revised 1 April 2009

Stein's method of exchangeable pairs is examined through five examples in relation to Poisson and normal distribution approximation. In particular, in the case where the exchangeable pair is constructed from a reversible Markov chain, we analyse how modifying the step size of the chain in a natural way affects the error term in the approximation acquired through Stein's method. It has been noted for the normal approximation that smaller step sizes may yield better bounds, and we obtain the first rigorous results that verify this intuition. For the examples associated to the normal distribution, the bound on the error is expressed in terms of the spectrum of the underlying chain, a characteristic of the chain related to convergence rates. The Poisson approximation using exchangeable pairs is less studied than the normal, but in the examples presented here the same principles hold.

1. Introduction

Stein's method has become a powerful tool in approximating probability distributions and proving central limit theorems. The various formulations of the method rely on exploiting the characterizing operator or 'Stein equation' of the distribution. The characterizing operator of a random variable X is an operator S such that, for a specified class of functions A, $\mathbb{E}Sf(Y) = 0$ for all f in A if and only if $Y \stackrel{d}{=} X$. Stein's method can be used to quantitatively bound the difference between two random variables, one of which has a known characterizing operator. In this paper we use theorems that obtain error terms from the characterizing operator through exchangeable pairs. There are other variations of Stein's method that exploit the characterizing operator in different ways, for example the zero bias transformation [23, 24], the size bias coupling [4, 7, 25], dependency graphs [1, 2], and other *ad hoc* methods [8, 11, 12, 14].

For a gentle, intuitive explanation of Stein's method for normal approximation, see [31]. A more rigorous introduction can be found in [36] and similar ideas with more references in [29]. To find an introduction to Stein's method of exchangeable pairs for Poisson approximation, see [13]. For a thorough modern introduction to Stein's method in general, see [6].

https://doi.org/10.1017/S096354830999023X Published online by Cambridge University Press

An exchangeable pair is a pair of identically distributed random variables (W, W') with the property that the distribution of (W, W') is equal to the distribution of (W', W). The typical method used to generate useful exchangeable pairs on a finite space Ω is through a Markov chain $\{X_0, X_1, \ldots\}$ on Ω reversible with respect to π. If W is a random variable on Ω, then it is easy to show that setting $W := W(X_0)$ and $W' := W(X_1)$, where X_0 is chosen according to π, is an exchangeable pair. This is not the only way to obtain an exchangeable pair: for exchangeable pairs from non-reversible Markov chains, see [19] and [30].

We will examine how modifying the step size of the underlying Markov chain in a natural way affects the error term in the approximation acquired through Stein's method. In the case where the underlying Markov chain is ergodic, the step size does not necessarily affect the rate of convergence to stationarity in a monotone way. However, the rate of convergence is related to the eigenvalues of the Markov chain, and in the examples associated to the normal distribution we are able to express the bound on the error in terms of the eigenvalues.

It will be obvious in the sequel that modifying the exchangeable pair has a profound effect on the error term. Most notably, in the theorems we use, Markov chains with larger steps require more computational work, and in the case of Poisson approximation, higher moment information. For the examples presented here, the chains that allow for the easiest computation always yield the best bound. For other examples it is difficult to compute the error term for any chain other than the most computationally simple in a form that yields information about the relative sizes of the bounds. Thus, it is difficult to take these examples and make a rigorous statement about the step size of the underlying Markov chain and the bound acquired from it in a general setting.

Section 2 introduces Stein's method of exchangeable pairs for normal approximation and explains why the effect of step size on the error term is not obvious. Sections 3, 4, and 5 each contain one example of Stein's method's approximation of respectively, the binomial (with $p = 1/2$), the Plancherel measure of the Hamming scheme, and the Plancherel measure of the irreducible representations of a group G by the normal distribution. Section 6 is tangent to Sections 3, 4, and 5 in that the bounds on the error from those sections are restated in terms of the eigenvalues of the chain. The final two sections examine the approximation of the binomial and the negative binomial by the Poisson distribution.

This is a small step in examining and refining Stein's method to be more widely applicable. This paper serves to illustrate the type of computations needed to apply the method and will hopefully yield some insight into the underlying theory behind Stein's method.

2. Normal approximation

For normal approximation, we use Stein's original theorem [36], not only because most other exchangeable pair formulations stem from it, but because in many situations it still yields the best results. Also, the theorem is stated in terms of the Kolmogorov metric, but the relative size of the bound on the error is determined by the same terms in an analogous theorem where the Wasserstein metric is used [15].

Theorem 2.1 ([36]). *Let (W, W') be an exchangeable pair of real random variables such that $\mathbb{E}(W' \mid W) = (1 - a)W$ with $0 < a < 1$. Also, let $\mathbb{E}(W) = 0$ and $\mathbb{E}(W^2) = 1$. Then,*

https://doi.org/10.1017/S096354830999023X Published online by Cambridge University Press

for all x_0 in $\mathbb{R}$,

$$\left|\mathbb{P}(W < x_0) - \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x_0} e^{\frac{-x^2}{2}}dx\right|$$
$$\leqslant \frac{\sqrt{\mathrm{Var}(\mathbb{E}[(W'-W)^2\,|\,W])}}{a} + (2\pi)^{-\frac{1}{4}}\sqrt{\frac{\mathbb{E}|W'-W|^3}{a}} \tag{2.1}$$
$$\leqslant \frac{\sqrt{\mathrm{Var}(\mathbb{E}[(W'-W)^2\,|\,W])}}{a} + \left[\frac{\mathbb{E}(W'-W)^4}{\pi a}\right]^{1/4}. \tag{2.2}$$

Remarks. **(1)** The first bound (2.1) is from [36], while (2.2) follows from the Cauchy–Schwarz inequality and the straightforward computation $\mathbb{E}(W'-W)^2 = 2a\,\mathrm{Var}(W)$ for an exchangeable pair (W, W') with $\mathbb{E}(W'\,|\,W) = (1-a)W$.

(2) A result holds [30] which contains error terms similar to Theorem 2.1 and often yields better rates, but requires almost sure bounds on $|W - W'|$.

(3) It has been shown [33] that Theorem 2.1 still holds without exchangeability, assuming instead that W and W' are equally distributed. However, it is unclear how useful this observation is in practice, as the exchangeability plays a critical role in defining the pair (W, W') and in computing the error from Theorem 2.1.

It has been casually noted that the error term from Theorem 2.1 should be small when a is small, or equivalently when W and W' are 'close'. One line of thought [30] that supports this idea is that if (W, W') is bivariate standard normal with covariance $(1-a)$, then $\mathrm{Var}(\mathbb{E}[(W'-W)^2\,|\,W]) = a^4\,\mathrm{Var}(W^2)$. This equation loosely implies that smaller values of a should yield better bounds if (W, W') are nearly bivariate normal. Another argument for the idea that the error term should be small when W and W' are close is that the bound from Theorem 2.1 follows from a Taylor series approximation of W about W'. A simple heuristic to illustrate this argument can be found in [29] and [31]. However, the Taylor approximation takes place exclusively in the numerator of the error term and so this argument does not take into account the denominator which also decreases when W and W' are close (the same observations can be made directly from the error term). In accordance with these remarks, we will take the value a generated by the exchangeable pair (W, W') of Theorem 2.1 to be a rough quantitative measure of the 'step size' as referred to in the Introduction.

Finally, we remark here that the families of chains from Sections 4 and 5 have a similar form. Both of these families of chains are canonically engineered to satisfy the linearity condition $\mathbb{E}[W'\,|\,W] = (1-a)W$. To elaborate, if, as described in the Introduction, $(W, W') = (W(X_0), W(X_1))$ (that is, $\{X_0, X_1, \ldots\}$ is a stationary Markov chain reversible with respect to some distribution π on a state space Ω and W is some function from Ω into $\mathbb{R}$), then we have

$$\mathbb{E}[W'\,|\,X_0 = i] = \sum_{j\in\Omega}\mathbb{P}(X_1 = j\,|\,X_0 = i)W(j) = (P\mathbf{W})_i,$$

where P is the transition matrix of the Markov chain and $\mathbf{W}$ is the vector $W_i = W(i)$. Roughly, the linearity condition implies

$$P\mathbf{W} = (1-a)\mathbf{W},$$

so that $\mathbf{W}$ must be an eigenvector of the transition matrix P.

https://doi.org/10.1017/S096354830999023X Published online by Cambridge University Press

In both of Sections 4 and 5, the random variable W under study is a member of an orthogonal (in the L^2 sense with respect to the measure π under study) family of functions on the state space. From this point, it is not difficult to canonically define a matrix indexed by $\Omega \times \Omega$ with row sums equal to one, which has the orthogonal family as eigenvectors and satisfies the detailed balance equations for the measure π (although the entries are not guaranteed to be positive). We omit a more detailed abstract formulation, but the main ideas can be found in [35].

3. First example

In this section, we examine the most basic example of Stein's method for normal approximation. It is well known that the binomial distribution with parameters n and p is approximately normal for large n. Stein's method can be used to obtain an error term in this approximation. The set-up is as follows. Let $(X_1, X_2, \ldots, X_n)$ be a random vector where each X_i is independent and equal to 1 with probability p $(0 < p < 1)$ and 0 with probability $(1 - p)$. Then take $X = \sum_{i=1}^{n} X_i$. In order to clearly illustrate the typical way we will analyse the rest of the examples, we will take $p = 1/2$ in this section.

The first thing we need to study the error term from Theorem 2.1 is the family of Markov chains that induce the exchangeable pair. Given a configuration of an n-dimensional 0–1 vector, the next step in the chain follows the rule of changing any fixed i coordinates with probability $b_i / \binom{n}{i}$, so that $\sum_{i=0}^{n} b_i = 1$ (b_i is the probability of moving to a configuration Hamming distance i away). Since the probability of going from a configuration x to a configuration y is the same as going from y to x, the random vector $(X_1, \ldots, X_n)$ with each coordinate an independent Bernoulli random variable with parameter $p = 1/2$ is clearly reversible with respect to these chains. Setting $X = \sum_{i=1}^{n} X_i$ and $X' = \sum_{i=1}^{n} X_i'$, where $(X_1', \ldots, X_n')$ is a step in the chain described above, induces an exchangeable pair. In order to apply Theorem 2.1, the random variable must have mean 0 and variance 1. Set

$$W = \frac{X - \mathbb{E}(X)}{\sqrt{\mathrm{Var}(X)}} = \frac{X - np}{\sqrt{np(1-p)}},$$

and define W' as W but with X' in place of X. The final hypothesis from Theorem 2.1 is the following lemma.

Lemma 3.1.

$$\mathbb{E}(W' \,|\, W) = \left(1 - 2\sum_{i=1}^{n} \left(\frac{i}{n}\right) b_i\right) W.$$

Proof. Let $X = \sum_{i=1}^{n} X_i$ and $X' = \sum_{i=1}^{n} X_i'$ as defined above. Then,

$$\begin{aligned}
\mathbb{E}(X' - X \,|\, X) &= \sum_{i=1}^{n} \mathbb{E}(X_i' - X_i \,|\, X) \\
&= \sum_{i:X_i=1} (-1)\mathbb{P}(X_i' = 0 \,|\, X_i = 1) + \sum_{i:X_i=0} \mathbb{P}(X_i' = 1 \,|\, X_i = 0) \\
&= (n - 2X) \sum_{i=1}^{n} \left(\frac{i}{n}\right) b_i.
\end{aligned}$$

https://doi.org/10.1017/S096354830999023X Published online by Cambridge University Press

Substituting $X = \sqrt{n/4}W + n/2$ and $X' = \sqrt{n/4}W' + n/2$ yields

$$\mathbb{E}(W' - W \mid W) = -2\sum_{i=1}^{n}\left(\frac{i}{n}\right)b_i W,$$

which is the lemma. □

Now Theorem 2.1 can be applied with $a = 2\sum_{i=1}^{n}(i/n)b_i$. In order to apply the theorem, we still need to compute the quantities $\mathrm{Var}(\mathbb{E}[(W' - W)^2 \mid W])$ and $\mathbb{E}(W' - W)^4$.

Lemma 3.2.

$$\mathrm{Var}(\mathbb{E}[(W' - W)^2 \mid W]) = 16B^2\,\mathrm{Var}(W^2),$$

where $B = \sum_{i=2}^{n}\frac{i(i-1)}{n(n-1)}b_i$.

Proof.

$$\begin{aligned}\mathbb{E}[(X' - X)^2 \mid X] &= \sum_{i=1}^{n}\mathbb{E}[(X_i' - X_i)^2 \mid X] + \sum_{\substack{i:X_i=1\\ j\neq i:X_j=0}}\mathbb{E}[(X_i' - X_i)(X_j' - X_j) \mid X]\\ &\quad + \sum_{\substack{i:X_i=1\\ j\neq i:X_j=1}}\mathbb{E}[(X_i' - X_i)(X_j' - X_j) \mid X]\\ &\quad + \sum_{\substack{i:X_i=0\\ j\neq i:X_j=0}}\mathbb{E}[(X_i' - X_i)(X_j' - X_j) \mid X]\\ &= nA + \big(X(X-1) + (n-X)(n-X-1) - 2X(n-X)\big)B,\end{aligned}$$

where $A = \sum_{i=1}^{n}(i/n)b_i$. Substituting $X = \sqrt{n/4}W + n/2$ and $X' = \sqrt{n/4}W' + n/2$ into the equation and solving appropriately yields $\mathbb{E}[(W' - W)^2 \mid W] = 4BW^2 + C$, where C is some constant. Taking variances proves the lemma. □

Lemma 3.3.

$$\mathbb{E}(W' - W)^4 = \frac{16}{n}[A + 3B(n-1) + (2 - 3n + n\mathbb{E}(W^4))D],$$

where $A = \sum_{i=1}^{n}(i/n)b_i$, *B is defined as in Lemma 3.2, and D is a constant.*

https://doi.org/10.1017/S096354830999023X Published online by Cambridge University Press

Proof. Let $Y_i = X_i' - X_i$. Then,

$$\begin{aligned}\mathbb{E}\big[(X'-X)^4 \,|\, X\big] &= \sum_{i=1}^{n} \mathbb{E}\big[Y_i^4 \,|\, X\big] + 4\sum_{\substack{i\\ j\neq i}} \mathbb{E}\big[Y_i^3 Y_j \,|\, X\big] \\ &\quad + 3\sum_{\substack{i\\ j\neq i}} \mathbb{E}\big[Y_i^2 Y_j^2 \,|\, X\big] + 6\sum_{\substack{i\\ j\neq i\\ l\neq i,j}} \mathbb{E}\big[Y_i^2 Y_j Y_l \,|\, X\big] \\ &\quad + \sum_{\substack{i\\ j\neq i\\ l\neq i,j\\ m\neq i,j,l}} \mathbb{E}\big[Y_i Y_j Y_l Y_m \,|\, X\big].\end{aligned}$$

Counting the number of each type and computing the expected values in a manner similar to the proof of Lemma 3.2 yields

$$\begin{aligned}\mathbb{E}[(X'-X)^4 \,|\, X] &= nA + (4B + 6(n-2)C)(n(n-1) - 4X(n-X)) \\ &\quad + [n(n-1)(n-2)(n-3) - 8((n-X)X(X-1)(X-2) \\ &\quad + X(n-X)(n-X-1)(n-X-2))]D + 3n(n-1)B.\end{aligned}$$

Here C and D are constants that vanish in the final expression, but C is the probability of any fixed three coordinates changing, and D is the same probability but with four coordinates. Substituting in W and W' as in the previous lemmas and taking the expected value of $\mathbb{E}\big[(W'-W)^4 \,|\, W\big]$ implies the lemma. □

Lemma 3.4. $\mathbb{E}[W^4] = 3 - 2/n$.

Proof. The moment generating function for W is

$$\phi(t) = \left(\frac{1}{2}\left[\exp\left(\left(\frac{1}{n}\right)^{1/2} t\right)\right] + \frac{1}{2}\left[\exp\left(-\left(\frac{1}{n}\right)^{1/2} t\right)\right]\right)^n,$$

and $\mathbb{E}[W^4] = \phi^{(4)}(0)$. □

Now that we have all the formulas needed to apply Theorem 2.1, we can prove the main result of this section.

Theorem 3.5. *Using the family of reversible Markov chains described previously, the error term given by Theorem 2.1 is minimal for* $b_0 + b_1 = 1$ *(and* $b_0 \neq 1$*). In this case the bound is* $[8/\pi n]^{1/4}$.

Proof. Because a is positive, it is sufficient to verify that $\mathrm{Var}(\mathbb{E}[(W'-W)^2 \,|\, W])/a^2$ and $\mathbb{E}(W'-W)^4/a$ are minimal for $b_0 + b_1 = 1$. First we examine $\mathrm{Var}(\mathbb{E}[(W'-W)^2 \,|\, W])/a^2$. By Lemmas 3.1, 3.2, and 3.4 and the fact that $a = 2A$, this term is equal to

$$8\left(\frac{n-1}{n}\right)\left(\frac{B^2}{A^2}\right),$$

which is non-negative and equal to zero when $b_0 + b_1 = 1$.

https://doi.org/10.1017/S096354830999023X Published online by Cambridge University Press

For the second term we use the second formulation in Theorem 2.1 and find the minimum of

$$\frac{\mathbb{E}(W' - W)^4}{a}. \tag{3.1}$$

By Lemmas 3.1, 3.3, and 3.4, (3.1) is equal to

$$\frac{8}{n}\left(1 + 3(n-1)\frac{B}{A}\right),$$

which is also minimized when $b_0 + b_1 = 1$, and is equal to $8/n$ in this case. □

Remarks. **(1)** It is important to notice that the error term from Theorem 2.1 depends on the Markov chain only through the term B/A, where

$$B = \sum_{i=1}^{n}\left(\frac{i(i-1)}{n(n-1)}\right)b_i \quad \text{and} \quad A = \sum_{i=1}^{n}\left(\frac{i}{n}\right)b_i.$$

We will be using this fact later in Section 6.

(2) The case $b_0 + b_1 = 1$ corresponds to the Markov chain that holds with probability b_0 and changes one coordinate chosen uniformly at random with probability b_1. Therefore, this chain has the smallest maximum step size over all the chains in the family under study. Also, as mentioned in the previous section, a quantitative measure of step size is the associated value of a from Theorem 2.1, which is equal to $2b_1/n$ in this optimal case. Restricting to the case where $b_0 = 0$, the chain yielding the best bound ($b_1 = 1$) has the smallest value of a. This restriction is not artificial; in lieu of the previous remark, the chain generated with $\mathbb{P}(T = t) = b'_t$, where $b'_0 = 0$, and $b'_i = b_i/(1 - b_0)$ for $i \neq 0$ has the same error term from Theorem 2.1 as the chain with $\mathbb{P}(T = t) = b_t$. In other words, manipulating the holding probability changes the parameter a, but can not yield bounds better than those obtained by choosing $b_0 = 0$.

4. Plancherel distribution of the Hamming scheme (or the binomial distribution in general)

In this section we examine the uniform distribution on the eigenvalues of the adjacency matrix for the Hamming graph. The Hamming graph $H(n, q)$ has vertex set X equal to n-tuples of $\{1, 2, \ldots, q\}$ (thus $|X| = q^n$) with an edge between two vertices if they differ in exactly one coordinate.

The following information about the adjacency matrix of the Hamming scheme can be found in [5] in the more general setting of association schemes. Let $v_i = (q-1)^i\binom{n}{i}$, the number of vertices that differ from a fixed vertex by i coordinates. The eigenvalues of $H(n, q)$ are $K_1(i) = n(q-1) - qi$ with multiplicity v_i for $i = 0, 1, 2, \ldots, n$. Choose i with probability $v_i/|X|$ and designate this the Plancherel distribution of the Hamming scheme. Let $W(i) = K_1(i)/\sqrt{v_1}$ be a random variable with unit variance. In order to define the family of Markov chains that induce the exchangeable pairs, we must define the q-Krawtchouk polynomials:

$$K_j(i) = \sum_{l=0}^{j}(-1)^l(q-1)^{j-l}\binom{i}{l}\binom{n-i}{j-l}.$$

https://doi.org/10.1017/S096354830999023X Published online by Cambridge University Press

Here and in what follows we freely use the convention $\binom{m}{r} = 0$ for $r > m$ or $r < 0$. Following Fulman [22], define

$$L_T(i,j) = \frac{v_j}{|X|} \sum_{r=0}^{n} \frac{K_r(i)K_r(T)K_r(j)}{v_r^2}. \tag{4.1}$$

For a given T in $\{1,\dots,n\}$, define a Markov chain on $\{0,\dots,n\}$ by the transition probability of moving from i to j as $L_T(i,j)$. Then $L_T(i,j)$ is a Markov chain on $\{0,\dots,n\}$ reversible with respect to the Plancherel distribution above ($\mathbb{P}(i) = v_i/|X|$). Following the usual set-up, choose i from the Plancherel distribution and then j with probability $L_T(i,j)$, and set the exchangeable pair $(W, W') = (W(i), W(j))$. Fulman [22] uses Theorem 2.1 with the exchangeable pair induced by the chain L_1 (defined by (4.1) with $T = 1$) to obtain a bound on the difference between the normal distribution and the Plancherel measure. We will show that over a large family of Markov chains, L_1 is the most local chain and the error term obtained in Theorem 2.1 using L_1 (as was done in [22]) is optimal over this family.

Another (somewhat more motivating) way of viewing the Hamming scheme is given in the following easily verified proposition.

Proposition 4.1 ([22]). *For W defined as above, W is equal in distribution to a binomial distribution with parameters n and $p = 1/q$ normalized to have mean 0 and variance 1.*

Thus the Plancherel distribution of the Hamming scheme can be defined as a binomial distribution. We will redefine the Markov chain $L_T(i,j)$ in terms of this characterization, but first we list some well-known properties of Krawtchouk polynomials found in [28].

Lemma 4.2. *For $j, l \in \{0,\dots,n\}$,*

$$\sum_{i=0}^{n} \frac{K_i(j)K_i(l)}{v_i} = \frac{|X|}{v_j}\delta_{j,l}.$$

Lemma 4.3. *For $j, l \in \{0,\dots,n\}$,*

$$\sum_{i=0}^{n} v_i K_j(i)K_l(i) = |X| v_j \delta_{j,l}.$$

Lemma 4.4. *For $j \in \{0,\dots,n\}$,*

$$(j+1)K_{j+1}(i) = ((n-j)(q-1) + j - qi)K_j(i) - (q-1)(n-j+1)K_{j-1}(i).$$

Lemma 4.5. *For $i, j \in \{0,\dots,n\}$,*

$$K_j(i) = (q-1)^{j-i}\frac{\binom{n}{j}}{\binom{n}{i}}K_i(j).$$

https://doi.org/10.1017/S096354830999023X Published online by Cambridge University Press

Finally, we need one more tool that equates the product of two Krawtchouk polynomials with a linear combination of Krawtchouk polynomials.

Lemma 4.6. *For $i, j, r \in \{0, \dots, n\}$,*

$$K_i(r)K_j(r) = \sum_{l=-j}^{j} A_{j,i+l}(i)K_{i+l}(r),$$

where

$$A_{j,i+l}(i) = \frac{\binom{n}{i}}{\binom{n}{i+l}} \sum_{k=0}^{j} \binom{j-k}{k-l}\binom{n-i}{k}\binom{i}{j-k}\frac{(q-2)^{j-2k+l}}{(q-1)^{l-k}}.$$

Proof. The first thing to note is that the Krawtchouk polynomials $K_i(r)$ for i non-negative integers form a basis for all polynomials in r, so that such a decomposition exists. Now, fix i and write $A_{j,i+l}(i) = A_{j,i+l}$. For $j = 0$, $A_{0,i} = 1$ and $A_{0,i+l} = 0$ for $l \neq 0$, which agrees with the lemma.

Also, since $K_1(r) = (q-1)(n-r) - r = (q-1)n - qr$, Lemma 4.4 implies

$$K_i(r)K_1(r) = i(q-2)K_i(r) + (i+1)K_{i+1}(r) + (q-1)(n-i+1)K_{i-1}(r), \tag{4.2}$$

which is also consistent with the lemma (the equality (4.2) was shown in [22]).

For $j \geqslant 1$, we use Lemma 4.4 and strong induction to obtain

$$\begin{aligned}
(j+1)K_{j+1}(r)K_i(r) &= ((n-j)(q-1) + j - qr)K_j(r)K_i(r) \\
&\quad - (q-1)(n-j+1)K_{j-1}(r)K_i(r) \\
&= \big((n-j)(q-1) + j - qr\big)\sum_{l=-j}^{j} A_{j,i+l}K_{i+l}(r) \\
&\quad - (q-1)(n-j+1)\sum_{l=-(j-1)}^{j-1} A_{j-1,i+l}K_{i+l}(r) \\
&= \big((n-j)(q-1) + j\big)\sum_{l=-j}^{j} A_{j,i+l}K_{i+l}(r) \\
&\quad - (q-1)(n-j+1)\sum_{l=-(j-1)}^{j-1} A_{j-1,i+l}K_{i+l}(r) \\
&\quad + \sum_{l=-j}^{j}[(l+i+1)K_{l+i+1}(r) + (q-1)(n-i-l+1)K_{i+l-1}(r) \\
&\quad - ((n-i-l)(q-1) + l + i)K_{i+l}(r)]A_{j,i+l},
\end{aligned} \tag{4.3}$$

where the final equality is by Lemma 4.4. For each l, the coefficient of $K_{i+l}(r)$ in (4.3) is

$$(i+l-j)(q-2)A_{j,i+l} + (q-1)(n-i-l)A_{j,i+(l+1)} \tag{4.4}$$

$$+ (i+l)A_{j,i+(l-1)} - (q-1)(n-j+1)A_{j-1,i+l}. \tag{4.5}$$

https://doi.org/10.1017/S096354830999023X Published online by Cambridge University Press

The lemma will follow if the expression above is equal to $(j+1)A_{j+1,i+l}$. To see this fact, re-index the sums in the definition of $A_{j,i}$ in (4.5) to begin at one, and equate summands. □

The fact from the previous lemma that the coefficients in the linear expansion are positive for $q \geqslant 2$ has been shown without explicit computation in [18] and restated in the monograph [3]. We use the positivity of the coefficients in the next theorem, which shows that they can be used to define a probability distribution.

Theorem 4.7. *For $q \geqslant 2$, the Markov chain $L_T(i,j)$ defined on $\{0,\dots,n\}$ has the same transition probabilities as the following chain. Given a* 0–1 *n-tuple with* i *ones, choose* T *coordinates at random. Replace every zero coordinate chosen to a one and for each one coordinate chosen, replace it with a zero with probability* $1/(q-1)$ *and let the coordinate remain as a one with probability* $(q-2)/(q-1)$. *The probability of going from* i *ones to* j *ones is* $L_T(i,j)$.

Proof. By summing over k equal to the number of zeros in the T coordinates chosen, the probability of going from i to $i+l$ in the chain described is

$$\sum_{k=0}^{T} \binom{T-k}{k-l} \left(\frac{(q-2)^{T-2k+l}}{(q-1)^{T-k}} \right) \left(\frac{\binom{n-i}{k}\binom{i}{T-k}}{\binom{n}{T}} \right).$$

Also,

$$\begin{aligned} L_T(i,j) &= \frac{v_j}{|X|} \sum_{r=0}^{n} \frac{K_r(i)K_r(T)K_r(j)}{v_r^2} \\ &= \frac{v_j}{|X|\binom{n}{i}\binom{n}{j}\binom{n}{T}(q-1)^{i+j+T}} \sum_{r=0}^{n} (q-1)^r \binom{n}{r} K_i(r)K_T(r)K_j(r) \\ &= \frac{v_j}{|X|\binom{n}{i}\binom{n}{j}\binom{n}{T}(q-1)^{i+j+T}} \sum_{r=0}^{n} v_r K_j(r) \sum_{l=-T}^{T} A_{T,i+l}K_{i+l}(r) \\ &= \frac{v_j}{|X|\binom{n}{i}\binom{n}{j}\binom{n}{T}(q-1)^{i+j+T}} \sum_{l=-T}^{T} A_{T,i+l} \sum_{r=0}^{n} v_r K_{i+l}(r)K_j(r). \end{aligned}$$

The first equality is by Lemma 4.5 and the second is by Lemma 4.6.

Applying Lemma 4.3 implies

$$L_T(i,i+l) = \frac{v_{i+l}}{\binom{n}{i}\binom{n}{T}(q-1)^{i+T}} A_{T,i+l}.$$

By Lemma 4.6, this equals the desired quantity. □

According to Proposition 4.1, the restriction $q \geqslant 2$ corresponds to a binomial distribution with parameters n and p with $1/2 \leqslant p < 1$. However, after normalizing to have mean equal to zero and unit variance, a binomial random variable with parameters n and $(1-p)$ is the negative of a binomial random variable with parameters n and p. Therefore, because the normal distribution is symmetric about zero, the following analysis can be applied to any binomial random variable.

https://doi.org/10.1017/S096354830999023X Published online by Cambridge University Press

The chains defined by (4.1) have now been fully described in terms of the binomial distribution. Now we will start to examine the error term from Theorem 2.1. The next lemma shows that the quantity a from Theorem 2.1 is equal to $qT/n(q-1)$ so that a is increasing as a function of T. Also, Theorem 4.7 implies that the maximum step size of L_T is T; smaller values of T make W and W' 'closer' in both senses described in Section 2.

Lemma 4.8 ([22]). $\mathbb{E}[W' \mid W] = \big(K_1(T)/v_1\big)W$.

Proof.

$$\begin{aligned}\mathbb{E}[W' \mid i] &= \frac{1}{\sqrt{v_1}} \sum_{j=0}^{n} L_T(i,j) K_1(j) \\ &= \frac{1}{\sqrt{v_1}\,|X|} \sum_{r=0}^{n} \frac{K_r(T)K_r(i)}{v_r^2} \sum_{j=0}^{n} v_j K_r(j) K_1(j) \\ &= \left(\frac{K_1(T)}{v_1}\right) W(i).\end{aligned}$$

The first equality is by definition and the second uses Lemma 4.3 directly. Since conditioning on i only depends on i through W, the lemma is proved. □

For this example, the error term (in the more general setting of association schemes) has already been computed using Theorem 2.1 in [22], so we only state the result we need. First we must define a function p_2 on $\{0,\dots,n\}$ using the following description. Start from a fixed X_0 in X and choose a coordinate uniformly at random. Replace the chosen coordinate by one of the remaining $q-1$ options different from the original value uniformly at random to obtain X_1. Perform the same operation on X_1 to obtain X_2, and define $p_2(j)$ to be the probability that X_2 has j coordinates different from X_0. Thus $p_2(0) = 1/n(q-1)$, $p_2(1) = (q-2)/n(q-1)$, $p_2(2) = (n-1)/n$, and $p_2(j) = 0$ for $j \geqslant 3$.

Theorem 4.9 ([22]). *Let W and a be defined as above and let T be fixed in $\{1,\dots,n\}$. Then, for all real x_0,*

$$\begin{aligned}&\left|\mathbb{P}(W < x_0) - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x_0} e^{\frac{-x^2}{2}}\,dx\right| \\ &\quad\leqslant \frac{v_1}{a}\sqrt{\sum_{j=1}^{n} \frac{p_2(j)^2}{v_j}\left(\frac{K_j(T)}{v_j} + 1 - \frac{2K_1(T)}{v_1}\right)^2} \\ &\qquad + \frac{\sqrt{v_1}}{\pi^{1/4}}\left[\sum_{j=0}^{n}\left(8 - \frac{6}{a}\left(1 - \frac{K_j(T)}{v_j}\right)\right)\frac{p_2(j)^2}{v_j}\right]^{1/4}.\end{aligned}$$

Armed with this theorem, we can examine how varying the value of T affects the error term. However, rather than examining the error for a fixed value of T, we define T to be a random variable on $\{0,\dots,n\}$ with $\mathbb{P}(T = t) = b_t$ (and $b_0 \neq 1$). Note that this modification does not affect the stationary distribution of the chain.

https://doi.org/10.1017/S096354830999023X Published online by Cambridge University Press

There are two reasons for using a random variable instead of a fixed value of T. The first reason is that using a random variable yields a larger family of Markov chains. The second reason is that when $q = 2$, the Plancherel distribution is a binomial distribution with parameters n and $p = 1/2$. In this case, Theorem 4.7 implies that the chain $L_t(i, j)$ chooses t coordinates with probability b_t and changes them all. That is, $L_t(i, j)$ follows the rule of moving a (Hamming) distance t away with probability b_t, and we recover the chain from Section 3.

We will show that the random variable T that minimizes the error term from Theorem 2.1 is induced by the Markov chain used in [22], which has $b_1 = 1$ (or alternatively $b_0 + b_1 = 1$ and $b_0 \neq 1$) and all other $b_t = 0$. The computation of the error term with T as a random variable closely follows the proof of Theorem 4.9 from [22]. To apply Theorem 2.1, we need to compute the terms in the bound on the error. It will be helpful to define (W_t, W_t') to be the exchangeable pair defined in the usual way from the Markov chain $L_t(i, j)$. Note that $W_t = W$, since using a different Markov chain in the family does not alter the stationary distribution.

The next two lemmas will generate all the terms needed to apply Theorem 2.1.

Lemma 4.10 ([22]).

$$\mathbb{E}[(W_t' - W_t)^2 \,|\, i] = v_1 \sum_{r=0}^{n} \frac{K_r(i)K_r(t)}{v_r^2} p_2(r) + \left(1 - \frac{2K_1(t)}{v_1}\right) W_t^2.$$

$$\operatorname{Var}(\mathbb{E}[(W_t' - W_t)^2 \,|\, i]) = v_1^2 \sum_{r=1}^{n} \frac{p_2(r)^2}{v_r} \left(1 + \left(\frac{K_r(t)}{v_r} - \frac{2K_1(t)}{v_1}\right)\right)^2.$$

$$\mathbb{E}[(W_t' - W_t)^4] = v_1^2 \left[\sum_{r=0}^{n} \left(8\left(1 - \frac{K_1(t)}{v_1}\right) - 6\left(1 - \frac{K_r(t)}{v_r}\right)\right) \frac{p_2(r)^2}{v_r}\right].$$

Lemma 4.11.

$$\mathbb{E}[W' \,|\, W] = \left(\sum_{t=0}^{n} b_t \frac{K_1(t)}{v_1}\right) W.$$

$$\mathbb{E}[(W' - W)^2 \,|\, i] = v_1 \sum_{r=0}^{n} \frac{K_r(i)\left(\sum_{t=0}^{n} b_t K_r(t)\right)}{v_r^2} p_2(r) + \left(1 - \frac{2\left(\sum_{t=0}^{n} b_t K_1(t)\right)}{v_1}\right) W^2.$$

$$\operatorname{Var}(\mathbb{E}[(W' - W)^2 \,|\, i]) = v_1^2 \sum_{r=1}^{n} \frac{p_2(r)^2}{v_r} \left(1 + \sum_{t=0}^{n} b_t \left(\frac{K_r(t)}{v_r} - \frac{2K_1(t)}{v_1}\right)\right)^2.$$

$$\mathbb{E}[(W' - W)^4] = v_1^2 \left[\sum_{r=0}^{n} \left(8\left(1 - \frac{\sum_{t=0}^{n} b_t K_1(t)}{v_1}\right) - 6\left(1 - \frac{\sum_{t=0}^{n} b_t K_r(t)}{v_r}\right)\right) \frac{p_2(r)^2}{v_r}\right].$$

Proof. We prove only the first equality; the proofs of the remaining are similar. By summing over t equal to the value of T chosen, we have

$$\mathbb{E}[W' \,|\, i] = \frac{1}{\sqrt{v_1}} \sum_{t=0}^{n} b_t \sum_{j=0}^{n} L_t(i, j) K_1(j).$$

The equality now follows from the proof of Lemma 4.8. □

https://doi.org/10.1017/S096354830999023X Published online by Cambridge University Press

Applying Lemma 4.11 to Theorem 2.1 proves the following theorem.

Theorem 4.12. *If W and $\{b_t\}_{t=0}^n$ are defined as above,*

$$a = 1 - \left(\sum_{t=0}^{n} b_t \frac{K_1(t)}{v_1}\right),$$

and all other variables are as in Theorem 4.9, then for all real x_0,

$$\left|\mathbb{P}(W < x_0) - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x_0} e^{\frac{-x^2}{2}} dx\right|$$
$$\leqslant \frac{v_1}{a} \sqrt{\sum_{j=1}^{n} \frac{p_2(j)^2}{v_j} \left(\frac{\sum_{t=0}^n b_t K_j(t)}{v_j} + 1 - \frac{2\left(\sum_{t=0}^n b_t K_1(t)\right)}{v_1}\right)^2}$$
$$+ \frac{\sqrt{v_1}}{\pi^{1/4}} \left[\sum_{j=0}^{n} \left(8 - \frac{6}{a}\left(1 - \frac{\sum_{t=0}^n b_t K_j(t)}{v_j}\right)\right) \frac{p_2(j)^2}{v_j}\right]^{1/4}.$$

We can now analyse the error term of Theorem 4.12 over $L_T(i, j)$, the family of Markov chains previously defined.

Theorem 4.13. *The Markov chain $L_T(i, j)$ that minimizes the error term from Theorem 4.12 is the chain with $\mathbb{P}(T = 1) = b_1 = 1$.*

Proof. The right-hand side of the inequality of Theorem 4.12 can be rewritten as

$$v_1 \sqrt{\sum_{j=1}^{n} \frac{p_2(j)^2}{v_j} \left(\frac{\left(\sum_{t=0}^n b_t \frac{K_j(t)}{v_j}\right) - 1}{a} + 2\right)^2} \tag{4.6}$$
$$+ \frac{\sqrt{v_1}}{\pi^{1/4}} \left[\sum_{j=0}^{n} \left(8 + 6\left(\frac{\left(\sum_{t=0}^n b_t \frac{K_j(t)}{v_j}\right) - 1}{a}\right)\right) \frac{p_2(j)^2}{v_j}\right]^{1/4}. \tag{4.7}$$

We will show that

$$\frac{\left(\sum_{t=0}^n b_t \frac{K_j(t)}{v_j}\right) - 1}{a} \tag{4.8}$$

is minimal for each j under some set $\mathcal{B} = \{b_t\}_{t=0}^n$, which implies that (4.7) is minimal under $\mathcal{B}$ (notice that $v_j \geqslant 0$ for all j). We will then show that the minimum value of (4.8) is no less than -2 for each j, so that (4.6) is also minimal under $\mathcal{B}$.

Since $p_2(j) = 0$ for $j \geqslant 3$, we only consider $j = 0, 1, 2$. For $j = 0$, $K_0(t) = v_0 = 1$, so that (4.8) does not depend on t. For $j = 1$, the numerator of (4.8) is equal to $-a$, so (4.8) is independent of t. For $j = 2$, a straightforward calculation (using $\sum_{t=0}^n b_t = 1$) yields

$$\frac{\left(\sum_{t=0}^n b_t \frac{K_2(t)}{v_2}\right) - 1}{a} + 2 = \left(\frac{q}{q-1}\right) \frac{\sum_{t=2}^n b_t \frac{t(t-1)}{n(n-1)}}{\sum_{t=1}^n b_t \left(\frac{t}{n}\right)}. \tag{4.9}$$

https://doi.org/10.1017/S096354830999023X Published online by Cambridge University Press

Since the summand in the numerator is 0 when $t = 1$ and positive for all other t values (where b_t is positive), the final term in (4.9) is minimal for $b_1 + b_0 = 1$ and $b_t = 0$ for $t > 1$. □

Remarks. **(1)** As in Section 3, Theorem 4.7 implies that the case $b_0 + b_1 = 1$ corresponds to the Markov chain L_T having step size at most one and with associated value of a from Theorem 2.1 equal to $\frac{qb_1}{(q-1)n}$. Among chains with $b_0 = 0$, the chain yielding the best bound has the smallest maximum step size and the smallest value of a. We refer to the remarks following Theorem 3.5 on this restriction.

(2) It is interesting to note that the size of the bound on the error in this section and in Section 3 both depend on the underlying chain through the same term (B/A from Section 3). This is obvious from the case $q = 2$, since this case is the same in both sections, but it is not clear why this should carry over to other values of q.

5. Plancherel distribution on a group

In this section we examine the Plancherel measure of the random walk generated by the conjugacy class of transpositions on S_n (the symmetric group on n symbols). First we describe the set-up for any group in order to state the theorems we will use in the utmost generality. Let G be a group and let C be a non-trivial conjugacy class of G such that $C = C^{-1}$. Define the random walk on G generated by C, a Markov chain with state space G, as follows. Given g in G, the next step in the chain is gh, where h is chosen uniformly at random from C. Now, denote the set of irreducible representations of G by $\mathrm{Irr}(G)$. From [17], for each character χ^λ of λ in $\mathrm{Irr}(G)$, there is an eigenvalue of the random walk on G generated by C given by $\chi^\lambda(C)/\dim(\lambda)$ (termed a character ratio) occurring with multiplicity $\dim(\lambda)^2$. A well-known fact from representation theory is (see, *e.g.*, [34])

$$\sum_{\lambda \in \mathrm{Irr}(G)} \dim(\lambda)^2 = |G|,$$

so we define the Plancherel measure of G to choose λ in $\mathrm{Irr}(G)$ with probability $\dim(\lambda)^2/|G|$. Then the unit variance random variable $W(\lambda) = |C|^{1/2}\chi^\lambda(C)/\dim(\lambda)$ is essentially an eigenvalue of the random walk above chosen uniformly at random.

Fulman [22] defines a Markov chain reversible with respect to the Plancherel measure of a group G with the probability of transitioning from λ to ρ given by

$$L_\tau(\lambda, \rho) = \frac{\dim(\rho)}{\dim(\lambda)\dim(\tau)} \frac{1}{|G|} \sum_{g \in G} \chi^\lambda(g)\chi^\tau(g)\overline{\chi^\rho(g)}.$$

Here τ is some representation of G. Following the usual set-up, define an exchangeable pair (W, W') by $W = W(\lambda)$, where λ is chosen from the Plancherel measure of G, and define $W' = W(\rho)$, where ρ is given by one step in the above chain. For the sake of continuity, we will postpone discussing properties of this Markov chain family until later in this section.

Using facts from representation theory, Fulman [22] proves the following lemma, which allows for the application of Theorem 2.1.

Lemma 5.1 ([22]). $\mathbb{E}(W' \mid W) = \big(\chi^\tau(C)/\dim(\tau)\big)W$.

https://doi.org/10.1017/S096354830999023X Published online by Cambridge University Press

Theorem 5.2 ([22]). *Let C be a conjugacy class of a finite group G such that $C = C^{-1}$ and fix a non-trivial irreducible representation τ of G whose character is real-valued. Let λ be a random irreducible representation chosen from the Plancherel measure of G. Let $W = |C|^{1/2}\chi^{\lambda}(C)/\dim(\lambda)$. Then, for all real x_0,*

$$\left|\mathbb{P}(W \leqslant x_0) - \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x_0} e^{-\frac{x^2}{2}}\,dx\right| \leqslant |C|\sqrt{\sum_{K\neq \mathrm{id}} \frac{p_2(K)^2}{|K|}\left(\frac{\frac{\chi^{\tau}(K)}{\dim(\tau)}-1}{a_\tau}+2\right)^2} + \left[\frac{|C|^2}{\pi}\sum_{K}\frac{p_2(K)^2}{|K|}\left(8+6\left(\frac{\frac{\chi^{\tau}(K)}{\dim(\tau)}-1}{a_\tau}\right)\right)\right]^{1/4}.$$

Here the sums are over conjugacy classes K of G, $a_\tau = 1 - \chi^{\tau}(C)/\dim(\tau)$, and $p_2(K)$ is the probability that the random walk on G generated by C started at the identity is at the conjugacy class K after two steps.

After proving this theorem, Fulman applies it with the choice of the irreducible representation τ corresponding to the partition $(n-1,1)$ (more on this notation later) to obtain a central limit theorem with an error term in the case where $G = S_n$ and C is the conjugacy class of i-cycles. We will show that this choice of τ minimizes the error term from Theorem 5.2 in the special case where C is the conjugacy class of transpositions (2-cycles).

Before going further we will explain the notation above and state some facts about the irreducible representations of S_n found in [34]. For S_n, the irreducible representations are indexed by partitions of the integer n, where the partition (n) corresponds to the trivial representation. Another way to index the irreducible representations (which can be more useful for combinatorial reasons) is to associate to each partition a 'tableau', which is a left-justified array of equally sized, aligned boxes (we will typically abuse notation and refer to λ as both the partition and the diagram). For a partition $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_k)$ of n, where $\lambda_1 \geqslant \cdots \geqslant \lambda_k > 0$, the associated tableau has λ_1 boxes in the first row, λ_2 boxes in the second row, and so on. Now, numbering the boxes one through n left to right, top to bottom (with this indexing, this is technically the largest standard Young tableau), we make the following definition.

Definition. Let $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_k)$ be a partition of n. Define the content of box i (in the labelling above) of λ to be $c_\lambda(i) = \mathrm{column}_\lambda(i) - \mathrm{row}_\lambda(i)$.

In order to clarify these definitions, Figure 1 is the labelled tableau for $\lambda = (5,3,3,1)$, a partition of $n = 12$. Figure 2 is the same tableau, but the box labelled i now has the value of the content $c_\lambda(i)$.

Because we are specializing to the conjugacy class C of transpositions, and $p_2(K)$ is the probability of being in the conjugacy class K after two steps in the random walk on S_n generated by C starting at the identity element, $p_2(K) = 0$ for many conjugacy classes. The following lemma formulates $\chi^{\lambda}(K)/\dim(\lambda)$ in terms of the contents of λ for conjugacy classes K, where $p_2(K) \neq 0$ (the conjugacy classes that contribute to the error term in Theorem 5.2). The lemma is

https://doi.org/10.1017/S096354830999023X Published online by Cambridge University Press

1	2	3	4	5
6	7	8		
9	10	11		
12				

Figure 1. Labelled tableau for $\lambda = (5, 3, 3, 1)$.

0	1	2	3	4
-1	0	1		
-2	-1	0		
-3				

Figure 2. $\lambda = (5, 3, 3, 1)$ with box i labelled $c_\lambda(i)$.

proved in the form shown here using Murphy's elements in [16], but can also be found in terms of the λ_i in [26].

Lemma 5.3 ([16]). *Let $\chi^\lambda(j)$ and $\chi^\lambda(2,2)$ be the character of the irreducible representation λ at the conjugacy class of a j-cycle and two 2-cycles, respectively. Then*

$$\frac{\chi^\lambda(\mathrm{id})}{\dim(\lambda)} = 1,$$
$$\frac{\chi^\lambda(2)}{\dim(\lambda)} = \frac{2}{n(n-1)} \sum_{i=1}^{n} c_\lambda(i),$$
$$\frac{\chi^\lambda(3)}{\dim(\lambda)} = \frac{3}{n(n-1)(n-2)} \left(\sum_{i=1}^{n} c_\lambda(i)^2 - \binom{n}{2} \right),$$
$$\frac{\chi^\lambda(2,2)}{\dim(\lambda)} = \frac{1}{6\binom{n}{4}} \left(\left(\sum_{i=1}^{n} c_\lambda(i) \right)^2 - 3 \sum_{i=1}^{n} c_\lambda(i)^2 + n(n-1) \right).$$

For the example from Figures 1 and 2 where $\lambda = (5, 3, 3, 1)$, the sum of the contents is equal to four and n is equal to twelve. Thus Lemma 5.3 implies, for example, $\chi^\lambda(2)/\dim(\lambda) = 2/33$.

We pause here to discuss some relevant properties of the chain L_λ. First, we reiterate the remarks at the end of Section 2; the definition of the family of Markov chains stems from the orthogonality relations of irreducible characters (see [21] and the references therein). Second, because it is not known how to combinatorially describe L_λ in general, it is not obvious how to relate the choice of λ to step size. However, some representations do have nice descriptions and we can use these to gain some intuition. From [20], if ν is the defining representation on S_n, then given an irreducible representation μ, L_ν follows the rule of removing a corner box of μ uniformly at random (so the resulting diagram of $n-1$ boxes remains a tableau) and moving it to a uniformly chosen concave open position of the altered diagram (to obtain a new tableau of size n). For ρ the trivial representation, ν the defining representation, and $\tau := (n-1, 1)$, we

https://doi.org/10.1017/S096354830999023X Published online by Cambridge University Press

have [34] $\chi^{\nu} = \chi^{\tau} + \chi^{\rho}$, $\dim(\rho) = 1$, $\dim(\nu) = n$, and $\dim(\tau) = n-1$, so that

$$L_{\tau} = \frac{nL_{\nu} - L_{\rho}}{n-1}.$$

Now, using the orthogonality relations of characters and the fact that $\chi^{\rho} \equiv 1$, it is easy to see that the chain L_{ρ} holds with probability one. Thus, the previous remarks imply that from an irreducible representation μ, the chain L_{τ} moves at most one box of μ. Also, again using orthogonality relations, it follows that for any irreducible representations μ and λ, we have $L_{\lambda}(\rho, \mu) = \delta_{\lambda\mu}$. That is, from the trivial representation the chain L_{λ} moves to λ with probability one, which will move more than one box if $\lambda \neq \tau$. Lemma 5.3 implies that $W(\mu) = \binom{n}{2}^{-1/2} \sum_{i=1}^{n} c_{\mu}(i)$, so that moving more boxes of a partition to obtain W' from W corresponds to a larger step size. This is one sense in which τ has the smallest step size among non-trivial irreducible representations.

As discussed previously, a more quantitative measure of step size of the chain L_{λ} is the value of a_{λ} from Theorem 5.2. From Lemma 5.1, we have

$$\mathbb{E}(W' \mid W) = \left(\frac{\chi^{\lambda}(2)}{\dim(\lambda)}\right) W = \left(\frac{2}{n(n-1)} \sum_{i=1}^{n} c_{\lambda}(i)\right) W.$$

By the definition of $c_{\lambda}(i)$, it is easy to see that a_{λ} is minimal over non-trivial irreducible representations for $\lambda = \tau$, and strictly increases as boxes of a partition are moved from higher to lower rows (this observation motivates Lemma 5.4 below). This is another sense in which τ has the smallest step size among non-trivial irreducible representations.

Now, analogously to Section 4, in order to show that the error term from Theorem 5.2 is minimal over non-trivial irreducible representations at $\tau = (n-1, 1)$, we will show that for each conjugacy class K from Lemma 5.3, the term $T_K(\lambda) = \left(\frac{\chi^{\lambda}(K)}{\dim(\lambda)} - 1\right)/a_{\lambda}$ is minimal and at least -2 for $\lambda = \tau$. Moreover, we will define an ordering on all irreducible representations such that the largest non-trivial irreducible representation in the ordering is $(n-1, 1)$ and then show that T_K is a decreasing function with respect to the ordering. This non-standard ordering is a coarsening of the usual dominance ordering found in [34].

Definition. Let $\lambda = (\lambda_1, \ldots, \lambda_k)$ and $\mu = (\mu_1, \ldots, \mu_m)$ be two irreducible representations of S_n. We say λ succeeds μ, denoted $\lambda \succ \mu$, if $\lambda_1 \geqslant \mu_1$, $\lambda_i = \mu_i$ for $i = 2, \ldots, k-1$, and $\lambda_k \leqslant \mu_k$. For $i > m$ $(i > k)$, take $\mu_i = 0$ $(\lambda_i = 0)$.

It is obvious that $(n-1, 1) = \tau \succ \lambda$ for all non-trivial irreducible representations λ, so that to prove the claims above it is enough to show that $\lambda \succ \mu$ implies $T_K(\lambda) \leqslant T_K(\mu)$ for each K from Lemma 5.3. The following lemma is a simpler characterization of the succession relation.

Lemma 5.4. *For two irreducible representations λ and μ of S_n, $\lambda \succ \mu$ if and only if μ can be transformed from λ only by moving blocks from the associated tableau of λ from the top row to either the bottom row of λ or to start a new row.*

Proof. If the defining relations of succession hold, and the two representations are not equal, then either $\lambda_k < \mu_k$ or $\mu_{k+1} \neq 0$. In the former case move a block from λ_1 to μ_k, in the latter start a new row. Inductively, continuing in this way will eventually lead to μ, since the new tableau still succeeds μ.

https://doi.org/10.1017/S096354830999023X Published online by Cambridge University Press

Conversely, if μ can be made from λ with the described method, then clearly the defining relations of succession must hold. □

Lemma 5.4 states that in order to prove $\lambda \succ \mu$ implies $T_K(\lambda) \leqslant T_K(\mu)$ (thus that $\tau = (n-1,1)$ minimizes the error term), it is enough to prove the statement in the case $\mu = (\lambda_1 - 1, \lambda_2, \ldots, \lambda_k + 1)$, where, in order to cover both actions described in Lemma 5.4, we break convention and permit $\lambda_k = 0$ (only if $\lambda_{k-1} \neq 0$). We make one final (non-standard) definition before we begin the proof of the claims above.

Definition. Let $\lambda = (\lambda_1, \ldots, \lambda_k)$ and $\mu = (\lambda_1 - 1, \lambda_2, \ldots, \lambda_k + 1)$. Define a new tableau, the joint of λ and μ, by

$$\lambda \,|\, \mu = (\lambda_1 - 1, \lambda_2, \ldots, \lambda_k).$$

Now that we have suitable definitions, we can begin to prove the lemmas that will be used to obtain the main result.

Lemma 5.5. *For any tableau λ, $T_{\mathrm{id}}(\lambda) = 0$ and $T_{(2)}(\lambda) = -1$.*

Proof. The first assertion follows from the fact that for any representation, the identity element is mapped to the identity matrix so that $\chi^\lambda(\mathrm{id}) = \dim(\lambda)$. The second is obtained directly from the definition of a_λ. □

The next lemma states a simpler criterion for determining which of $T_{(3)}(\lambda)$ and $T_{(3)}(\mu)$ is larger.

Lemma 5.6. *With λ and μ as above, $T_{(3)}(\mu) - T_{(3)}(\lambda)$ is non-negative if and only if $f(\lambda)$ (defined below) is also non-negative.*

$$f(\lambda) = (\lambda_1 + \lambda_k - k)\left(\sum_{i=1}^{n-1} c_{\lambda|\mu}(i) - \binom{n}{2}\right) + (\lambda_1 - 1)(\lambda_k + 1 - k) - \sum_{i=1}^{n-1} c_{\lambda|\mu}^2(i) + \binom{n}{2} + 2\binom{n}{3}.$$

Proof.

$$T_{(3)}(\mu) - T_{(3)}(\lambda) = \frac{\left(1 - \frac{1}{2\binom{n}{3}}\left(\sum_i c_\lambda^2(i) - \binom{n}{2}\right)\right)a_\mu - \left(1 - \frac{1}{2\binom{n}{3}}\left(\sum_i c_\mu^2(i) - \binom{n}{2}\right)\right)a_\lambda}{a_\lambda a_\mu}. \tag{5.1}$$

Notice that $\sum_i c_\lambda(i)$ is maximized for $\lambda = (n)$, in which case it is $\binom{n}{2}$ and strictly less for all other tableaux. Thus, for all tableaux λ under consideration, $a_\lambda > 0$, so the non-negativity of (5.1) is

https://doi.org/10.1017/S096354830999023X Published online by Cambridge University Press

determined by the numerator. The numerator of (5.1) is equal to

$$\frac{\sum_i (c_\mu^2(i) - c_\lambda^2(i))}{2\binom{n}{3}} + \frac{\sum_i (c_\lambda(i) - c_\mu(i))}{\binom{n}{2}} + \frac{\sum_i c_\lambda^2(i) \sum_j c_\mu(j) - \sum_i c_\lambda(i) \sum_j c_\mu^2(j) + \binom{n}{2} \sum_i (c_\lambda(i) - c_\mu(i))}{2\binom{n}{3}\binom{n}{2}}. \tag{5.2}$$

Rewriting the sums

$$\sum_i c_\lambda^t(i) = \sum_i c_{\lambda|\mu}^t(i) + (\lambda_1 - 1)^t \quad \text{and} \quad \sum_i c_\mu^t(i) = \sum_i c_{\lambda|\mu}^t(i) + (\lambda_k + 1 - k)^t,$$

for $t = 1, 2$, and then simplifying implies that (5.2) is equal to

$$\begin{aligned} &\frac{(\lambda_k + 1 - k)^2 - (\lambda_1 - 1)^2}{2\binom{n}{3}} + \frac{(\lambda_1 - 1) - (\lambda_k + 1 - k)}{\binom{n}{2}} \\ &\quad + \frac{\big((\lambda_1 - 1)^2 - (\lambda_k + 1 - k)^2\big) \sum_i c_{\lambda|\mu}(i)}{2\binom{n}{3}\binom{n}{2}} \\ &\quad + \frac{(\lambda_1 - 1)(\lambda_k + 1 - k)\big((\lambda_1 - 1) - (\lambda_k + 1 - k)\big)}{2\binom{n}{3}\binom{n}{2}} \\ &\quad - \frac{((\lambda_1 - 1) - (\lambda_k + 1 - k)) \sum_i c_{\lambda|\mu}^2(i)}{2\binom{n}{3}\binom{n}{2}} \\ &\quad + \frac{\binom{n}{2}((\lambda_1 - 1) - (\lambda_k + 1 - k))}{2\binom{n}{3}\binom{n}{2}}. \end{aligned}$$

Multiplying by $\big(2\binom{n}{3}\binom{n}{2}\big)/((\lambda_1 - 1) - (\lambda_k + 1 - k))$ proves the lemma (since $\lambda_1 > \lambda_k$ and $k \geqslant 2$, this multiplication does not affect the non-negativity of the term). □

Now we can prove the following lemma.

Lemma 5.7. *Let λ and μ be two irreducible representations on the symmetric group. If $\lambda \succ \mu$, then $T_{(3)}(\lambda) \leqslant T_{(3)}(\mu)$.*

Proof. Lemma 5.6 implies that in order to show the monotonicity of $T_{(3)}$ with respect to the succession relation, we only need to show that $f(\lambda) \geqslant 0$ for all tableaux $\lambda = (\lambda_1, \ldots, \lambda_k)$ with $\lambda_1 > \lambda_2$ and $\lambda_k < \lambda_{k-1}$ (where λ_k may possibly be zero). In order to prove the lemma, we use induction on n, the number being partitioned. In each of the three cases below, we associate to each tableau λ of size $n + 1$ a tableau ϕ of size n, where $f(\phi) \geqslant 0$ by the induction hypothesis. It can be easily verified that in the case of $n = 3$, the only non-trivial tableau satisfying $\lambda_1 > \lambda_2$ and $\lambda_k < \lambda_{k-1}$ is $\lambda = (2, 1, 0)$, and $f(\lambda) \geqslant 0$.

https://doi.org/10.1017/S096354830999023X Published online by Cambridge University Press

Case 1: $\lambda_k \neq 0$.
For this case, let $\phi = (\lambda_1, \ldots, \lambda_{k-1}, \lambda_k - 1)$ be a partition of n and let ψ be equal to the partition $(\lambda_1 - 1, \ldots, \lambda_{k-1}, \lambda_k)$. Then we have

$$\begin{aligned} f(\lambda) &= (\lambda_1 + \lambda_k - k)\left(\sum_{i=1}^{n} c_{\lambda|\mu}(i) - \binom{n+1}{2}\right) \\ &\quad + (\lambda_1 - 1)(\lambda_k + 1 - k) - \sum_{i=1}^{n} c_{\lambda|\mu}^2(i) + \binom{n+1}{2} + 2\binom{n+1}{3}, \\ f(\phi) &= (\lambda_1 + (\lambda_k - 1) - k)\left(\sum_{i=1}^{n-1} c_{\phi|\psi}(i) - \binom{n}{2}\right) \\ &\quad + (\lambda_1 - 1)((\lambda_k - 1) + 1 - k) - \sum_{i=1}^{n-1} c_{\phi|\psi}^2(i) + \binom{n}{2} + 2\binom{n}{3}. \end{aligned}$$

Since $f(\phi)$ is non-negative by the induction hypothesis, we will show that $f(\phi) - f(\lambda) \leqslant 0$, which will prove the lemma for this case. Simplifying this expression using the identity $\binom{n+1}{k} = \binom{n}{k} + \binom{n}{k-1}$ and the fact that

$$\sum_{i=1}^{n-1} (c_{\phi|\psi}(i))^t = \sum_{i=1}^{n} (c_{\lambda|\mu}(i))^t - (\lambda_k - k)^t \qquad t = 1, 2,$$

we obtain

$$f(\phi) - f(\lambda) = (n-1)(\lambda_1 - 1) + (n - \lambda_1 + 1)(\lambda_k - k) - \sum_{i=1}^{n} c_{\lambda|\mu}(i) - \binom{n}{2}.$$

Now we bound the sum of the contents:

$$\begin{aligned} \sum_{i=1}^{n} c_{\lambda|\mu}(i) &\geqslant \sum_{i=1}^{\lambda_1 - 2} i + \sum_{i=1}^{\lambda_k} (i - k) + (n - (\lambda_1 - 1) - \lambda_k)(1 - (k-1)) \\ &= \frac{(\lambda_1 - 1)(\lambda_1 - 2)}{2} + \frac{\lambda_k(\lambda_k - 3)}{2} + (2 - k)(n - \lambda_1 + 1). \end{aligned}$$

The inequality follows because the first term on the right-hand side is the sum of the contents of row one, the second term is the sum of the contents of row k, and the third term is the number of boxes not in rows one or k times the minimum content of those boxes.

The inequality above shows that $f(\phi) - f(\lambda) \leqslant g(\lambda_1, \lambda_k)$, where the function g is defined by

$$\begin{aligned} g(\lambda_1, \lambda_k) &= (n-1)(\lambda_1 - 1) + (n - \lambda_1 + 1)(\lambda_k - 2) \\ &\quad - \frac{(\lambda_1 - 1)(\lambda_1 - 2)}{2} - \frac{\lambda_k(\lambda_k - 3)}{2} - \binom{n}{2}. \end{aligned}$$

Let $\mathcal{D} = \{(\lambda_1, \lambda_k) : \lambda_k \leqslant n + 1 - \lambda_1, 1 \leqslant \lambda_k \leqslant \lambda_1 - 2\}$ and notice that the allowable values of (λ_1, λ_k) are a subset of $\mathcal{D}$. A straightforward (but tedious) analysis shows that the maximum of g over the domain $\mathcal{D}$ is non-positive, which proves the lemma for this case.

https://doi.org/10.1017/S096354830999023X Published online by Cambridge University Press

Case 2: $\lambda_k = 0, \lambda_{k-1} \neq 1$.
In this case, let $\phi = (\lambda_1, \ldots, \lambda_{k-1} - 1, 0)$ and then let ψ be equal to the partition $(\lambda_1 - 1, \ldots, \lambda_{k-1} - 1, 1)$. Then let

$$f(\phi) = (\lambda_1 - k)\left(\sum_{i=1}^{n-1} c_{\phi|\psi}(i) - \binom{n}{2}\right) + (\lambda_1 - 1)(1-k) - \sum_{i=1}^{n-1} c_{\phi|\psi}^2(i) + \binom{n}{2} + 2\binom{n}{3},$$

and notice that $f(\phi) \geqslant 0$ by the induction hypothesis. After simplifying, using the fact that

$$\sum_{i=1}^{n-1}(c_{\phi|\psi}(i))^t = \sum_{i=1}^{n}(c_{\lambda|\mu}(i))^t - (\lambda_{k-1} - k + 1)^t, \quad t = 1, 2,$$

we obtain

$$f(\phi) - f(\lambda) = n(\lambda_1 - k) + (\lambda_{k-1} - \lambda_1 + 1)(\lambda_{k-1} - k + 1) - n^2. \tag{5.3}$$

The right-hand side of equation (5.3) is increasing in λ_1, so that the maximum is attained for $\lambda_1 = n$. In this case it is easy to see (5.3) is non-positive.

Case 3: $\lambda = (\lambda_1, \ldots, \lambda_{k-2}, 1, 0)$.
In this case, let $\phi = (\lambda_1, \ldots, \lambda_{k-2}, 0)$ and let ψ be equal to the partition $(\lambda_1 - 1, \ldots, \lambda_{k-2}, 1)$. Analogously to the previous two cases, define

$$f(\phi) = (\lambda_1 - (k-1))\left(\sum_{i=1}^{n-1} c_{\phi|\psi}(i) - \binom{n}{2}\right) + (\lambda_1 - 1)(1 - (k-1)) - \sum_{i=1}^{n-1} c_{\phi|\psi}^2(i) + \binom{n}{2} + 2\binom{n}{3},$$

which implies

$$f(\phi) - f(\lambda) = (n + k)(\lambda_1 - n) + \sum_{i=1}^{n-1} c_{\phi|\psi}(i) - \binom{n}{2} + 3 - 2k - \lambda_1.$$

Using that $\lambda_1 \leqslant n$ and thus $\sum_{i=1}^{n-1} c_{\phi|\psi}(i) \leqslant \binom{n-1}{2}$, it is easy to see the term is negative. □

Now we move on to the final conjugacy class needed to prove the result. As before, we have the following lemma, which states a simpler criterion for determining which of $T_{(2,2)}(\lambda)$ and $T_{(2,2)}(\mu)$ is larger.

Lemma 5.8. *With λ and μ as above, $T_{(2,2)}(\mu) - T_{(2,2)}(\lambda)$ is non-negative if and only if $h(\lambda)$ (defined below) is also non-negative.*

$$h(\lambda) = 2(\lambda_1 + \lambda_k - k)\left(\binom{n}{2} - \sum_{i=1}^{n-1} c_{\lambda|\mu}(i)\right) - 2(\lambda_1 - 1)(\lambda_k + 1 - k) - 2\binom{n}{2}\sum_{i=1}^{n-1} c_{\lambda|\mu}(i) + \left(\sum_{i=1}^{n-1} c_{\lambda|\mu}(i)\right)^2 + 3\sum_{i=1}^{n-1} c_{\lambda|\mu}^2(i) + 6\binom{n}{4} - 2\binom{n}{2}.$$

https://doi.org/10.1017/S096354830999023X Published online by Cambridge University Press

Proof.

$$T_{(2,2)}(\mu) - T_{(2,2)}(\lambda) = \frac{\left(1 - \frac{\chi^{\lambda}(2,2)}{\dim(\lambda)}\right)a_\mu - \left(1 - \frac{\chi^{\mu}(2,2)}{\dim(\mu)}\right)a_\lambda}{a_\lambda a_\mu}. \tag{5.4}$$

As in Lemma 5.6, the non-negativity of (5.4) is determined by the numerator. Using the formulas from Lemma 5.3 and rearranging, the numerator of (5.4) is equal to

$$\begin{aligned}&\frac{\left(\sum_i c_\mu(i)\right)^2 - \left(\sum_i c_\lambda(i)\right)^2 - 3\left(\sum_i [c_\mu^2(i) - c_\lambda^2(i)]\right)}{6\binom{n}{4}} + \frac{\sum_i (c_\lambda(i) - c_\mu(i))}{\binom{n}{2}}\\ &\quad + \frac{\left[\left(\sum_i c_\lambda(i)\right)^2 - 3\sum_i c_\lambda^2(i) + 2\binom{n}{2}\right]\sum_j c_\mu(j)}{6\binom{n}{4}\binom{n}{2}}\\ &\quad - \frac{\left[\left(\sum_i c_\mu(i)\right)^2 - 3\sum_i c_\mu^2(i) + 2\binom{n}{2}\right]\sum_j c_\lambda(j)}{6\binom{n}{4}\binom{n}{2}}.\end{aligned}$$

Analogously to Lemma 5.6 for 3-cycles, simplifying the term by rewriting

$$\sum_i c_\lambda^t(i) = \sum_i c_{\lambda|\mu}^t(i) + (\lambda_1 - 1)^t \quad \text{and} \quad \sum_i c_\mu^t(i) = \sum_i c_{\lambda|\mu}^t(i) + (\lambda_k + 1 - k)^t,$$

for $t = 1, 2$, yields a term proportional to $h(\lambda)$. In this case, the constant of proportionality is $\left(6\binom{n}{4}\binom{n}{2}\right)/((\lambda_1 - 1) - (\lambda_k + 1 - k))$, which is positive and so does not affect the non-negativity. □

Now we can prove the following lemma.

Lemma 5.9. *Let λ and μ be two irreducible representations on the symmetric group. If $\lambda \succ \mu$, then $T_{(2,2)}(\lambda) \leqslant T_{(2,2)}(\mu)$.*

Proof. We will exactly follow the strategy of the proof of Lemma 5.7 with the function h replacing f.

Case 1: $\lambda_k \neq 0$.
For this case, let $\phi = (\lambda_1, \ldots, \lambda_{k-1}, \lambda_k - 1)$ be a partition of n, and let ψ be equal to the partition $(\lambda_1 - 1, \ldots, \lambda_{k-1}, \lambda_k)$. Then, following the proof of Lemma 5.7,

$$\begin{aligned}(h(\phi) - h(\lambda))/2 &= (\lambda_1 - 1)(\lambda_k - k + 1) + \lambda_k\binom{n}{2} - (k+1)\binom{n}{2}\\ &\quad + (n - \lambda_k + k + 1)\sum_{i=1}^{n} c_{\lambda|\mu}(i) - n(\lambda_1 + \lambda_k - k - 1) - 3\binom{n}{3}\\ &\leqslant (\lambda_1 - 1)(\lambda_k - k + 1) + \binom{n}{2}(\lambda_k - k - 1)\\ &\quad + (n - \lambda_k + k + 1)\binom{n-1}{2} - n(\lambda_1 + \lambda_k - k - 1) - 3\binom{n}{3}\\ &= (\lambda_1 - 2)(\lambda_k - k) - \lambda_1(n-1) \leqslant 0.\end{aligned}$$

https://doi.org/10.1017/S096354830999023X Published online by Cambridge University Press

Here the first inequality follows because $n - \lambda_k + k + 1 \geqslant 0$ and $\sum_i c_{\lambda|\mu}(i) \leqslant \binom{n-1}{2}$, and the final inequality comes from $\lambda_k - k \leqslant n - 1$ and $\lambda_1 \geqslant 2$.

Case 2: $\lambda_k = 0$, $\lambda_{k-1} \neq 1$.
In this case, let $\phi = (\lambda_1, \ldots, \lambda_{k-1} - 1, 0)$, and then let ψ be equal to the partition $(\lambda_1 - 1, \ldots, \lambda_{k-1} - 1, 1)$. Then we have

$$\begin{aligned}(h(\phi) - h(\lambda))/2 &= (\lambda_1 - k)(\lambda_{k-1} - k + 1 - n) + \binom{n}{2}(\lambda_{k-1} - k + 1) \\ &\quad + (n - \lambda_{k-1} + k - 1)\sum_{i=1}^{n} c_{\lambda|\mu}(i) - (\lambda_{k-1} - k + 1)^2 - 3\binom{n}{3} + n \\ &\leqslant (\lambda_1 - k)(\lambda_{k-1} - k + 1 - n) + \binom{n}{2}(\lambda_{k-1} - k + 1) \\ &\quad + (n - \lambda_{k-1} + k - 1)\binom{n-1}{2} - (\lambda_{k-1} - k + 1)^2 - 3\binom{n}{3} + n \\ &= (\lambda_1 - \lambda_{k-1} - 2)(\lambda_{k-1} - k + 1 - n).\end{aligned}$$

The inequality follows because $n - \lambda_{k-1} + k - 1 \geqslant 0$ and $\sum_i c_{\lambda|\mu}(i) \leqslant \binom{n-1}{2}$. If $\lambda_1 - \lambda_{k-1} - 2 \geqslant 0$, then this case is shown as $\lambda_{k-1} - k + 1 - n \leqslant 0$. If not, then $\lambda_1 = \lambda_{k-1} + 1$, which implies $\lambda_j = \lambda_{k-1}$ for $j = 2 \cdots k - 1$ (since $\lambda_1 \geqslant \lambda_2 + 1$). In this case, $(k-1)\lambda_{k-1} = n$ and

$$\sum_i c_{\lambda|\mu}(i) = \sum_{i=1}^{k-1}\sum_{j=1}^{\lambda_{k-1}} j - i = \frac{n}{2}(\lambda_{k-1} + 1 - k).$$

Then we have

$$\begin{aligned}(h(\phi) - h(\lambda))/2 &= (\lambda_{k-1} + 1 - k)(\lambda_{k-1} - k + 1 - n) + \binom{n}{2}(\lambda_{k-1} - k + 1) \\ &\quad + \frac{n}{2}(\lambda_{k-1} + 1 - k)(n - \lambda_{k-1} + k - 1) - (\lambda_{k-1} - k + 1)^2 - 3\binom{n}{3} + n.\end{aligned}$$

This is a downward-facing parabola in λ_{k-1} with roots equal to $n + k - 1$ and $n + k - 4$, neither of which is a possible value of λ_{k-1} since k must be at least three.

Case 3: $\lambda = (\lambda_1, \ldots, \lambda_{k-2}, 1, 0)$.
Let $\phi = (\lambda_1, \ldots, \lambda_{k-2}, 0)$ and then let ψ be equal to the partition $(\lambda_1 - 1, \ldots, \lambda_{k-2}, 1)$. Then we have

$$\begin{aligned}(h(\phi) - h(\lambda))/2 &= (n + k - 3)\left(\sum_{i=1}^{n} c_{\lambda|\mu}(i) + k - \lambda_1 - \binom{n-1}{2}\right) \\ &\quad + n(4 - k) - 2\lambda_1 + k - (2 - k)^2 \qquad (5.5) \\ &\leqslant (n - 1)(4 - \lambda_1) + k(2 - \lambda_1) \qquad (5.6)\end{aligned}$$

To see the inequality, notice that $n + k - 3 \geqslant 0$ and $\sum_i c_{\lambda|\mu}(i) \leqslant \binom{n-1}{2}$.

The term (5.6) is clearly negative for $\lambda_1 \geqslant 4$, and the smaller cases can be individually handled starting from (5.5) (for example, $\lambda_1 = 2$ implies $\sum_i c_{\lambda|\mu}(i) = -\binom{n}{2}$ and $k = n + 1$). □

https://doi.org/10.1017/S096354830999023X Published online by Cambridge University Press

The main result of this section is stated in the following.

Theorem 5.10. *If $\lambda \succ \mu$ for non-trivial irreducible representations λ and μ, then the error term from Theorem 5.2 associated to λ is no larger than the error term associated to μ. In particular, the error term is minimal for $\tau = (n-1, 1)$.*

Proof. By the previous lemmas and remarks, the theorem will be proved if $T_{(3)}(\tau)$ and $T_{(2,2)}(\tau)$ are no less than -2 (for K equal to the identity or transposition conjugacy class, T_K is constant). Using the formulas from Lemma 5.3, or the fact [34] that, for any permutation g, $\chi^\tau(g)$ is equal to the number of fixed points of g minus one and $\dim(\tau) = n-1$, we have

$$T_{(3)}(\tau) = -3/2,$$
$$T_{(2,2)}(\tau) = -2.$$

□

6. Eigenvalue characterization

Heuristically, increasing the step size of a Markov chain has the effect of making the chain become 'random' faster. This suggests that for an ergodic chain, the step size may be related to the rate of convergence to stationarity. From basic facts about reversible Markov chains on a finite state space [9], all the chains previously defined are ergodic if and only if they are irreducible and aperiodic. In this case, the rate of convergence to stationarity is determined by the eigenvalues, where the eigenvalues with the largest moduli make the largest asymptotic contributions. Because of this relationship, we will express the error terms from the past sections in terms of the eigenvalues of the Markov chains used to induce the exchangeable pairs.

We first examine the chains from Section 4 (and hence also Section 3 as previously mentioned). For a fixed value of t, the eigenvalues for $L_t(i, j)$ have been computed in [22], but we include the proof since it is illustrative. Recall the definitions and notation from Section 4.

Lemma 6.1 ([22]). *For fixed $t \in \{0, 1, \ldots, n\}$, the eigenvalues of $L_t(i, j)$ are $K_s(t)/v_s$ for $s = 0 \cdots n$.*

Proof. By definitions and Lemma 4.3,

$$\sum_{j=0}^{n} L_t(i,j)K_s(j) = \frac{1}{q^n}\sum_{r=0}^{n}\frac{K_r(i)K_r(t)}{v_r^2}\sum_{j=0}^{n} v_j K_s(j)K_r(j)$$
$$= \left(\frac{K_s(t)}{v_s}\right)K_s(i).$$

In other words, if M is the transition matrix of $L_t(i, j)$ and $\mathbf{v}$ is the vector with coordinate j equal to $K_s(j)$, then $M\mathbf{v} = (K_s(t)/v_s)\mathbf{v}$. □

A proof of the next lemma follows along the lines of the proof of Lemma 4.11.

https://doi.org/10.1017/S096354830999023X Published online by Cambridge University Press

Lemma 6.2. *The eigenvalues of $L_T(i,j)$ as defined in Section 4 are, for $s = 0, \ldots, n$,*

$$\lambda_s = \sum_{t=0}^{n} b_t \left(\frac{K_s(t)}{v_s} \right).$$

Recall from the proof of Theorem 4.13 that the bound on the error depends only on the following term for $s = 0, 1, 2$:

$$\frac{\left(\sum_{t=0}^{n} b_t \frac{K_s(t)}{v_s}\right) - 1}{1 - \left(\sum_{t=0}^{n} b_t \frac{K_1(t)}{v_1}\right)}. \tag{6.1}$$

By Lemma 6.2, (6.1) can be rewritten as $(\lambda_s - 1)/(1 - \lambda_1)$ for $s = 0, 1, 2$. For $s = 0$ and $s = 1$, (6.1) does not depend on values b_t, so we have the following theorem.

Theorem 6.3. *The size of the error term given by Stein's method via the family of chains from Section 4 is a monotone increasing function of*

$$\frac{\lambda_2 - 1}{1 - \lambda_1},$$

where λ_1 and λ_2 are defined as in Lemma 6.2.

A natural problem that arises is to ascertain under what setting the eigenvalues that determine the error term will be the eigenvalues with largest moduli (not including $\lambda_0 = 1$), as these are the eigenvalues with the largest contribution to the rate of convergence. It follows immediately from definitions that in the case $b_0 + b_1 = 1$, where the error term is minimal, $\lambda_s = 1 - \frac{b_1 qs}{n(q-1)}$, so that the ordering of the eigenvalues corresponds to the subscript notation. In this case the eigenvalues that affect the error term will be the largest in moduli if and only if $b_1 \leqslant \frac{2n(q-1)}{q(n+2)}$.

For the chains L_τ from Section 5, the eigenvalues have already been computed in [22] using orthogonality relations. Recall the definitions and notation from Section 5.

Lemma 6.4 ([22]). *The eigenvalues of L_τ as defined in Section 5 are, for conjugacy classes C of the group G,*

$$\lambda_C = \frac{\chi^\tau(C)}{\dim \tau}.$$

Also, from Theorem 5.2 and the remarks following it, the bound on the error term depends only on the following term for the conjugacy classes $K = (\mathrm{id}), (2), (3)$, and $(2,2)$:

$$\frac{\frac{\chi^\tau(K)}{\dim \tau} - 1}{1 - \frac{\chi^\tau(2)}{\dim \tau}}. \tag{6.2}$$

By Lemma 6.4, (6.2) can be rewritten as $(\lambda_K - 1)/(1 - \lambda_{(2)})$ for $K = (\mathrm{id}), (2), (3)$, and $(2,2)$. For $K = (\mathrm{id})$ and $K = (2)$, (6.2) does not depend on the irreducible representation τ used to generate L_τ, so we have the following theorem.

https://doi.org/10.1017/S096354830999023X Published online by Cambridge University Press

Theorem 6.5. *The size of the error term given by Stein's method via the family of chains from Section 5 is a monotone increasing function of*

$$\frac{\lambda_K - 1}{1 - \lambda_{(2)}},$$

where λ_K is defined as in Lemma 6.4 and $K = (3)$ or $K = (2)(2)$.

Once again it is natural to ask, for a given irreducible representation τ, whether the three eigenvalues that affect the error term are those with the largest moduli (not including $\lambda_{(\mathrm{id})} = 1$). For $\tau = (n-1, 1)$, the representation where the error term is minimal, it is well known [34] that

$$\frac{\chi^\tau(K)}{\dim \tau} = \frac{F_K - 1}{n-1},$$

where F_K is defined to be the number of fixed points of K. In this case, the eigenvalues that affect the error term are the eigenvalues with the second, third, and fourth largest moduli.

7. Poisson approximation

For the final two sections we change focus from approximation by the normal distribution to approximation by the Poisson distribution. A metric routinely used for integer-supported random variables X and Y is the total variation distance defined by

$$\|X - Y\|_{TV} = \frac{1}{2}\left|\sum_{k\in\mathbb{Z}} \big(\mathbb{P}(X = k) - \mathbb{P}(Y = k)\big)\right|.$$

We will examine in the following theorem how the step size of the Markov chain that induces an exchangeable pair affects the error term.

Theorem 7.1 ([13]). *Let $W = \sum_i X_i$, a sum of random variables, such that $\mathbb{E}(W) = \lambda$. Let (W, W') be an exchangeable pair, c any constant, and let Poi_λ denote the Poisson distribution with mean λ. Then, for C_λ a constant only depending on λ,*

$$\begin{aligned}\|W - \mathrm{Poi}_\lambda\|_{TV} &\\ \leqslant C_\lambda\big[\mathbb{E}|\lambda - c\,\mathbb{P}(W' = W + 1 \,|\, \{X_i\})| + \mathbb{E}|W - c\,\mathbb{P}(W' = W - 1 \,|\, \{X_i\})|\big].&\end{aligned} \tag{7.1}$$

Remarks. **(1)** It has been shown [33] that a result similar to Theorem 7.1, but with additional error terms, still holds assuming only that W and W' are equally distributed.

(2) Ideally, c should be chosen so that we have the approximate equalities

$$\mathbb{P}(W' = W + 1 \,|\, \{X_i\}) \approx \frac{\lambda}{c}, \tag{7.2}$$

$$\mathbb{P}(W' = W - 1 \,|\, \{X_i\}) \approx \frac{W}{c}. \tag{7.3}$$

It is shown in [13] that, intuitively, the existence of such a constant is likely, a heuristic that is reinforced in the examples presented there. In fact, if $(W' - W) \in \{-1, 0, 1\}$ and $\mathbb{E}(W' - W \,|\, W) = -a(W - \mathbb{E}(W))$, it is easy to see that for the choice of $c = 1/a$ we have the same error in the approximate equalities (7.2) and (7.3). This is in general a useful guide for the choice of the constant c (for more on this line of thought, see [32]).

https://doi.org/10.1017/S096354830999023X Published online by Cambridge University Press

One of the main technical details in analysing (7.1) for different exchangeable pairs is the choice of the constant c. It would be preferable to have a systematic method of choosing this constant based on the exchangeable pair so that the results here are not contrived. Ideally, we would choose the constant c to minimize the error terms from Theorem 7.1, or more feasibly, their Cauchy–Schwarz bound (choose c such that the expectation of the terms in the absolute value signs are zero). However, in the examples presented here, we will choose the constant c to yield the best-possible bound under the constraint that the terms in the absolute value signs are positive. Admittedly, part of the reason for this restriction is technical convenience, but, in practice, choosing the constant in this way is typical (see the examples of [13]). In the next section we will compare the error terms using both of these strategies in a small example.

Both of the examples presented here are sums of i.i.d. random variables (Bernoulli and geometric). Even the simplest introduction of dependence (*e.g.*, the hypergeometric distribution) yield results that make the type of analysis in this paper difficult. Because we are in the setting of independence, we can use the same family of Markov chains for both examples. Given a vector $(X_1, \ldots, X_n)$ of non-negative integer-valued i.i.d. random variables, the next step in the chain follows the rule of choosing k coordinates uniformly at random and replacing them with k new i.i.d. random variables (with the same distribution as the original). It is not hard to see that this chain is reversible with respect to vectors of i.i.d. random variables and hence generates an exchangeable pair. Extending this exchangeable pair to the sum of the components of the vector allows for the application of Theorem 7.1.

Finally, notice that under this chain it is not clear how modifying the number of coordinates chosen to be selected (*i.e.*, varying k) will affect the error term.

8. Binomial distribution

It is well known that the binomial distribution with parameters n and p converges to a Poisson distribution with mean λ as n tends to infinity if np tends to λ. For simplicity, in this section we consider the case where $p = 1/n$, so that $\lambda = 1$. In this case we will show that, among the exchangeable pairs associated with the family of Markov chains described in Section 7, the term from Theorem 7.1 is minimized when $k = 1$. First we will prove some lemmas that will be used to compute the error term from the theorem.

Lemma 8.1. *Let $\mathbb{P}_k$ denote probability under the chain that substitutes k coordinates as described in Section 7. Then*

$$\mathbb{P}_k(W' = W + 1 \,|\, \{X_i\}) = \sum_{i=0}^{k-1} \binom{k}{i+1} \frac{(n-1)^{k-i-1}}{n^k} \frac{\binom{W}{i}\binom{n-W}{k-i}}{\binom{n}{k}},$$

$$\mathbb{P}_k(W' = W - 1 \,|\, \{X_i\}) = \sum_{i=1}^{k} \binom{k}{i-1} \frac{(n-1)^{k-i+1}}{n^k} \frac{\binom{W}{i}\binom{n-W}{k-i}}{\binom{n}{k}}.$$

Proof. Let the random variable Y be the number of ones in the k coordinates chosen. Then $\mathbb{P}_k(W' = W + 1 \,|\, Y = i)$ is the probability of $i + 1$ ones in the binomial distribution with

https://doi.org/10.1017/S096354830999023X Published online by Cambridge University Press

parameters k and $p = 1/n$, which implies

$$\begin{aligned}\mathbb{P}_k(W' = W+1 \,|\, \{X_i\}) &= \sum_{i=0}^{k-1} \mathbb{P}_k(W' = W+1 \,|\, Y = i)\mathbb{P}_k(Y = i) \\ &= \sum_{i=0}^{k-1} \binom{k}{i+1} \frac{(n-1)^{k-i-1}}{n^k} \frac{\binom{W}{i}\binom{n-W}{k-i}}{\binom{n}{k}}.\end{aligned}$$

Conditioning and summing over Y also yields the expression for $\mathbb{P}_k(W' = W-1 \,|\, \{X_i\})$ in the lemma. □

For the remainder of the section define

$$c_k = \left(\frac{n}{k}\right)\left(\frac{n}{n-1}\right)^{k-1}. \tag{8.1}$$

The next two lemmas prove a useful property of the constant c_k.

Lemma 8.2. $1 - c_k\mathbb{P}_k(W' = W+1 \,|\, \{X_i\}) \geqslant 0.$

Proof. Conditioning and summing over the random variable Y equal to the number of ones chosen in the k coordinates,

$$\begin{aligned}\mathbb{P}_k(W' = W+1 \,|\, \{X_i\}) &= \sum_{i=0}^{k-1} \binom{k}{i+1}\left(\frac{n-1}{n}\right)^{k-i-1}\left(\frac{1}{n}\right)^{i+1} \mathbb{P}_k(Y = i) \\ &= \sum_{i=0}^{k-1} \left(\frac{k}{n(i+1)}\right)\left(\frac{n-1}{n}\right)^{k-1}\binom{k-1}{i}\left(\frac{1}{n-1}\right)^{i} \mathbb{P}_k(Y = i) \\ &\leqslant \sum_{i=0}^{k-1} \left(\frac{k}{n}\right)\left(\frac{n-1}{n}\right)^{k-1} \mathbb{P}_k(Y = i) \\ &\leqslant \frac{k}{n}\left(\frac{n-1}{n}\right)^{k-1}.\end{aligned}$$

Here the first inequality follows from the fact that $k \leqslant n$. □

Lemma 8.3. $W - c_k\mathbb{P}_k(W' = W-1 \,|\, \{X_i\}) \geqslant 0.$

Proof. For $W = 0$, the lemma is trivially true. For $W \neq 0$, we condition and sum over the random variable Y equal to the number of ones chosen in the k coordinates to obtain

$$\begin{aligned}\mathbb{P}_k(W' = W-1 \,|\, \{X_i\}) &= \sum_{i=1}^{k} \mathbb{P}_k(W' = W-1 \,|\, Y = i)\left[\frac{\binom{W}{i}\binom{n-W}{k-i}}{\binom{n}{k}}\right] \\ &= \sum_{i=1}^{k} \binom{k}{i-1}\left(\frac{n-1}{n}\right)^{k-i+1}\left(\frac{1}{n}\right)^{i-1}\left(\frac{Wk}{n}\right)\left[\frac{\binom{W-1}{i-1}\binom{n-W}{k-i}}{i\binom{n-1}{k-1}}\right] \\ &\leqslant \left(\frac{Wk}{n}\right)\left(\frac{n-1}{n}\right)^{k-1}\sum_{i=1}^{k}\binom{k}{i-1}\frac{1}{n(n-1)^{i-2}}\left[\frac{\binom{W-1}{i-1}\binom{(n-1)-(W-1)}{(k-1)-(i-1)}}{\binom{n-1}{k-1}}\right] \\ &\leqslant \left(\frac{Wk}{n}\right)\left(\frac{n-1}{n}\right)^{k-1}.\end{aligned}$$

https://doi.org/10.1017/S096354830999023X Published online by Cambridge University Press

To see the final inequality, notice that for each summand, the second part of the product is a probability of the hypergeometric distribution and the first part of the product is at most one. □

The previous two lemmas show that for $c = c_k$ defined by (8.1), the terms within the absolute values in (7.1) are positive. Under this constraint, note that the error terms are decreasing in c_k and that $1 - c_k \mathbb{P}_k(W' = W + 1 \,|\, W = 0) = 0$. These observations imply that among constants satisfying Lemmas 8.2 and 8.3, the error from Theorem 7.1 is minimized for each k when c_k is defined as (8.1). As discussed in the previous section, this is a natural way to choose the constant in the approximation that allows for the analysis done here.

We pause here to show in a simple example the difference in the error terms from Theorem 7.1 when choosing the constant c according to the two approaches outlined in Section 7. First, we will determine the error terms using the strategy we take here in the case $k = 1$. For the following computations, recall that $p = 1/n$. We have

$$1 - c_1 \mathbb{P}_1(W' = W + 1 \,|\, \{X_i\}) = 1 - (n - W)p = pW,$$
$$W - c_1 \mathbb{P}_1(W' = W - 1 \,|\, \{X_i\}) = W - W(1 - p) = pW,$$

which implies that the error from Theorem 7.1 is equal to $2p$ (from [13], $C_\lambda = 1$ for $\lambda \leqslant 1$). Choosing instead $c_1' = n/(1 - p)$, so that $c_1' \mathbb{P}_1(W' = W + 1) = 1$ (this was discussed as the alternative system of choosing c), we obtain

$$1 - c_1' \mathbb{P}_1(W' = W + 1 \,|\, \{X_i\}) = 1 - \frac{(n - W)p}{1 - p} = \frac{p(W - 1)}{1 - p},$$
$$W - c_1' \mathbb{P}_1(W' = W - 1 \,|\, \{X_i\}) = W - W = 0,$$

which implies that the error from Theorem 7.1 is equal to

$$\frac{p\mathbb{E}|W - 1|}{1 - p} \leqslant \frac{p}{\sqrt{1 - p}}.$$

In the limit, the two error terms differ in quality only by a constant and c_1 is asymptotically equal to c_1'. Although the Cauchy–Schwarz approach using c_1' asymptotically yields a better constant, computing the appropriate moment information for general k using this scheme is much more difficult than the strategy we choose. Also, this small example suggests that the Cauchy–Schwarz approach will yield superior asymptotic rates only in the constant, so that it is not worth the extra effort of computing the more complicated (and higher) moment information needed in order to undertake the type of analysis presented in this paper. Finally, we note that using the chain here (with $k = 1$), it is possible to use intermediate terms in the proof of Theorem 7.1 with the constant c_1 to obtain the superior upper bound of p [13]; however, this approach does not carry over to the chains with larger step size.

Moving forward, in order to apply the theorem, we need to take the expected value of the terms in Lemma 8.1. The next lemma has a nice expression for the expectation we need.

Lemma 8.4.

$$\mathbb{E}\left[\binom{W}{i}\binom{n - W}{k - i}\right] i!(k - i)! = n \cdots (n - k + 1)\frac{(n - 1)^{k-i}}{n^k}.$$

https://doi.org/10.1017/S096354830999023X Published online by Cambridge University Press

Proof.

$$\mathbb{E}(s^W r^{n-W}) = r^n \mathbb{E}\left[\left(\frac{s}{r}\right)^W\right] = \left(\frac{s}{n} + r\left(\frac{n-1}{n}\right)\right)^n. \tag{8.2}$$

Taking i derivatives with respect to s and $k-i$ derivatives with respect to r of (8.2) and evaluating at $r = s = 1$ implies the lemma. □

The final lemma in this section establishes bounds on the error from Theorem 7.1.

Lemma 8.5. *Both of* $\mathbb{E}[c_k \mathbb{P}_k(W' = W+1 \,|\, \{X_i\})]$ *and* $\mathbb{E}[c_k \mathbb{P}_k(W' = W-1 \,|\, \{X_i\})]$ *are bounded above by*

$$\left(\frac{n-1}{n}\right).$$

Proof. From Lemmas 8.1 and 8.4 we have

$$\begin{aligned}
\mathbb{E}[c_k \mathbb{P}_k(W' = W+1 \,|\, \{X_i\})] &= c_k \sum_{i=0}^{k-1} \binom{k}{i+1} \binom{k}{i} \left(\frac{(n-1)^{2k-2i-1}}{n^{2k}}\right) \\
&= c_k \left(\frac{n-1}{n}\right)^k \left(\frac{k}{n}\right) \sum_{i=0}^{k-1} \binom{k-1}{i}^2 \left(\frac{(n-1)^{k-i-1}}{n^{k-1}(n-1)^i}\right) \left(\frac{k}{(i+1)(k-i)}\right) \\
&= \left(\frac{n-1}{n}\right) \sum_{i=0}^{k-1} \binom{k-1}{i} \left(\frac{(n-1)^{k-i-1}}{n^{k-1}}\right) \left(\binom{k-1}{i} \frac{1}{(n-1)^i}\right) \left(\frac{k}{(i+1)(k-i)}\right) \\
&\leqslant \left(\frac{n-1}{n}\right).
\end{aligned}$$

The inequality follows from the fact that each summand is the product of two terms no larger than one and a probability.

For the remaining term, exchangeability implies

$$\mathbb{E}[c_k \mathbb{P}_k(W' = W-1 \,|\, \{X_i\})] = \mathbb{E}[c_k \mathbb{P}_k(W' = W+1 \,|\, \{X_i\})],$$

which proves the lemma. □

Theorem 8.6. *For the values of* c_k *defined previously, the error term from Theorem 7.1 is minimized for* $k = 1$ *and is equal to* $2p$.

Proof. The final bound was computed previously in this section, and the fact that it is minimal follows directly from Lemma 8.5 and the easily verified fact

$$\mathbb{E}[c_1 \mathbb{P}_1(W' = W+1 \,|\, \{X_i\})] = \mathbb{E}[c_1 \mathbb{P}_1(W' = W-1 \,|\, \{X_i\})] = \frac{n-1}{n}.$$ □

https://doi.org/10.1017/S096354830999023X Published online by Cambridge University Press

9. Negative binomial distribution

The final example presented in this paper is the approximation of the negative binomial distribution by the Poisson. A random variable X has the geometric distribution with parameter p if $\mathbb{P}(X = i) = (1-p)^i p$ for all non-negative integers i. Classically, the random variable X is viewed as the number of failures before the first success in a sequence of independent Bernoulli trials, each with probability of success equal to p. The random variable W is negative binomial with parameters r and p if $W = \sum_{i=1}^{r} X_i$, where the X_i are independent geometric random variables with parameter p. By viewing W as the number of failures before r successes have occurred in a sequence of Bernoulli trials, it is easy to see that for all non-negative integers i, $\mathbb{P}(W = i) = \binom{r+i-1}{i}(1-p)^i p^r$. We will use Theorem 7.1 to approximate W by Poi(λ), where λ is the mean of W equal to $r(1-p)/p$ (the mean of a geometric random variable is $(1-p)/p$).

For fixed λ, $p = r/(\lambda + r)$, so

$$\mathbb{P}(W = i) = \frac{\lambda^i}{i!}\frac{(r+i-1)!}{(r-1)!(\lambda+r)^i}\left(1+\frac{\lambda}{r}\right)^{-r}. \tag{9.1}$$

As r goes to infinity, the distribution converges to a Poisson distribution with mean λ. However, for fixed λ, p approaches one as r goes to infinity, so that when p is small the negative binomial will not be approximately Poisson. Because of this fact, in this example we will not obtain a result as straightforward as Theorem 8.6. For some values of p, the optimal error term does not occur with the smallest step size. We will prove all supporting lemmas for general p, but the final theorem will have a natural restriction on the value of p. For this case we will show that among the exchangeable pairs associated with the family of Markov chains described in Section 7, the term from Theorem 7.1 is minimized when $k = 1$. First we will prove some lemmas that will be used to compute the error term from the theorem.

Lemma 9.1. *Let $\mathbb{P}_k$ denote probability under the chain that substitutes k coordinates as described in Section 7. Then*

$$\mathbb{P}_k(W' = W+1 \,|\, \{X_i\}) = \binom{r}{k}^{-1} \sum_{\substack{\{i_1,\ldots,i_k\} \\ \subseteq\{1,\ldots,r\}}} \binom{k+\sum_j X_{i_j}}{k-1}(1-p)^{(\sum_j X_{i_j}+1)}p^k,$$

$$\mathbb{P}_k(W' = W-1 \,|\, \{X_i\}) = \binom{r}{k}^{-1} \sum_{\substack{\{i_1,\ldots,i_k\} \\ \subseteq\{1,\ldots,r\}}} \binom{k+\sum_j X_{i_j}-2}{k-1}(1-p)^{(\sum_j X_{i_j}-1)}p^k.$$

Proof. This is immediate from conditioning and summing over the subset of $(X_1, \ldots, X_r)$ chosen. □

For the remainder of the section define

$$c_k = \lambda a_k \left(k\binom{k+a_k-2}{k-1}(1-p)^{a_k}p^{k-1}\right)^{-1}, \tag{9.2}$$

https://doi.org/10.1017/S096354830999023X Published online by Cambridge University Press

where a_k is the maximum of one and integer part of $(k-2)(1-p)/p$. The next lemma states a useful property of a_k; the proof can be found in [27], but it is elementary so we will include it.

Lemma 9.2. *Let r be a positive integer, let x be any non-negative integer, and let a be the integer part of $(r-1)(1-p)/p$. If $f(y) = \binom{r+y-1}{r-1}(1-p)^y p^r$, then $f(x) \geqslant f(x-1)$ for $x \leqslant a$, and is strictly decreasing otherwise. In particular, a is the mode of a negative binomial random variable with parameters r and p.*

Proof. The ratio of f evaluated at consecutive integers is given by

$$\frac{f(x+1)}{f(x)} = \left(\frac{x+r}{x+1}\right)(1-p).$$

Comparing this ratio to one implies the lemma. □

The next two lemmas prove a useful property of the constant c_k as defined by (9.2).

Lemma 9.3. $\lambda - c_k \mathbb{P}_k(W' = W+1 \,|\, \{X_i\}) \geqslant 0.$

Proof.

$$c_k \mathbb{P}_k(W' = W+1 \,|\, \{X_i\}) = \lambda \left(\frac{a_k}{k}\right)\binom{r}{k}^{-1} \sum_{\substack{\{i_1,\dots,i_k\} \\ \subseteq\{1,\dots,r\}}} \frac{\binom{k+\sum_j X_{i_j}}{k-1}(1-p)^{(\sum_j X_{i_j}+1)} p^k}{\binom{k+a_k-2}{k-1}(1-p)^{a_k} p^{k-1}}.$$

Define b to be the maximum of one and the integer part of $(k-1)(1-p)/p$ and note that $b=1$ implies $a_k = 1$. Lemma 9.2 implies

$$\begin{aligned} c_k \mathbb{P}_k(W' = W+1 \,|\, \{X_i\}) &\leqslant \lambda \left(\frac{a_k}{k}\right)\binom{r}{k}^{-1} \sum_{\substack{\{i_1,\dots,i_k\} \\ \subseteq\{1,\dots,r\}}} \frac{\binom{k+b-1}{k-1}(1-p)^b p^k}{\binom{k+a_k-2}{k-1}(1-p)^{a_k} p^{k-1}} \\ &= \lambda \left(\frac{(k+b-1)p}{k}\right) \frac{\binom{k+b-2}{k-2}(1-p)^b}{\binom{k+a_k-2}{k-2}(1-p)^{a_k}} \leqslant \lambda. \end{aligned}$$

The final inequality follows from the definition of b and by applying Lemma 9.2 with $r = k-1$. □

Lemma 9.4. $W - c_k \mathbb{P}_k(W' = W-1 \,|\, \{X_i\}) \geqslant 0.$

https://doi.org/10.1017/S096354830999023X Published online by Cambridge University Press

Proof. The cases $W = 0$ or $k = 1$ are simple to verify, so assume otherwise. By Lemma 9.1 and the definition of c_k,

$$\begin{aligned} &c_k \mathbb{P}_k(W' = W - 1 \,|\, \{X_i\}) \\ &\quad = \sum_{\substack{\{i_1,\dots,i_k\} \\ \subseteq\{1,\dots,r\}}} \frac{\left(\sum_j X_{i_j}\right)}{\binom{r-1}{k-1}} \frac{\binom{k+\sum_j X_{i_j}-2}{k-2}(1-p)^{\left(\sum_j X_{i_j}\right)} p^{k-1}}{\binom{k+a_k-2}{k-2}(1-p)^{a_k} p^{k-1}} \\ &\quad \leqslant \binom{r-1}{k-1}^{-1} \sum_{\substack{\{i_1,\dots,i_k\} \\ \subseteq\{1,\dots,r\}}} \left(\sum_j X_{i_j}\right) = \sum_{i=1}^{r} X_i = W. \end{aligned}$$

An application of Lemma 9.2 with $r = k - 1$ implies the inequality. □

The previous two lemmas show that for $c = c_k$ defined by (9.2), the terms within the absolute values in (7.1) are positive. Also note that

$$c_k \mathbb{P}_k(W' = W - 1 \,|\, X_1 = a_k, X_i = 0; i > 1) = a_k = W \neq 0,$$

so that among constants satisfying Lemmas 9.3 and 9.4, the error from Theorem 7.1 is minimized for each k when c_k is defined as (9.2).

To apply the theorem, we need to take the expected value of the term in Lemma 9.1. The next lemma has a nice expression for the expectation we need.

Lemma 9.5. *If Y is a random variable distributed as negative binomial with parameters p and k, then*

$$\mathbb{E}\left[\binom{k+Y}{k-1}(1-p)^Y\right] = \frac{\sum_{l=0}^{k-1} \binom{k}{l+1}\binom{k+l-1}{l}\left(\frac{(1-p)^2}{p(2-p)}\right)^l}{(2-p)^k}.$$

Proof. By the definition of expected value,

$$\begin{aligned} \mathbb{E}\left[\binom{k+Y}{k-1}(1-p)^Y\right] &= \sum_{i\geqslant 0}(1-p)^i\binom{k+i}{k-1}(1-p)^i\binom{k+i-1}{k-1}p^k \\ &= \frac{\sum_i \binom{k+i}{k-1}\binom{k+i-1}{k-1}((1-p)^2)^i(p(2-p))^k}{(2-p)^k}. \end{aligned} \tag{9.3}$$

If Z is a random variable distributed as negative binomial with parameters and $q = p(2-p)$ and k, then (9.3) can be written as $\mathbb{E}\left[\binom{k+Z}{k-1}\right]/(2-p)^k$. Using the fact that Z is the sum of independent geometric random variables, we have

$$\mathbb{E}(s^{k+Z}) = s^k\left(\frac{q}{1-(1-q)s}\right)^k = q^k s^k\left(\frac{1}{1-(1-q)s}\right)^k.$$

https://doi.org/10.1017/S096354830999023X Published online by Cambridge University Press

Taking $k-1$ derivatives with respect to s and dividing by $(k-1)!$ implies

$$\mathbb{E}\left[\binom{k+Z}{k-1}s^{Z+1}\right] = \frac{q^k}{(k-1)!}\sum_{l=0}^{k-1}\binom{k-1}{l}\frac{k!}{(l+1)!}s^{l+1}(1-q)^l\frac{(k+l-1)!}{(k-1)!}\left(\frac{1}{1-(1-q)s}\right)^{k+l}. \quad (9.4)$$

Finally, substituting $s=1$ into (9.4) implies the lemma. □

The final results of this section will be stated in two cases. The first case will pertain to 'small' values of k where $(k-1)(1-p)/p<1$, and the 'large' case to all other values of k. For fixed k and λ, the small case is in some sense the typical case as p should be near one in order for W to be approximately Poisson. In this case, there is no need for further restrictions on the value of p in order to prove results analogous to the previous section. However, in the large case additional assumptions will be made. We will first state and prove results for the small case, then discuss the additional assumptions and prove results for the large case.

Lemma 9.6. *For* $(k-1)(1-p)/p<1$, *both*

$$\mathbb{E}[c_k\mathbb{P}_k(W'=W+1\,|\,\{X_i\})] \quad \textit{and} \quad \mathbb{E}[c_k\mathbb{P}_k(W'=W-1\,|\,\{X_i\})]$$

are bounded above by

$$\frac{r(1-p)}{2-p}.$$

Proof. For $k=1,2$, the lemma is easy to verify, so assume $k\geqslant 3$. Let Y be a random variable distributed as negative binomial with parameters p and k. Then we have

$$\begin{aligned} c_k\mathbb{E}[\mathbb{P}_k(W'=W+1\,|\,\{X_i\})] &= c_k(1-p)p^k\mathbb{E}\left[\tbinom{k+Y}{k-1}(1-p)^Y\right] \\ &= c_k(1-p)p^k\frac{\sum_{l=0}^{k-1}\binom{k}{l+1}\binom{k+l-1}{l}\left(\frac{(1-p)^2}{p(2-p)}\right)^l}{(2-p)^k} \\ &= \tfrac{r(1-p)}{2-p}\left(\tfrac{1}{k(2-p)^{k-1}}\right)\textstyle\sum_{l=0}^{k-1}\binom{k}{l+1}\left(\frac{1-p}{p}\right)^l\binom{k+l-1}{k-1}\left(\frac{1-p}{2-p}\right)^l. \end{aligned} \quad (9.5)$$

Noting first that $(k-1)(1-p)<1$, an application of Lemma 9.2 implies that $\binom{k+l-1}{k-1}\left(\frac{1-p}{2-p}\right)^l$ is at most one, yielding the following inequality:

$$\begin{aligned} c_k\mathbb{E}[\mathbb{P}_k(W'=W+1\,|\,\{X_i\})] &\leqslant \frac{r(1-p)}{2-p}\left(\frac{1}{k(2-p)^{k-1}}\right)\sum_{l=0}^{k-1}\binom{k}{l+1}\left(\frac{1-p}{p}\right)^l \\ &= \left(\frac{r(1-p)}{2-p}\right)\left(\frac{1-p^k}{k(1-p)(p(2-p))^{k-1}}\right). \end{aligned}$$

https://doi.org/10.1017/S096354830999023X Published online by Cambridge University Press

From the previous lines, it is enough to show that, for all $3 \leqslant k \leqslant r$, the following term is at most one:

$$\frac{1-p^k}{k(1-p)(p\,(2-p))^{k-1}}. \tag{9.6}$$

The difference of (9.6) applied at $k+1$ and k is positively proportional to

$$k\left(\sum_{i=0}^{k} p^i\right) - (k+1)p\,(2-p)\left(\sum_{i=0}^{k-1} p^i\right). \tag{9.7}$$

We will show that this difference is at most zero, which implies that (9.6) is decreasing in k, so it is enough to show the lemma holds in the case where $k=3$. Note that $(k-1)(1-p)/p < 1$ implies $k < 1/(1-p)$, so

$$\begin{aligned} k\left(\sum_{i=0}^{k} p^i\right) - (k+1)p\,(2-p)\left(\sum_{i=0}^{k-1} p^i\right) &= \left(\sum_{i=0}^{k-1} p^i\right)(k(1-p)^2 - p\,(2-p)) + kp^k \\ &< \left(\sum_{i=0}^{k-1} p^i\right)(1-3p+p^2) + kp^k. \end{aligned} \tag{9.8}$$

The small-k condition for $k \geqslant 3$ implies, in particular, that $2/3 < p < 1$, so $1-3p+p^2 < 0$, which, starting from (9.8), yields

$$\begin{aligned} &k\left(\sum_{i=0}^{k} p^i\right) - (k+1)p\,(2-p)\left(\sum_{i=0}^{k-1} p^i\right) \\ &\quad< (1-3p+p^2) + (1-3p+p^2)(k-1)p^{k-1} + kp^k \\ &\quad= (1-3p+p^2) + p^{k-1}(k(1-p)^2 - (1-3p+p^2)) \\ &\quad\leqslant 1-3p+p^2+p^k(2-p) \leqslant 1-3p+p^2+2p^3-p^4. \end{aligned} \tag{9.9}$$

The penultimate inequality follows from the fact noted above that $k < 1/(1-p)$, and the final inequality since $k \geqslant 3$. From this point it is a straightforward calculus exercise to show that the final term in (9.9) is negative for $2/3 < p < 1$.

For the remaining term, exchangeability implies

$$\mathbb{E}[c_k \mathbb{P}_k(W' = W-1 \,|\, \{X_i\})] = \mathbb{E}[c_k \mathbb{P}_k(W' = W+1 \,|\, \{X_i\})], \tag{9.10}$$

which proves the lemma. □

In order to prove a result analogous to Lemma 9.6 for the case of $k \geqslant 1/(1-p)$, the value of p will be restricted. Ideally, the values of p under consideration should coincide with the values of p where the Poisson distribution is a good approximation to the negative binomial distribution. First note that $\mathrm{Var}(W) = \lambda + \lambda^2/r$, so it is not unreasonable to assume that $\lambda^2 \leqslant r$. We will instead use a stronger restriction: for the remainder of the section take $3\lambda^2 e^{\lambda} \leqslant r$. This may seem like a demanding constraint, but from (9.1) it is clear that in order for the negative binomial distribution to resemble a Poisson distribution, e^{λ} should be close to $(1+\lambda/r)^r$. The next lemma shows that the assumption on r is not unreasonable given the previous statement; it can be proved by standard analysis using the Taylor expansion of the appropriate functions.

https://doi.org/10.1017/S096354830999023X Published online by Cambridge University Press

Lemma 9.7. *For $r > \lambda^2 > 0$,*

$$\left(\frac{e^\lambda \lambda^2}{7r}\right) \leqslant e^\lambda - \left(1 + \frac{\lambda}{r}\right)^r \leqslant \left(\frac{e^\lambda \lambda^2}{2r}\right).$$

Remark. It has been shown [10] that $\|W - \mathrm{Poi}_\lambda\|_{TV} \leqslant \lambda/r$, which implies that for some values of p and r the restriction $3\lambda^2 e^\lambda \leqslant r$ is an overly demanding constraint. It is an interesting problem to consider what is the minimum constraint that will yield results analogous to Section 8 and how it relates to the proximity of the negative binomial to the Poisson distribution.

Lemma 9.8. *For $3\lambda^2 e^\lambda \leqslant r$ and $k \geqslant 1/(1-p)$, both of $\mathbb{E}[c_k \mathbb{P}_k(W' = W + 1 \,|\, \{X_i\})]$ and $\mathbb{E}[c_k \mathbb{P}_k(W' = W - 1 \,|\, \{X_i\})]$ are bounded above by*

$$\frac{r(1-p)}{2-p}.$$

Proof. Let Y be a random variable distributed as negative binomial with parameters p and k. Then, continuing from (9.5), for $k \geqslant 2$,

$$\begin{aligned} &c_k \mathbb{E}[\mathbb{P}_k(W' = W + 1 \,|\, \{X_i\})] \\ &\quad = \frac{r(1-p)}{2-p}\left(\frac{(k-1)(1-p)}{k(2-p)^{k-1}}\right)\sum_{l=0}^{k-1}\binom{k}{l+1}\left(\frac{1-p}{p(2-p)}\right)^l \frac{\binom{k+l-1}{k-1}(1-p)^l}{\binom{k+a_k-2}{k-2}(1-p)^{a_k}}. \end{aligned}$$

Using the definition of a_k, Lemma 9.2, and an argument similar to the use of the constant b in the proof of Lemma 9.3 (using the fact that $k \geqslant 1/(1-p)$), we obtain an upper bound of $1/p$ on the appropriate fraction in each summand, which yields the following inequality:

$$\begin{aligned} &c_k \mathbb{E}[\mathbb{P}_k(W' = W + 1 \,|\, \{X_i\})] \\ &\quad \leqslant \frac{r(1-p)}{2-p}\left(\frac{(k-1)}{k(2-p)^{k-2}}\right)\sum_{l=0}^{k-1}\binom{k}{l+1}\left(\frac{1-p}{p(2-p)}\right)^{l+1} \\ &\quad = \left(\frac{r(1-p)}{2-p}\right)\left(\frac{(k-1)}{k(2-p)^{k-2}}\right)\left(\left(\frac{(1-p)}{(2-p)p}+1\right)^k - 1\right). \end{aligned}$$

From the previous lines, it is enough to show that, for all $2 \leqslant k \leqslant r$,

$$\left(\frac{(k-1)}{k(2-p)^{k-2}}\right)\left(\left(\frac{(1-p)}{(2-p)p}+1\right)^k - 1\right) \leqslant 1. \tag{9.11}$$

The ratio of successive terms is equal to

$$\left(\frac{k^2}{k^2-1}\right)\frac{\left(\left(\frac{1+p-p^2}{p(2-p)^2}\right)\left(\frac{1+p-p^2}{p(2-p)}\right)^k - (2-p)^{-1}\right)}{\left(\frac{1+p-p^2}{p(2-p)}\right)^k - 1},$$

https://doi.org/10.1017/S096354830999023X Published online by Cambridge University Press

which is at least one. Thus it is enough to show the inequality (9.11) for $k = r$. Now, substituting $p = r/(\lambda + r)$ into (9.11) yields

$$\left(\frac{r-1}{r}\right)\left(1-\frac{\lambda}{2\lambda+r}\right)^{r-2}\left(\left(1+\frac{\lambda(\lambda+r)}{r(2\lambda+r)}\right)^{r}-1\right) \leqslant e^{-\lambda}\left(\frac{2\lambda+r}{\lambda+r}\right)^{2\lambda+2}\left(e^{\lambda}-1\right). \tag{9.12}$$

For the inequality we use the fact that $(1 + x/n)^n \leqslant (1 + x/(n + 1))^{n+1} \leqslant e^x$ if $n + x$ is positive and $n \geqslant 1$. The term (9.12) is clearly decreasing in r; by using the restriction on the value of r and then the inequality $\log(1 + x) \leqslant x$, we have

$$\begin{aligned}\left(\frac{r-1}{r}\right)\left(1-\frac{\lambda}{2\lambda+r}\right)^{r-2}&\left(\left(1+\frac{\lambda(\lambda+r)}{r(2\lambda+r)}\right)^{r}-1\right)\\ &\leqslant \exp\left\{(2\lambda+2)\log\left(1+\frac{1}{1+3\lambda e^{\lambda}}\right)\right\}\left(1-e^{-\lambda}\right)\\ &\leqslant \exp\left\{\frac{2\lambda+2}{1+3\lambda e^{\lambda}}\right\}\left(1-e^{-\lambda}\right).\end{aligned} \tag{9.13}$$

Taking the natural logarithm of (9.13), we have

$$\frac{2\lambda+2}{1+3\lambda e^{\lambda}}+\log\left(1-e^{-\lambda}\right)=\frac{2\lambda+2}{1+3\lambda e^{\lambda}}-\sum_{i\geqslant 1}\frac{e^{-i\lambda}}{i}. \tag{9.14}$$

The final expression is smaller than any partial sum, and it is easy to see that, by only taking one term in the sum, (9.14) is negative for $\lambda \geqslant 2$, and taking three terms yields the proper inequality for $\lambda \geqslant 1$.

For the remaining term, the equation (9.10) continues to hold in this case, which proves the lemma. □

Theorem 9.9. *For the values of c_k defined previously, and the set of k where either $3\lambda^2 e^{\lambda} \leqslant r$ and $k \geqslant 1/(1-p)$ or $k < 1/(1-p)$, the error term from Theorem 7.1 is minimized for $k = 1$.*

Proof. This follows directly from Lemmas 9.6 and 9.8 and the easily verified fact

$$\mathbb{E}[c_1\mathbb{P}_1(W' = W+1\,|\,\{X_i\})] = \mathbb{E}[c_1\mathbb{P}_1(W' = W-1\,|\,\{X_i\})] = \frac{r(1-p)}{2-p}.$$ □

Acknowledgements

The author thanks Jason Fulman for the suggestion to write on this topic, and an anonymous referee for detailed comments which greatly improved this work.

References

[1] Arratia, R., Goldstein, L. and Gordon, L. (1989) Two moments suffice for Poisson approximations: The Chen–Stein method. *Ann. Probab.* **17** 9–25.

https://doi.org/10.1017/S096354830999023X Published online by Cambridge University Press

[2] Arratia, R., Goldstein, L. and Gordon, L. (1990) Poisson approximation and the Chen–Stein method. *Statist. Sci.* **5** 403–434.
[3] Askey, R. (1975) *Orthogonal Polynomials and Special Functions*, SIAM, Philadelphia, PA.
[4] Baldi, P., Rinott, Y. and Stein, C. (1989) A normal approximation for the number of local maxima of a random function on a graph. In *Probability, Statistics, and Mathematics*, Academic Press, Boston, MA, pp. 59–81.
[5] Bannai, E. and Ito, T. (1984) *Algebraic Combinatorics I: Association Schemes*, Benjamin/Cummings, Menlo Park, CA.
[6] Barbour, A. D. and Chen, L. H. Y., eds (2005) *An Introduction to Stein's Method*, Vol. 4 of *Lecture Notes Series, Institute for Mathematical Sciences, National University of Singapore*, Singapore University Press.
[7] Barbour, A. D., Holst, L. and Janson, S. (1992) *Poisson Approximation*, Vol. 2 of *Oxford Studies in Probability*, Clarendon Press, New York.
[8] Bolthausen, E. (1984) An estimate of the remainder in a combinatorial central limit theorem. *Z. Wahrsch. Verw. Gebiete* **66** 379–386.
[9] Brémaud, P. (1999) *Markov Chains: Gibbs Fields, Monte Carlo Simulation, and Queues*, Vol. 31 of *Texts in Applied Mathematics*. Springer, New York.
[10] Brown, T. C. and Phillips, M. J. (1999) Negative binomial approximation with Stein's method. *Method. Comput. Appl. Probab.* **1** 407–421.
[11] Chatterjee, S. (2008) A new method of normal approximation. *Ann. Probab.* **36** 1584–1610.
[12] Chatterjee, S. (2009) Fluctuations of eigenvalues and second order Poincaré inequalities. *Probab. Theory Rel. Fields* **143** 1–40.
[13] Chatterjee, S., Diaconis, P. and Meckes, E. (2005) Exchangeable pairs and Poisson approximation. *Probab. Surv.* **2** 64–106 (electronic).
[14] Chen, L. H. Y. and Shao, Q.-M. (2001) A non-uniform Berry–Esseen bound via Stein's method. *Probab. Theory Rel. Fields* **120** 236–254.
[15] Chen, L. H. Y. and Shao, Q.-M. (2005) Stein's method for normal approximation. In *An Introduction to Stein's Method*, Singapore University Press, pp. 1–59.
[16] Diaconis, P. and Greene, C. (1989) Applications of Murphy's elements. Technical Report 335, Stanford University.
[17] Diaconis, P. and Shahshahani, M. (1981) Generating a random permutation with random transpositions. *Z. Wahrsch. Verw. Gebiete* **57** 159–179.
[18] Eagleson, G. K. (1969) A characterization theorem for positive definite sequences on the Krawtchouk polynomials. *Austral. J. Statist.* **11** 29–38.
[19] Fulman, J. (2004) Stein's method and non-reversible Markov chains. In *Stein's Method: Expository Lectures and Applications*, Vol. 46 of *IMS Lecture Notes Monograph Series*, Institute of Mathematical Statistics, pp. 69–77.
[20] Fulman, J. (2005) Stein's method and Plancherel measure of the symmetric group. *Trans. Amer. Math. Soc.* **357** 555–570 (electronic).
[21] Fulman, J. (2008) Convergence rates of random walk on irreducible representations of finite groups. *J. Theoret. Probab.* **21** 193–211.
[22] Fulman, J. (2008) Stein's method and random character ratios. *Trans. Amer. Math. Soc.* **360** 3687–3730.
[23] Goldstein, L. (2005) Berry–Esseen bounds for combinatorial central limit theorems and pattern occurrences, using zero and size biasing. *J. Appl. Probab.* **42** 661–683.
[24] Goldstein, L. and Reinert, G. (1997) Stein's method and the zero bias transformation with application to simple random sampling. *Ann. Appl. Probab.* **7** 935–952.
[25] Goldstein, L. and Rinott, Y. (1996) Multivariate normal approximations by Stein's method and size bias couplings. *J. Appl. Probab.* **33** 1–17.
[26] Ingram, R. E. (1950) Some characters of the symmetric group. *Proc. Amer. Math. Soc.* **1** 358–369.

https://doi.org/10.1017/S096354830999023X Published online by Cambridge University Press

[27] Johnson, N. L., Kemp, A. W. and Kotz, S. (2005) *Univariate Discrete Distributions*, 3rd edn, Wiley Series in Probability and Statistics, Wiley-Interscience, Hoboken, NJ.
[28] MacWilliams, F. J. and Sloane, N. J. A. (1977) *The Theory of Error-Correcting Codes I*, Vol. 16 of *North-Holland Mathematical Library*, North-Holland, Amsterdam.
[29] Reinert, G. (1998) Couplings for normal approximations with Stein's method. In *Microsurveys in Discrete Probability* (Princeton 1997), Vol. 41 of *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, AMS, Providence, RI, pp. 193–207.
[30] Rinott, Y. and Rotar, V. (1997) On coupling constructions and rates in the CLT for dependent summands with applications to the antivoter model and weighted U-statistics. *Ann. Appl. Probab.* **7** 1080–1105.
[31] Rinott, Y. and Rotar, V. (2000) Normal approximations by Stein's method. *Decis. Econ. Finance* **23** 15–29.
[32] Röllin, A. (2007) Translated Poisson approximation using exchangeable pair couplings. *Ann. Appl. Probab.* **17** 1596–1614.
[33] Röllin, A. (2008) A note on the exchangeability condition in Stein's method. *Statist. Probab. Lett.* **78** (13) 1800–1806.
[34] Sagan, B. E. (2001) *The Symmetric Group: Representations, Combinatorial Algorithms, and Symmetric Functions*, 2nd edn, Vol. 203 of *Graduate Texts in Mathematics*, Springer, New York,
[35] Seneta, E. (2001) Characterization by orthogonal polynomial systems of finite Markov chains. *J. Appl. Probab.* **38A** 42–52.
[36] Stein, C. (1986) *Approximate Computation of Expectations*, Vol. 7 of *IMS Lecture Notes Monograph Series*, Institute of Mathematical Statistics, Hayward, CA.

https://doi.org/10.1017/S096354830999023X Published online by Cambridge University Press