Convergence Rates for Generalized Descents

John Pike

Department of Mathematics
University of Southern California, U.S.A.

jpike@usc.edu

Abstract

d-descents are permutation statistics that generalize the notions of descents and inversions. It is known that the distribution of d-descents of permutations of length n satisfies a central limit theorem as n goes to infinity. We provide an explicit formula for the mean and variance of these statistics and obtain bounds on the rate of convergence using Stein's method.

1 Introduction

For $\pi \in S_n$, the symmetric group on $[n] = \{1, 2, \ldots, n\}$, and $1 \leq d < n$, we say that a pair (i, j) with $i < j \leq i + d$ and $\pi(i) > \pi(j)$ is a d-descent of π and write $\mathrm{Des}_d(\pi) = |\{(i, j) \in [n]^2 : i < j \leq i + d,\ \pi(i) > \pi(j)\}|$, the number of d-descents of π. (Note that $\mathrm{Des}_1(\pi) = \mathrm{Des}(\pi)$ is the number of descents in π and $\mathrm{Des}_{n-1}(\pi) = \mathrm{Inv}(\pi)$ is the number of inversions.) These permutation statistics first appeared in [2] where they were related to the Betti numbers of Hessenberg varieties. In 2008, Miklós Bóna used Janson's criterion to show that the distribution of the number of d-descents of permutations of length n converges to a normal distribution as n goes to infinity [1]. In 2004, Jason Fulman was able to provide convergence rates for the cases $d = 1$ and $d = n - 1$ using Stein's method techniques [6, 10]. In this paper, we carry out analogous computations to get convergence rates for general d-descents both for arbitrary fixed values of d and when d grows with n (excluding a certain exceptional regime). Because of a recent theorem due to Adrian Röllin regarding the necessity of exchangeability conditions in Stein's method [9], we are able to avoid some of the technical arguments used in Fulman's proof. We also improve upon Bóna's formula for the variance of d-descents. Essentially, this paper serves to compile and generalize the results of [1] and [6] and to clarify the underlying arguments for future reference. It also illustrates the utility and applications of Röllin's theorem.

2 Mean and Variance of d-descents

To begin, we define $Y_{n,d}$ to be the random variable on S_n (equipped with uniform probability measure P) given by $Y_{n,d}(\pi) = \mathrm{Des}_d(\pi)$. Observe that for $n \geq 2$, $1 \leq d < n$, the number of $(i,j) \in [n]^2$ with $i < j \leq i + d$ is

$$N_{n,d} = d(n-d) + (d-1) + (d-2) + \cdots + 2 + 1 = \frac{2nd - d^2 + d}{2}.$$

Ordering such pairs lexicographically - that is, $(i,j) \prec (r,s)$ if $i < r$ or $i = r$ and $j < s$ - and indexing them by $\{(i_k, j_k)\}_{k=1}^{N_{n,d}}$, we can write $\mathrm{Des}_d(\pi) = \sum_{k=1}^{N_{n,d}} 1(\pi(i_k) > \pi(j_k))$ where $1(\pi(i) > \pi(j)) = \begin{cases} 1, & \pi(i) > \pi(j) \\ 0, & \pi(i) < \pi(j) \end{cases}$ is the indicator of the event $\{\pi(i) > \pi(j)\}$. Letting $X_{n,d}^k$ be the random variable defined by $X_{n,d}^k(\pi) = 1(\pi(i_k) > \pi(j_k))$ gives $Y_{n,d} = \sum_{k=1}^{N_{n,d}} X_{n,d}^k$. Since $P(\pi(i_k) > \pi(j_k)) = \frac{1}{2}$ for all $(i,j) \in \{(i_k, j_k)\}_{k=1}^{N_{n,d}}$, the $X_{n,d}^k$'s are Bernoulli($\frac{1}{2}$), thus $E[X_{n,d}^k] = \frac{1}{2}$ and $\mathrm{Var}(X_{n,d}^k) = \frac{1}{4}$. Accordingly,

$$E[Y_{n,d}] = \sum_{k=1}^{N_{n,d}} E[X_{n,d}^k] = \frac{N_{n,d}}{2} = \frac{2nd - d^2 + d}{4}$$

and

$$\mathrm{Var}(Y_{n,d}) = \sum_{k=1}^{N_{n,d}} \mathrm{Var}(X_{n,d}^k) + 2 \sum_{1 \leq k < l \leq N_{n,d}} \mathrm{Cov}(X_{n,d}^k, X_{n,d}^l)$$
$$= \frac{N_{n,d}}{4} + 2 \sum_{1 \leq k < l \leq N_{n,d}} \mathrm{Cov}(X_{n,d}^k, X_{n,d}^l).$$

Now when $\{i_k, j_k\} \cap \{i_l, j_l\} = \varnothing$, $X_{n,d}^k$ and $X_{n,d}^l$ are independent, so $\mathrm{Cov}(X_{n,d}^k, X_{n,d}^l) = 0$. As such, our assumptions on the ordering of the indices imply that the only nonzero summands correspond to the cases $i_k = i_l$, $j_k = j_l$, and $j_k = i_l$. Because $\mathrm{Cov}(X_{n,d}^k, X_{n,d}^l) = E[X_{n,d}^k X_{n,d}^l] - E[X_{n,d}^k]E[X_{n,d}^l] = E[X_{n,d}^k X_{n,d}^l] - \frac{1}{4}$, we just have to compute $E[X_{n,d}^k X_{n,d}^l]$ for each of these cases and count the number of ways each case can occur.

For the case $i_k = i = i_l$, we have that $E[X_{n,d}^k X_{n,d}^l] = P(\pi(i) > \pi(j_k), \pi(j_l)) = \frac{1}{3}$ and the number of triplets $(i, j_k, j_l) \in [n]^3$ with $i < j_k < j_l \leq i + d$ is

$$(n-d)\binom{d}{2} + \sum_{t=2}^{d-1} \binom{t}{2} = \frac{(d^2 - d)(3n - 2d - 2)}{6}.$$

If $j_k = j = j_l$, then $E[X_{n,d}^k X_{n,d}^l] = P(\pi(j) < \pi(i_k), \pi(i_l)) = \frac{1}{3}$ and the number of triplets $(i_k, i_l, j) \in [n]^3$ with $i_k < i_l < j \leq i_k + d$ is also the same as in the $i_k = i = i_l$ case.

If $i_k < j_k = m = i_l < j_l$, then $E[X_{n,d}^k X_{n,d}^l] = P(\pi(i) > \pi(m) > \pi(j_l)) = \frac{1}{6}$. Let M denote the number of triplets $(i_k, m, j_l) \in [n]^3$ with $i_k < m \leq i_k + d$ and $m < j_l \leq m + d$.

If $2d \leq n$, then there are d^2 choices for (i_k, j_l) when $d < m \leq n - d$, there are $(m-1)d$ choices for (i_k, j_l) when $m \leq d$, and there are $d(n-m)$ choices for (i_k, j_l) when $m > n - d$. Thus if $2d \leq n$, then

$$M = d^2(n - 2d) + 2d \sum_{t=1}^{d-1} t = nd^2 - d^3 - d^2.$$

For $2d > n$, there are $(m-1)d$ choices for (i_k, j_l) when $m \leq n - d$, there are $d(n-m)$ choices for (i_k, j_l) when $m > d$, and there are $(m-1)(n-m)$ choices for (i_k, j_l) when $n - d < m \leq d$. Thus if $2d > n$, then

$$M = \sum_{m=n-d+1}^{d} (m-1)(n-m) + 2d \sum_{t=1}^{n-d-1} t$$
$$= \frac{6n^2 d - 6nd^2 - 12nd - n^3 + 3n^2 - 2n + 2d^3 + 6d^2 + 4d}{6}.$$

The variance is thus

$$\operatorname{Var}(Y_{n,d}) = \frac{N_{n,d}}{4} + 2 \sum_{1 \leq k < l \leq N_{n,d}} \operatorname{Cov}(X_{n,d}^k, X_{n,d}^l)$$
$$= \frac{N_{n,d}}{4} + 2[2(\frac{1}{3} - \frac{1}{4})\frac{(d^2 - d)(3n - 2d - 2)}{6} + (\frac{1}{6} - \frac{1}{4})M]$$
$$= \frac{1}{72}(12nd^2 + 6nd - 8d^3 - 9d^2 - d - 12M).$$

Substituting the values of M established in the preceding paragraph gives an explicit formula for $\operatorname{Var}(Y_{n,d})$.

To summarize, we have:

Theorem 1. *If $Y_{n,d}$ is the number of d-descents in a random permutation of length n, then*

$$E[Y_{n,d}] = \frac{2nd - d^2 - d}{4}$$

and

$$\operatorname{Var}(Y_{n,d}) = \begin{cases} \frac{6nd + 4d^3 + 3d^3 - d}{72}, & 2d \leq n \\ \frac{2n^3 - 12n^2 d - 6n^2 + 24nd^2 + 30nd + 4n - 12d^3 - 21d^2 - 9d}{72}, & 2d > n \end{cases}.$$

- Observe that $d = 1$ yields $\text{Var}(\text{Des}) = \frac{n+1}{12}$ and $d = n - 1$ gives $\text{Var}(\text{Inv}) = \frac{n(n-1)(2n+5)}{72}$.

- The value of the variance in the $2d \leq n$ case is slightly different than that reported in [1] because a 2 was omitted when recording the number of pairs in 2.2.3. This minor error does not affect the validity of the subsequent results.

- The author has been unable to find another instance of a general formula for the variance of d-descents in the literature, though as mentioned above, Miklós Bóna worked out the $2d \leq n$ case completely and the preceding calculations follow the same basic reasoning. The general idea for calculating the variance can also be found in Lemma 4.3.1 in [6], but the calculations are only carried out for $d = 1$ and $d = n - 1$.

3 Convergence Rates

This section is basically a generalization of Fulman's derivation of the rate of convergence for descents and inversions [6]. The crux of his proof relies upon the following theorem due to Rinott and Rotar [8].

Theorem 2 (Rinott and Rotar). *Let W, W' be an exchangeable pair of real-valued random variables such that $E[W'|W] = (1 - \lambda)W$ for some $\lambda \in (0, 1)$. If there exists a constant A such that $|W' - W| \leq A$ almost surely, then for all $x \in \mathbb{R}$,*

$$|P(W \leq x) - \Phi(x)| \leq \frac{12}{\lambda}\sqrt{\text{Var}(E[(W' - W)^2|W]} + 48\frac{A^3}{\lambda} + 8\frac{A^2}{\sqrt{\lambda}}$$

where $\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt$ is the standard normal c.d.f.

A substantial portion of Fulman's argument involves proving that the pair (W, W') he constructed is indeed exchangeable. The derivation in this paper uses essentially the same pair of random variables, but we are spared the onus of establishing exchangeability thanks to the following adaptation of Rinott and Rotar's result discovered by Adrian Rllin [9].

Theorem 3 (Rllin). *Suppose that W and W' are a pair of real-valued random variables having common law such that $E[W] = 0$, $\text{Var}(W) = 1$, and $E[W'|W] = (1 - \lambda)W$ for some $\lambda \in (0, 1)$. Suppose moreover that there is a constant A such that $|W' - W| \leq A$ almost surely. Then for all $x \in \mathbb{R}$,*

$$|P(W \leq x) - \Phi(x)| \leq \frac{12}{\lambda}\sqrt{\text{Var}(E[(W' - W)^2|W]} + 32\frac{A^3}{\lambda} + 6\frac{A^2}{\sqrt{\lambda}}$$

where Φ is the standard normal c.d.f.

- The original version of Rllin's theorem is a little more general. The above is an immediate corollary which is sufficient for our purposes.

To facilitate the ensuing arguments, we define a modified version of $Y_{n,d}$ as follows: For each $n \geq 2$, $1 \leq d < n$, consider the real, skew-symmetric $n \times n$ matrix $M(n,d) = [M_{i,j}(n,d)]_{i,j=1}^{n}$ given by

$$M_{i,j}(n,d) = \begin{cases} -1, & i < j \leq i + d \\ 1, & j < i \leq i + d \\ 0, & \text{otherwise} \end{cases}$$

and define the random variable $Z_{n,d}$ by $Z_{n,d}(\pi) = \sum_{i<j} M_{\pi(i),\pi(j)}(n,d)$. (For notational convenience, we will often abbreviate $M_{i,j}(n,d) = M_{i,j}$ when there is no danger of confusion.) Also, define $\mathrm{Asc}_d(\pi) = |\{(i,j) \in [n]^2 : i < j \leq i+d,\ \pi(i) < \pi(j)\}|$ to be the number of d-ascents of π so that $N_{n,d} = \mathrm{Des}_d(\pi) + \mathrm{Asc}_d(\pi)$.

Then

$$Z_{n,d}(\pi^{-1}) = \sum_{i<j} M_{\pi^{-1}(i),\pi^{-1}(j)}(n,d)$$
$$= \left| \{(i,j) : i < j,\ \pi^{-1}(j) < \pi^{-1}(i) \leq \pi^{-1}(j) + d\} \right|$$
$$- \left| \{(i,j) : i < j,\ \pi^{-1}(i) < \pi^{-1}(j) \leq \pi^{-1}(i) + d\} \right|.$$

Taking $r = \pi^{-1}(i)$, $s = \pi^{-1}(j)$ in the first term and $s = \pi^{-1}(i)$, $r = \pi^{-1}(j)$ in the second term, we see that

$$Z_{n,d}(\pi^{-1}) = |\{(r,s) : \pi(r) < \pi(s),\ s < r \leq s + d\}|$$
$$- |\{(r,s) : \pi(s) < \pi(r),\ s < r \leq s + d\}|$$
$$= \mathrm{Des}_d(\pi) - \mathrm{Asc}_d(\pi) = 2\mathrm{Des}_d(\pi) - N_{n,d},$$

hence $Z_{n,d}(\pi) = 2\mathrm{Des}_d(\pi^{-1}) - \frac{2nd - d^2 - d}{2}$.

Writing $\tilde{Y}_{n,d}(\pi) = Y_{n,d}(\pi^{-1})$, we have $Z_{n,d} = 2\tilde{Y}_{n,d} - \frac{2nd - d^2 - d}{2}$. Now $\tilde{Y}_{n,d}$ and $Y_{n,d}$ have the same distribution because replacing π with π^{-1} merely amounts to relabeling the sample space. As such, $E[\tilde{Y}_{n,d}] = E[Y_{n,d}]$, so it follows from Theorem 1 and the linearity of expectation that $E[Z_{n,d}] = 2E[Y_{n,d}] - \frac{2nd - d^2 - d}{2} = 0$. Thus if we let

$$W_{n,d} = \frac{Z_{n,d}}{\sqrt{\mathrm{Var}(Z_{n,d})}} = \frac{2\tilde{Y}_{n,d} - \frac{2nd-d^2-d}{2}}{\sqrt{4\mathrm{Var}(\tilde{Y}_{n,d})}} = \frac{\tilde{Y}_{n,d} - \frac{2nd-d^2-d}{4}}{\sqrt{\mathrm{Var}(\tilde{Y}_{n,d})}},$$

then $W_{n,d}$ has mean zero and variance one. Moreover, since $\tilde{Y}_{n,d}$ and $Y_{n,d}$ have the same distribution, $W_{n,d}$ is distributed as $\frac{Y_{n,d} - \frac{2nd-d^2-d}{4}}{\sqrt{\mathrm{Var}(Y_{n,d})}}$.

At this point, we need to construct a complementary random variable W' in order to apply Theorem 3. To this end, let $\sigma_I \in S_n$ be the cycle $\sigma_I = (I, I+1, \ldots, n)$ and define $W'_{n,d}$ by choosing I uniformly from $[n] = \{1, 2, \ldots, n\}$ and setting $W'_{n,d}(\pi) = W_{n,d}(\pi\sigma_I)$ where permutation multiplication is performed from right to left. For example, taking $n = 6$, $I = 4$, if π is given in two-line notation by

$$\pi = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 6 & 1 & 4 & 2 & 5 & 3 \end{pmatrix}, \text{ then } \pi\sigma_I = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 6 & 1 & 4 & 5 & 3 & 2 \end{pmatrix}.$$

Notice that $W_{n,d}$ and $W'_{n,d}$ have the same distribution since right multiplication by σ_I with I chosen uniformly from $[n]$ corresponds to a "random-to-end" shuffle of π, and successive shuffles define a Markov chain with uniform stationary distribution [3]. Thus choosing π uniformly from S_n, choosing I uniformly from $[n]$, and taking the composition $\pi' = \pi\sigma_I$ is equivalent to choosing π' uniformly from S_n. In addition, we have the following lemma.

Lemma 1. $E[W'_{n,d}|W_{n,d}] = (1 - \frac{2}{n})W_{n,d}$

Proof.

$$E[W'_{n,d} - W_{n,d}|\pi] = \frac{1}{n} \sum_{i=1}^{n} [W_{n,d}(\pi\sigma_i) - W_{n,d}(\pi)]$$

$$= \frac{1}{n\sqrt{\text{Var}(Z_{n,d})}} \sum_{i=1}^{n-1} \sum_{j<k} [M_{\pi(\sigma_i(j)),\pi(\sigma_i(k))} - M_{\pi(j),\pi(k)}]$$

(because $\sigma_n = id$). Now $(\sigma_i(j), \sigma_i(k)) = (j, k)$ when $j, k < i$, $(\sigma_i(j), \sigma_i(k)) = (j, k+1)$ when $j < i \le k < n$, $(\sigma_i(j), \sigma_i(k)) = (j, i)$ when $j < i$, $k = n$, $(\sigma_i(j), \sigma_i(k)) = (j+1, k+1)$ when $i \le j < k < n$, and $(\sigma_i(j), \sigma_i(k)) = (j+1, i)$ when $i \le j < n$, $k = n$. Thus, after some careful bookkeeping, we see that for each $1 \le i < n$,

$$\{(\pi(\sigma_i(j)), \pi(\sigma_i(k))) : j < k\} \setminus \{(\pi(j), \pi(k)) : j < k\}$$
$$= \{(\pi(j+1), \pi(i)) : j \ge i\} = \{(\pi(j), \pi(i)) : j > i\}$$

and

$$\{(\pi(j), \pi(k)) : j < k\} \setminus \{(\pi(\sigma_i(j)), \pi(\sigma_i(k))) : j < k\} = \{(\pi(i), \pi(k)) : k > i\},$$

hence

$$E[W'_{n,d} - W_{n,d}|\pi] = \frac{1}{n\sqrt{\text{Var}(Z_{n,d})}} \sum_{i=1}^{n-1} \sum_{j<k} [M_{\pi(\sigma_i(j)),\pi(\sigma_i(k))} - M_{\pi(j),\pi(k)}]$$

$$= \frac{1}{n\sqrt{\text{Var}(Z_{n,d})}} \sum_{i=1}^{n-1} \left(\sum_{i<j} -M_{\pi(i),\pi(j)} - \sum_{i<k} M_{\pi(i),\pi(k)} \right)$$

$$= \frac{1}{n\sqrt{\text{Var}(Z_{n,d})}} \sum_{i=1}^{n-1} \sum_{i<j} -2M_{\pi(i),\pi(j)}$$

$$= -\frac{2}{n}\left(\frac{1}{\sqrt{\text{Var}(Z_{n,d})}} \sum_{i<j} M_{\pi(i),\pi(j)}\right) = -\frac{2}{n}W_{n,d}.$$

As $W_{n,d}$ is a function of π and thus is $\sigma(\pi)$-measurable, we have $E[W'_{n,d}-W_{n,d}|W_{n,d}] = E[E[W'_{n,d}-W_{n,d}|\pi]|W_{n,d}] = E[-\frac{2}{n}W_{n,d}|W_{n,d}] = -\frac{2}{n}W_{n,d}$, hence $E[W'_{n,d}|W_{n,d}] = E[W'_{n,d}-W_{n,d}|W_{n,d}] + W_{n,d} = (1-\frac{2}{n})W_{n,d}$. $\qquad\square$

It is worth remarking that this construction of a complementary random variable by applying some shuffling scheme to the input might be useful in analyzing other permutation statistics. Indeed, the preceding arguments show that for any random variable $X = X(g)$ defined on a finite group G with uniform probability measure, if μ is a probability measure on G that is not concentrated on a coset of a subgroup of G (see [4]), $W = \frac{X-E[X]}{\sqrt{\text{Var}(X)}}$, and W' is defined by $W'(g) = W(hg)$ where h is drawn from μ, then the assumptions of Theorem 3 are satisfied whenever

$$E[W - W'|g] = \sum_{h \in G}[W(g) - W(hg)]\mu(h) = \lambda W$$

for some $\lambda \in (0, 1)$.

More generally, since every Markov chain with finite state space Ω and transition kernel $K(x,y) = P(X_{n+1} = y|X_n = x)$ has a *random mapping representation* - that is, a Λ-valued random variable Z (with distribution μ) and a mapping $f : \Omega \times \Lambda \to \Omega$ such that $K(x,y) = P(f(x,Z) = y)$ - this procedure for finding a complementary random variable applies to all temporally homogeneous, finite state space, ergodic Markov chains by setting $W'(\omega) = W(f(Z,\omega))$ where Z is drawn from μ. (See [7] for a discussion of random mapping representations.)

If W and W' are distributed as consecutive steps in a *reversible* Markov chain in equilibrium, then, in obvious notation, $P(W = x, W' = y) = \pi(x)K(x,y) = \pi(y)K(y,x) = P(W = y, W' = x)$, hence (W, W') is an exchangeable pair. The point here is that Rollin's observations allow one to apply the machinery of Stein's method using a pair of similarly constructed random variables without requiring that the underlying chain is reversible provided that the above conditions are satisfied.

Of course, one still needs to verify that $E[W'|\omega] = (1-\lambda)W$, in order to conclude that $E[W'|W] = E[E[W'|\omega]|W] = E[(1-\lambda)W|W] = (1-\lambda)W$. Because

$$E[W'|\omega] = E[W(f(\omega, Z))] = \sum_{y \in \Omega} W(y)P(f(\omega, Z) = y) = \sum_{y \in \Omega} K(\omega, y)W(y),$$

the condition $E[W'|\omega] = (1-\lambda)W$ implies that W is a (right) eigenfunction of K corresponding to the eigenvalue $1 - \lambda$. Thus this method is applicable precisely when the

statistics in question are eigenfunctions of a Markov chain for which the underlying distribution is stationary.

Lemma 1 and the above observation immediately yield the following corollary

Corollary 1. *For $n > 1$, $d = 1, \ldots, n-1$, the functions $W_{n,d}(\pi)$ are eigenfunctions for the "random-to-end" shuffle corresponding to the eigenvalue $1 - \frac{2}{n}$.*

Lemma 1 says we can take $\lambda = \frac{2}{n}$ in the statement of Theorem 3. A sharp value for the A term in Theorem 3 is given by the following lemma and a corollary gives a simplified version that suffices for the purposes of this paper.

Lemma 2.

$$
\left|W'_{n,d} - W_{n,d}\right| \leq
\begin{cases}
\dfrac{6\sqrt{2}}{\sqrt{(6n-1)d^{-1}+3+4d}}, & 2d \leq n \\[3ex]
\dfrac{6\sqrt{2}}{\sqrt{(2n^3+4n-6n^2)d^{-2}+(30n-12n^2-9)d^{-1}+24n-21-12d}}, & 2d > n
\end{cases}
$$

and this is the best bound possible.

Proof. For any permutation $\pi = \pi_1, \pi_2, \ldots, \pi_n$ and any $I \in [n]$,

$$
|W_{n,d}(\pi\sigma_I) - W_{n,d}(\pi)| = \left| \frac{\mathrm{Des}_d(\sigma_I^{-1}\pi^{-1}) - \frac{2nd-d^2-d}{4}}{\sqrt{\mathrm{Var}(Y_{n,d})}} - \frac{\mathrm{Des}_d(\pi^{-1}) - \frac{2nd-d^2-d}{4}}{\sqrt{\mathrm{Var}(Y_{n,d})}} \right|
$$

$$
= \frac{1}{\sqrt{\mathrm{Var}(Y_{n,d})}} \left| \mathrm{Des}_d(\sigma_I^{-1}\pi^{-1}) - \mathrm{Des}_d(\pi^{-1}) \right|.
$$

Now the number of d-descents in π^{-1} is equal to the number of pairs $(j,k) \in [n]^2$ with $j < k \leq j + d$ such that $\pi^{-1}(j) > \pi^{-1}(k)$. Reindexing $[n]$ by $m \mapsto \pi(m)$, we see that this is equal to the number of pairs (j,k) with $\pi_j < \pi_k \leq \pi_j + d$ such that $j > k$ - that is,

$$
\mathrm{Des}_d(\pi^{-1}) = \sum_{j=1}^{n-1} \left| \{k : \pi_j < \pi_k \leq \pi_j + d, \ k < j\} \right|.
$$

Similarly, the number of d-descents in the inverse of $\pi' = \pi\sigma_I = \pi_1, \ldots, \pi_{I-1}, \pi_{I+1}, \ldots, \pi_n, \pi_I$ is

$$
\mathrm{Des}_d((\pi')^{-1}) = \sum_{j=1}^{n-1} \left| \{k : \pi'_j < \pi'_k \leq \pi'_j + d, \ k < j\} \right|.
$$

Since the relative ordering of the terms in the sequence π' is the same as in π except that the Ith term in π is the nth term in π', we have $\left| \{k : \pi'_j < \pi'_k \leq \pi'_j + d, \ k < j\} \right| = \left| \{k : \pi_j < \pi_k \leq \pi_j + d, \ k < j\} \right|$ whenever $|\pi_j - \pi_I| > d$, $\left| \{k : \pi'_j < \pi'_k \leq \pi'_j + d, \ k < j\} \right| = \left| \{k : \pi_j < \pi_k \leq \pi_j + d, \ k < j\} \right| - 1$ for j such that $\pi_j < \pi_I \leq \pi_j + d$ (of which there are at most d), and $\left| \{k : \pi'_j < \pi'_k \leq \pi'_j + d, \ k < I\} \right| \leq \left| \{k : \pi_j < \pi_k \leq \pi_j + d, \ k < I\} \right| + d$. Thus it follows from the above representations of $\mathrm{Des}_d(\pi^{-1})$ and $\mathrm{Des}_d((\pi')^{-1})$ that

$\left|\operatorname{Des}_d((\pi')^{-1}) - \operatorname{Des}_d(\pi^{-1})\right| \le d$. Moreover, the exact same reasoning shows that when $\pi = id$, $I = 1$, we get $\left|\operatorname{Des}_d((\pi')^{-1}) - \operatorname{Des}_d(\pi^{-1})\right| = d$, so this is the best bound possible. Therefore, we have the tight bound

$$\left|W_{n,d}(\pi\sigma_I) - W_{n,d}(\pi)\right| = \frac{1}{\sqrt{\operatorname{Var}(Y_{n,d})}}\left|\operatorname{Des}_d(\sigma_I^{-1}\pi^{-1}) - \operatorname{Des}_d(\pi^{-1})\right| \le \frac{d}{\operatorname{Var}(Y_{n,d})}$$

and the result follows from Theorem 1. $\qquad\square$

As the bound in Lemma 2 is pretty unwieldy, we record the following corollary.

Corollary 2.

$$\left|W'_{n,d} - W_{n,d}\right| \le \begin{cases} C_1 d^{\frac{1}{2}} n^{-\frac{1}{2}}, & d(n) \le \sqrt{n} \\ C_2 d^{\frac{1}{2}}, & \sqrt{n} < d(n) \le \frac{n}{2} \\ C_3 n^{-\frac{1}{2}}, & d(n) > \frac{n}{2} \\ C(d) n^{-\frac{1}{2}}, & d \text{ fixed} \end{cases}$$

where C_1, C_2, C_3, and $C(d)$ are universal constants independent of n. Moreover, these bounds are of the best possible order.

Proof. When $d \le \frac{n}{2}$, noting that $\frac{6\sqrt{2}}{\sqrt{6nd^{-1}+7d}} \le \frac{6\sqrt{2}}{\sqrt{(6n-1)d^{-1}+3+4d}} \le \frac{3\sqrt{2}}{\sqrt{nd^{-1}+d}}$, we see that if $d = d(n) \le \frac{n}{2}$, we can take $A = C\sqrt{\frac{d}{n+d^2}}$ with $C \ge 3\sqrt{2}$. When $d \le \sqrt{n}$, $\frac{d}{n+d^2} = \Theta(dn^{-1})$, so $\left|W'_{n,d} - W_{n,d}\right| \le C_1 d^{\frac{1}{2}} n^{-\frac{1}{2}}$ for some constant C_1 and this is the best possible order. When $\sqrt{n} < d \le \frac{n}{2}$, $\frac{d}{n+d^2} = \Theta(d^{-1})$, so $\left|W'_{n,d} - W_{n,d}\right| \le C_2 d^{-\frac{1}{2}}$ for some constant C_2 and this is the best possible order. (Note that $A = C\sqrt{\frac{d}{n+d^2}}$ is maximized when $d = \sqrt{n}$, in which case $A = \Theta(n^{-\frac{1}{4}})$.)

Now a little calculus shows that $\operatorname{Var}(Y_{n,d})$ increases monotonically with d for $\frac{n}{2} \le d < n$, so $\frac{n^3}{144} < \frac{2n^3+15n^2-2n}{288} = \operatorname{Var}(Y_{n,d})|_{d=\frac{n}{2}} < \frac{d}{\operatorname{Var}(Y_{n,d})} \le \operatorname{Var}(Y_{n,d})|_{d=n-1} = \frac{2n^3+3n^2-5n}{72} < \frac{n^3}{12}$. Accordingly, for $\frac{n}{2} < d < n$, we have $\sqrt{3}n^{-\frac{1}{2}} = \frac{n\sqrt{12}}{2\sqrt{n^3}} < \frac{d}{\sqrt{Var(Y_{n,d})}} < \frac{n\sqrt{144}}{\sqrt{n^3}} = 12n^{-\frac{1}{2}}$. Therefore, when $\frac{n}{2} < d < n$, $\left|W'_{n,d} - W_{n,d}\right| \le C_3 n^{-\frac{1}{2}}$ for some constant C_3 and this is the best possible order.

Finally, for fixed d, the preceding analysis shows that $\frac{d}{\sqrt{\operatorname{Var}(Y_{n,d})}} = \Theta(n^{-\frac{1}{2}})$, so the best possible bound is $\left|W'_{n,d} - W_{n,d}\right| \le C(d) n^{-\frac{1}{2}}$ for some constant $C(d)$ that does not depend on n. $\qquad\square$

In order to apply Theorem 3, it remains only to bound $\operatorname{Var}(E[(W'_{n,d} - W_{n,d})^2 | W_{n,d}])$. To accomplish this task, we first establish a simplifying lemma which can be found in [6], but whose proof is included for the sake of completeness.

Lemma 3.

$$\operatorname{Var}(E[(W'_{n,d} - W_{n,d})^2 | W_{n,d}]) \le \operatorname{Var}(E[(W'_{n,d} - W_{n,d})^2 | \pi])$$

Proof. The conditional version of Jensen's inequality states that if φ is convex and $E[|X|], E[|\varphi(X)|]$ are finite, then $\varphi(E[X|\mathcal{F}]) \leq E[\varphi(X)|\mathcal{F}]$ (see section 4.1 in [5]). Taking expectations gives

$$E[\varphi(E[X|\mathcal{F}])] \leq E[E[\varphi(X)|\mathcal{F}]] = E[\varphi(X)].$$

Letting $X = E[(W'_{n,d} - W_{n,d})^2|\pi]$, $\varphi(x) = x^2$, and $\mathcal{F} = \sigma(W_{n,d}) \subseteq \sigma(\pi)$, we see that

$$E[E[(W'_{n,d} - W_{n,d})^2|W_{n,d}]^2] = E[E[E[(W'_{n,d} - W_{n,d})^2|\pi]|W_{n,d}]^2] \leq E[E[(W'_{n,d} - W_{n,d})^2|\pi]^2].$$

It follows that

$$\mathrm{Var}(E[(W'_{n,d} - W_{n,d})^2|W_{n,d}]) = E[E[(W'_{n,d} - W_{n,d})^2|W_{n,d}]^2] - E[(W'_{n,d} - W_{n,d})^2]^2$$
$$\leq E[E[(W'_{n,d} - W_{n,d})^2|\pi]^2] - E[(W'_{n,d} - W_{n,d})^2]^2 = \mathrm{Var}(E[(W'_{n,d} - W_{n,d})^2|\pi]).$$

$\square$

The above lemma and some simplifying observations imply:

Lemma 4.

$$\mathrm{Var}(E[(W'_{n,d} - W_{n,d})^2|W_{n,d}]) \leq \begin{cases} K_1 d^2 n^{-3}, & d(n) \leq \sqrt{n} \\ K_2 d^{-2} n^{-1}, & \sqrt{n} < d(n) \leq \frac{n}{2} \\ K_3 n^{-3}, & d(n) > \frac{n}{2} \\ K(d) n^{-3}, & d \text{ fixed} \end{cases}$$

where K_1, K_2, K_3, and $K(d)$ are universal constants which do not depend on n.

Proof. We see from the proof of Lemma 1 that

$$E[(W'_{n,d} - W_{n,d})^2|\pi] = \frac{1}{n} \sum_{i=1}^{n} [W_{n,d}(\pi\sigma_i) - W_{n,d}(\pi)]^2$$

$$= \frac{1}{\mathrm{Var}(Z_{n,d})} \frac{1}{n} \sum_{i=1}^{n-1} \left(\sum_{i<j} -2M_{\pi(i),\pi(j)} \right)^2$$

$$= \frac{1}{\mathrm{Var}(Y_{n,d})} \frac{1}{n} \sum_{i=1}^{n-1} \left(\sum_{i<j} M_{\pi(i),\pi(j)} \right)^2$$

$$= \frac{1}{n\mathrm{Var}(Y_{n,d})} \left(\sum_{i=1}^{n-1} \sum_{i<j} M^2_{\pi(i),\pi(j)} + 2 \sum_{i=1}^{n-1} \sum_{i<j<k} M_{\pi(i),\pi(j)} M_{\pi(i),\pi(k)} \right)$$

$$= \frac{1}{n\mathrm{Var}(Y_{n,d})} \left(\sum_{i<j} M^2_{\pi(i),\pi(j)} + 2 \sum_{i<j<k} M_{\pi(i),\pi(j)} M_{\pi(i),\pi(k)} \right).$$

The reasoning used in establishing the distribution of $Z_{n,d}$ shows that

$$\sum_{i<j} M^2_{\pi(i),\pi(j)} = \left|\{(i,j) \in [n]^2 : i < j, |\pi(i) - \pi(j)| \le d\}\right| = N_{n,d}$$

for all $\pi \in S_n$, so, writing $\sigma_{n,d} = \sqrt{\mathrm{Var}(Y_{n,d})}$, it follows from Lemma 3 that

$$\mathrm{Var}(E[(W'_{n,d} - W_{n,d})^2 | W_{n,d}]) \le \mathrm{Var}(E[(W'_{n,d} - W_{n,d})^2 | \pi])$$

$$= \mathrm{Var}\left(\frac{1}{n\sigma^2_{n,d}}\left(\sum_{i<j} M^2_{\pi(i),\pi(j)} + 2\sum_{i<j<k} M_{\pi(i),\pi(j)} M_{\pi(i),\pi(k)}\right)\right)$$

$$= \frac{4}{n^2\sigma^4_{n,d}}\mathrm{Var}\left(\sum_{i<j<k} M_{\pi(i),\pi(j)} M_{\pi(i),\pi(k)}\right)$$

$$= \frac{4}{n^2\sigma^4_{n,d}}\sum_{\substack{i<j_1<k_1 \\ i_2<j_2<k_2}} \mathrm{Cov}(M_{\pi(i_1),\pi(j_1)} M_{\pi(i_1),\pi(k_1)}, M_{\pi(i_2),\pi(j_2)} M_{\pi(i_2),\pi(k_2)}).$$

Now for all $(i_1, j_1), (i_2, j_2)$, $M_{\pi(i_1),\pi(j_1)} M_{\pi(i_1),\pi(k_1)}$ and $M_{\pi(i_2),\pi(j_2)} M_{\pi(i_2),\pi(k_2)}$ have common law, so $E[M_{\pi(i_1),\pi(j_1)} M_{\pi(i_1),\pi(k_1)}] E[M_{\pi(i_2),\pi(j_2)} M_{\pi(i_2),\pi(k_2)}] > 0$, hence

$$\mathrm{Cov}(M_{\pi(i_1),\pi(j_1)} M_{\pi(i_1),\pi(k_1)}, M_{\pi(i_2),\pi(j_2)} M_{\pi(i_2),\pi(k_2)}$$
$$\le E[M_{\pi(i_1),\pi(j_1)} M_{\pi(i_1),\pi(k_1)} M_{\pi(i_2),\pi(j_2)} M_{\pi(i_2),\pi(k_2)}]$$
$$\le P(M_{\pi(i_1),\pi(j_1)} M_{\pi(i_1),\pi(k_1)} M_{\pi(i_2),\pi(j_2)} M_{\pi(i_2),\pi(k_2)} \ne 0).$$

Since $M_{\pi(i_1),\pi(j_1)} M_{\pi(i_1),\pi(k_1)}$ and $M_{\pi(i_2),\pi(j_2)} M_{\pi(i_2),\pi(k_2)}$ are independent when $\{i_1, j_1, k_1\} \cap \{i_2, j_2, k_2\} = \varnothing$, it follows that $\mathrm{Var}(E[(W'_{n,d} - W_{n,d})^2 | W_{n,d}])$ is bounded above by

$$\frac{4}{n^2\sigma^4_{n,d}}\sum_{\substack{i_1<j_1<k_1 \\ i_2<j_2<k_2 \\ \{i_1,j_1,k_1\}\cap\{i_2,j_2,k_2\}\ne\varnothing}} P(M_{\pi(i_1),\pi(j_1)} M_{\pi(i_1),\pi(k_1)} M_{\pi(i_2),\pi(j_2)} M_{\pi(i_2),\pi(k_2)} \ne 0).$$

We first observe that when $2d > n$, the proof of Corollary 1 shows that $\frac{n^3}{144} < \sigma^2_{n,d}$, so, since the above sum contains $O(n^5)$ terms, all of which are at most 1, $\mathrm{Var}(E[(W'_{n,d} - W_{n,d})^2 | W_{n,d}]) \le \frac{K_3}{n^3}$ for some constant K_3.

As such, we need only worry about the summands for the $2d \le n$ case. Here we note that the sum can be broken up according to the nature of the intersection $\{i_1, j_1, k_1\} \cap \{i_2, j_2, k_2\}$ so that $P(M_{\pi(i_1),\pi(j_1)} M_{\pi(i_1),\pi(k_1)} M_{\pi(i_2),\pi(j_2)} M_{\pi(i_2),\pi(k_2)} \ne 0)$ is constant on each of these sets. For example, there are $\binom{n}{5}$ terms with $i_1 = i_2$ and $\{j_1, k_1\} \cap \{j_2, k_2\} = \varnothing$, and for each such term, the event that the product of the corresponding matrix entries is nonzero is equal to the event that within a random row of $M = M(n, d)$, four off-diagonal entries are chosen uniformly without replacement and are all nonzero. Letting A_i be the number of nonzero off-diagonal entries in row i of M, we have $A_i = \min\{i - 1, d\} + \min\{d, n - i\}$, so

$$P(M_{\pi(i_1),\pi(j_1)} M_{\pi(i_1),\pi(k_1)} M_{\pi(i_2),\pi(j_2)} M_{\pi(i_2),\pi(k_2)} \neq 0)$$

$$= \frac{1}{n} \sum_{i=1}^{n} \binom{A_i}{4} \binom{n-1}{4}^{-1} = \frac{(n-5)!}{n!} \sum_{i=1}^{n} \frac{A_i!}{(A_i-4)!}$$

$$= \frac{(n-5)!}{n!} \left[\sum_{i=1}^{d} \frac{(i+d-1)!}{(i+d-4)!} + (n-2d)\frac{(2d)!}{(2d-4)!} + \sum_{i=n-d+1}^{n} \frac{(n+d-i)!}{(n+d-i-4)!} \right]$$

$$= \frac{90nd^4 - 240nd^3 + 220nd^2 - 60nd - 98d^5 + 180d^4 + 50d^3 - 180d^2 + 48d}{5n(n-1)(n-2)(n-3)(n-4)}$$

$$= O(n^{-4}d^4).$$

The total contribution of such terms is thus $O(nd^4)$. The other cases may be handled similarly and the summands are still seen to contribute no more than $O(nd^4)$.

Now for $d \leq \sqrt{n}$, $\sigma_{n,d}^2 = O(nd)$, so $\mathrm{Var}(E[(W'_{n,d} - W_{n,d})^2|W_{n,d}]) \leq \frac{4}{n^2\sigma_{n,d}^4}O(nd^4) \leq K_1 d^2 n^{-3}$ for some constant K_1. When $\sqrt{(n)} < d \leq \frac{n}{2}$, $\sigma_{n,d}^2 = O(d^3)$, so that $\mathrm{Var}(E[(W'_{n,d} - W_{n,d})^2|W_{n,d}]) \leq K_2 d^{-2}n^{-1}$ for some constant K_2.

Finally, since $\sigma_{n,d}^2 \geq \frac{nd}{12}$, thus $\mathrm{Var}(E[(W'_{n,d} - W_{n,d})^2|W_{n,d}]) \leq \frac{4}{n^2\sigma_{n,d}^4}O(nd^4) = O(d^2n^{-3})$ when $2d \leq n$, and $\mathrm{Var}(E[(W'_{n,d} - W_{n,d})^2|W_{n,d}]) \leq K_3 n^{-3}$ when $2d > n$, for all fixed values of d, we have $\mathrm{Var}(E[(W'_{n,d} - W_{n,d})^2|W_{n,d}]) \leq K(d)n^{-3}$ for some constant $K(d)$ that does not depend on n. $\quad\square$

We are now in a position to bound the rate of convergence of d-descents.

Theorem 4. *The number of d-descents in a random permutation of length n satisfies*

$$\left| P\left(\frac{\mathrm{Des}_d - \mu_{n,d}}{\sigma_{n,d}} \leq x \right) - \Phi(x) \right| \leq \begin{cases} M_1 d^{\frac{3}{2}} n^{-\frac{1}{2}}, & d(n) \leq \sqrt{(n)} \\ M_2 n d^{-\frac{3}{2}}, & \sqrt{(n)} < d(n) \leq \frac{n}{2} \\ M_3 n^{-\frac{1}{2}}, & 2d(n) > n \\ M(d) n^{-\frac{1}{2}}, & d \text{ fixed} \end{cases}$$

where Φ is the standard normal c.d.f., M_1, M_2, M_3, and $M(d)$ are constants which do not depend on n, $\mu_{n,d} = \frac{2nd - d^2 - d}{4}$, and $\sigma_{n,d} = \sqrt{\mathrm{Var}(Y_{n,d})}$ is given by Theorem 1.

Proof. Apply Theorem 3 to the pair $(W_{n,d}, W'_{n,d})$ - recalling that $W_{n,d}$ is distributed as $\frac{Y_{n,d} - \frac{2nd-d^2-d}{4}}{\sqrt{\mathrm{Var}(Y_{n,d})}} = \frac{\mathrm{Des}_d - \mu_{n,d}}{\sigma_{n,d}}$ - where λ, A, and $\mathrm{Var}(E[(W'_{n,d} - W_{n,d})^2|W_{n,d}])$ are given by Lemma 1, Corollary 1, and Lemma 4, respectively. $\quad\square$

- Examination of the $d(n) \leq \frac{n}{2}$ cases shows that this method does not yield useful rates when $d(n) \in \Omega(n^{\frac{1}{3}}) \bigcap O(n^{\frac{2}{3}})$: In order that the statement is not completely vacuous, the rates must be $o(1)$. Bóna's paper shows that a central limit theorem holds in these cases, but the rates remain unknown.

- In each case, the $\frac{1}{\lambda}\sqrt{\mathrm{Var}(E[(W'_{n,d} - W_{n,d})^2 | W_{n,d}])}$ term and the $\frac{A^2}{\sqrt{\lambda}}$ term are of the same order, so, since the bounds on A were tight, the bounds from Lemma 4 are sufficient for getting the best possible order using the pair $(W_{n,d}, W'_{n,d})$. The largest term in each case comes from $\frac{A^3}{\lambda}$.

4   Conclusion

For the most part, we were able to successfully extend the work of Fulman in [6] to the more general setting of d-descents. In particular, Theorem 4 shows that the distribution of d-descents in a random permutation converges to the normal distribution on the order of $\frac{1}{\sqrt{n}}$ when d is fixed or $d(n) = \Theta(n)$. When $d(n) = o(n)$, we still get a central limit theorem with an error term unless $d(n) \in \Omega(n^{\frac{1}{3}}) \bigcap O(n^{\frac{2}{3}})$. In this case, we need to appeal to Bóna's work using Janson's criterion to get asymptotic normality and the rate is unknown. The problem in this regime is that $\frac{d(n)}{\sqrt{\mathrm{Var}(Y_{n,d})}}$ has a spike around $d(n) = \sqrt{n}$. The only way around this obstacle using the methods of this paper would be to find another pair of random variables related to Des_d that satisfy the assumptions of Theorem 3 and yield a larger value of λ and/or a smaller value of A. The author has been unable to find such a pair and at present it is not clear whether the rates for the $2d(n) \leq n$ cases derived in this paper are artifacts of the methodology or a true reflection of the actual dependence of the rates on the growth of $d(n)$. The bounds in the $d(n) \in \Omega(n^{\frac{1}{3}}) \bigcap O(n^{\frac{2}{3}})$ cases are certainly not optimal and it would not be surprising to learn that an order $O(n^{-\frac{1}{2}})$ convergence rate holds for all choices of $d(n)$, though it seems that another approach is needed to determine whether this is the case.

Beyond broadening the known results on the distribution of d-descents of a random permutation and establishing a relationship between these statistics and certain eigenfunctions of the random-to-end shuffle, this paper has demonstrated a method that may be useful for getting rates in other cases involving real-valued statistics defined on groups. For example, the construction of the pair (W, W') outlined in the remark following Lemma 1 may be relevant to some statistics involving metrics on groups. Indeed, $\mathrm{Des}_{n-1}(\pi) = \mathrm{Inv}(\pi) = \tau(\pi)$ is the Kendall tau metric (see [3], Ch. 6). It might also yield central limit theorems for statistics concerning group representations. In fact, as mentioned in the discussion following Lemma 1, the method applies more generally to the study of eigenfunctions (corresponding to real eigenvalues) of ergodic Markov chains in general so there are many potential applications of the technique. Finally, the preceding analysis shows that Rllin's theorem can greatly simplify many arguments using Stein's method by removing the condition of exchangeability and thus extends the reach of this powerful tool.

Acknowledgement

References

[1] M. Bóna, Generalized descents and normality. Electronic Journal of Combinatorics **15** (2008), no. 1.

[2] F. De Mari, M.A. Shayman, Generalized Eulerian numbers and the topology of the Hessenberg variety of a matrix. *Acta Appl. Math.* **12** (1988), no. 3, 213-235.

[3] P. Diaconis, Group representations in probability and statistics, *Institute of Mathematical Statistics Lecture Notes*, **11**, 1988.

[4] P. Diaconis, Random walks on groups: Characters and geometry, *Groups St. Andrews 2001 in Oxford*, vol. 1, Cambridge University Press, 2000, pp. 120-142.

[5] R. Durrett, *Probability: Theory and Examples*, Ed. 3, Brooks/Cole Publishing Company, 2005.

[6] J. Fulman, Stein's method and non-reversible Markov chains. *Stein's Method: Expository Lectures and Applications.* IMS Lecture Notes - Monogr. Ser., Vol. 46. 66-74. 2004.

[7] D. Levin, Y. Peres, and E. Wilmer, Markov Chains and Mixing Times, American Mathematical Society, 2009.

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

[9] A. Rllin, A note on the exchangeability condition in stein's method, *Statistics and Probability Letters*, **78** (2008), 1800-1806.

[10] C. Stein, Approximate computation of expectations, *Institute of Mathematical Statistics Lecture Notes*, **7**, 1986.