Asymptotic Analysis of (3, 2, 1)-Shell Sort

R.T. Smythe,[1] J.A. Wellner[2]*
[1]*Department of Statistics, Oregon State University, Corvallis, OR 97331; e-mail: smythe@stat.orst.edu*
[2]*Statistics, University of Washington, Box 354322, Seattle, WA 98195-4322; e-mail: jaw@stat.washington.edu*

Received 13 July 2001; accepted 12 January 2002

DOI 10.1002/rsa.10046

ABSTRACT: We analyze the (3, 2, 1)-Shell Sort algorithm under the usual random permutation model. © 2002 Wiley Periodicals, Inc. Random Struct. Alg., 21: 59–75, 2002

Keywords: additive functional; algorithm; local limit theorem; Markov chain; Poissonization

1. INTRODUCTION

Shell Sort is an algorithm for sorting a list of numbers in stages. The algorithm, proposed by Shell [16], was given in Knuth's 1973 book [6] and has generated a substantial literature in the years since (many references are given in Sedgewick [14] and Mahmoud [10]). Shell Sort is a generalization of insertion sort. It has little overhead and is easy to implement, and is thus a practical choice for sorting moderate-sized lists. In sorting n random keys, Shell Sort is able to improve considerably on the $\Theta(n^2)$ average running time of insertion sort (Sedgewick [14] gives detailed results for different parameter choices.)

In a recent paper (Smythe and Wellner [17]) we gave a probabilistic analysis of

Correspondence to: R. T. Smythe

*Research supported in part by National Science Foundation grant DMS-95-32039 and NIAID grant 2R01-AI291968-04.

two-stage Shell Sort. For the special case of (2,1)-Shell Sort, this led to a rederivation of the limiting result of Louchard [9], and the approach was generalized to give limiting results for $(h, 1)$-Shell Sort, for arbitrary h. In the present work, we use the tools developed in the previous paper to analyze a particular case of three-stage Shell Sort, (3, 2, 1)-Shell Sort. In analyzing $(h, k, 1)$-Shell Sort, the case in which h and k are relatively prime is of most interest, and (3, 2, 1) is the simplest such example. It turns out that the last stage of (3, 2, 1)-Shell Sort, which consists of sorting a list that is both 2-sorted and 3-sorted, makes an asymptotically normal number of comparisons. Recent work of Janson and Knuth [5] gives detailed results on the mean number of comparisons for $(h, k, 1)$-Shell Sort; for the (3, 2, 1) case, we are able to find the variance as well.

Section 2 gives a brief description of the algorithm; more details may be found, e.g., in Sedgewick and Flajolet [15] or Mahmoud [10]. In Section 3 we develop a characterization of the number of comparisons in the final stage of (3, 2, 1)-Shell Sort as a functional of a generalized hypergeometric process. Using some basic results from the theory of empirical processes and spacings, we give a heuristic argument that the limiting number of comparisons would be expected to have a normal distribution. Section 4 introduces a kind of Poissonization argument and invokes a local limit theorem for Markov chains to confirm that the limiting number of comparisons is indeed normally distributed. The key step here involves the verification of the fact that a limit of a conditional distribution is the conditional distribution of the limit: this turns out to be surprisingly difficult, and the more technical parts of the proof are relegated to two appendices.

In Section 5 we discuss briefly possibilities for generalizing the arguments given here to $(h, 2, 1)$-Shell Sort.

2. THE ALGORITHM

Insertion sort proceeds by progressively adding keys to an already sorted file. Shell Sort performs several stages of insertion sort. Assume the given data to be a linear array structure of size n. For a k-stage Shell Sort, let $t_k, t_{k-1}, \ldots, t_1$ $(= 1)$ be a decreasing sequence of positive integers. The first stage in sorting n keys sorts (by ordinary insertion) keys that are t_k positions apart in the list, creating t_k sorted subarrays of length at most $\lceil n/t_k \rceil$. In the second stage, t_{k-1} subarrays of keys that are t_{k-1} apart are sorted, and so on down to the last stage, which performs an insertion sort of the entire array.

As an example, we apply (3, 2, 1)-Shell Sort to sort the array

$$3 \quad 12 \quad 6 \quad 10 \quad 5 \quad 9 \quad 8 \quad 1 \quad 11 \quad 4 \quad 7 \quad 2.$$

The first stage creates 3 sorted lists of length 4:

$$\begin{array}{cccc} 3 & 4 & 8 & 10 \\ 1 & 5 & 7 & 12 \\ 2 & 6 & 9 & 11. \end{array}$$

The second stage takes the resulting 3-sorted list,

$$3 \quad 1 \quad 2 \quad 4 \quad 5 \quad 6 \quad 8 \quad 7 \quad 9 \quad 10 \quad 12 \quad 11,$$

and creates two sorted lists of length 6:

$$\begin{array}{cccccc} 2 & 3 & 5 & 8 & 9 & 12 \\ 1 & 4 & 6 & 7 & 10 & 11. \end{array}$$

We now have a list that is both 3-sorted and 2 sorted:

$$2 \quad 1 \quad 3 \quad 4 \quad 5 \quad 6 \quad 8 \quad 7 \quad 9 \quad 10 \quad 12 \quad 11,$$

and the final stage sorts this list.

As in the prior work, the concept of an *inversion* in a permutation plays a key role in the analysis. Let $\Pi_n = (\pi_1, \ldots, \pi_n)$ be a permutation of $(1, \ldots, n)$. The pair (π_i, π_j) is an inversion if π_i and π_j are out of their natural order, that is, if $\pi_i > \pi_j$ when $i < j$.

The notation $Z_{(j)}$ will be slightly abused to denote the jth-order statistics among $Z_1, \ldots, Z_r$. (It would be more accurate to use $Z_{(j:r)}$, as $Z_{(j:r)}$ and $Z_{(j:s)}$ may differ for $r \neq s$; however, the second subscript will generally be obvious, and will be dropped for convenience.)

3. ANALYSIS OF (3, 2, 1)-SHELL SORT

In the analysis of (3, 2, 1)-Shell Sort, it is notationally convenient to assume we are given a linear array of size $3n$. We assume the usual random permutation model for the data, in which the ranks of the data are equally likely to be any of the permutations of $\{1, 2, \ldots, 3n\}$, each occurring with probability $1/(3n)!$. We may assume that our data are $3n$ real numbers from a continuous probability distribution, and because the probability integral transform preserves ordering of the data, we may (and will) assume that the probability distribution is uniform on (0,1). Prior to any sorting, we will denote our raw array by

$$X_1, Y_1, Z_1, X_2, Y_2, Z_2, \ldots, X_n, Y_n, Z_n,$$

where the X's, Y's, and Z's may be taken to be mutually independent. The first stage of the algorithm sorts the X's, Y's, and Z's separately. Analogously to the case described in Smythe and Wellner [17], if C_n denotes the number of comparisons made by (linear) insertion sort to sort n random keys, the initial stage of (3, 2, 1)-Shell Sort makes three runs of insertion sort on the subarrays $X_1, \ldots, X_n$, $Y_1, \ldots, Y_n$, and $Z_1, \ldots, Z_n$, requiring

$$C_n^1 + C_n^2 + C_n^3$$

comparisons, where $C_j^1 \overset{\mathcal{D}}{=} C_j^2 \overset{\mathcal{D}}{=} C_j^3$ and C^1, C^2, C^3 are independent. We then have the 3-sorted list

$$X_{(1)}, Y_{(1)}, Z_{(1)}, \ldots, X_{(n)}, Y_{(n)}, Z_{(n)}.$$

The next stage of the algorithm 2-sorts this list: we make two lists,

$$X_{(1)}, Z_{(1)}, Y_{(2)}, X_{(3)}, Z_{(3)}, Y_{(4)}, \ldots \tag{3.1}$$

and

$$Y_{(1)}, X_{(2)}, Z_{(2)}, Y_{(3)}, X_{(4)}, Z_{(4)}, \ldots, \tag{3.2}$$

and we sort each of these lists. Each of these lists is 3-sorted; let $\tilde{C}_n$ denote the number of comparisons made by insertion sort to sort a 3-sorted array of length n. The second stage of the algorithm thus requires

$$\tilde{C}^1_{\lceil 3n/2 \rceil} + \tilde{C}^2_{\lfloor 3n/2 \rfloor}$$

comparisons, where $\tilde{C}^1_j \overset{\mathcal{D}}{=} \tilde{C}^2_j$ and $\tilde{C}^1$, $\tilde{C}^2$ are independent. The final stage of the algorithm sorts the list, which is now both 3-sorted and 2-sorted. Denote the keys in the list (3.1) by $U_1, U_2, \ldots$ and those in (3.2) by $V_1, V_2, \ldots$. When we are about to insert $V_{(j)}$, we place it among

$$\{U_{(1)}, \ldots, U_{(j)}\} \cup \{V_{(1)}, \ldots, V_{(j-1)}\}.$$

The so-called sentinel version of insertion sort makes

$$C(\Pi_n) = n + I(\Pi_n)$$

comparisons to sort a permutation Π_n on n letters with $I(\Pi_n)$ inversions. Thus the overall number of comparisons S_n made by (3, 2, 1)-Shell Sort is given by the sum

$$S_n = C^1_n + C^2_n + C^3_n + \tilde{C}^1_{\lceil 3n/2 \rceil} + \tilde{C}^2_{\lfloor 3n/2 \rfloor} + 3n + I_{3n}, \tag{3.3}$$

where all the terms are independent, and I_{3n} denotes the number of inversions in the list that is both 2-sorted and 3-sorted. It is known (Lent and Mahmoud [8]) that for linear search,

$$\frac{C_n - \frac{1}{4}n^2}{n^{3/2}} \overset{\mathcal{D}}{\longrightarrow} \mathcal{N}\left(0, \frac{1}{36}\right).$$

The second stage requires sorting of 3-sorted lists, for which the number of comparisons is given by the $\tilde{C}$ terms of (3.3). The asymptotic distribution of $\tilde{C}^1_j$ was characterized in Theorem 2 of Smythe and Wellner [17] (no explicit expression is known for the limiting distribution). Our focus in this paper is on the last stage, i.e., the limiting distribution of I_{3n}.

Regard the X's, Y's, and Z's as three independent sets of n i.i.d. observations, uniformly distributed on (0, 1), giving rise to a set of $3n$ points in (0, 1). Associate with each of these points a triple giving the parity of the numbers of X's, Y's, and Z's that precede the point; for example, the triple OEO means that the number of X-predecessors is odd, the number of Y-predecessors is even, and the number of Z-predecessors is odd. The first point will thus be labeled EEE, and the next will be OEE, EOE, or EEO according as the smallest key is an X, Y, or Z resp. Let

$$N_1 \equiv \text{number of points of types } OOE, EEO,$$

$$N_2 \equiv \text{number of points of types } OEE, EOO,$$

$$N_3 \equiv \text{number of points of types } OOO, EEE,$$

$$N_4 \equiv \text{number of points of types } OEO, EOE.$$

Then $N_1 + N_2 + N_3 + N_4 = 3n$, so by symmetry of the N_i's

$$0 = \mathrm{Cov}(N_1, N_1 + N_2 + N_3 + N_4) = \mathrm{Var}(N_1) + 3\,\mathrm{Cov}(N_1, N_2),$$

and this implies, again using symmetry,

$$\mathrm{Cov}(N_1 + N_4, N_2 + N_4) = 3\,\mathrm{Cov}(N_1, N_2) + \mathrm{Var}(N_4) = 0.$$

Thus $N_1 + N_4$, $N_2 + N_4$, and $N_3 + N_4$ are uncorrelated. Further, we can show, using basic empirical process theory and results on spacings for uniform order statistics (cf., for example, Pyke [12]) that each of $N_1 + N_4$, $N_2 + N_4$, and $N_3 + N_4$ has an asymptotically normal distribution with mean equal to $3n/2$ and variance equal to $3n/8$. Since the sum of these variables is $3n + 2N_4$, this provides a strong hint, though not yet a proof, that N_4 will have an asymptotically normal distribution. Next we show that N_4 is precisely equal to I_{3n}, the number of inversions.

Proposition 1. *An inversion in the* 2-*sorted and* 3-*sorted list occurs when, and only when, the key causing the inversion is of type OEO or EOE. Hence* I_{3n}, *the number of inversions in the* 2-*sorted and* 3-*sorted list, is equal to* N_4, *the number of keys of type* $\{OEO, EOE\}$.

The following lemma is needed to characterize inversions.

Lemma. In a 2-sorted and 3-sorted list $\{\pi_i\}_{i=1}^{3n}$, the only possible inversions are when $\pi_i > \pi_{i+1}$.

Proof of the Lemma. Given i, any $k > i + 1$ satisfies $2a + 3b = k - i$ for some nonnegative integers a, b. ■

Proof of Proposition 1. The 2-sorted and 3-sorted list is derived by merging the sorted versions of lists (3.1) and (3.2).

A. Suppose $X_{(k)}$ causes an inversion, where k is assumed odd. If the next key after $X_{(k)}$ in the merged list is $Y_{(j)}$ for some j, then $Y_{(j)} < X_{(k)}$ and j must be odd, since $X_{(k)}$ appears in the sorted version of list (3.1) and $Y_{(j)}$ in that of list (3.2). Thus exactly j of the Y's are less than $X_{(k)}$. However, the total number of keys less than $X_{(k)}$ is odd, as there is one more predecessor of $X_{(k)}$ in the ordered version of (3.2) than in (3.1), owing to the inversion. Thus the type of $X_{(k)}$ must be EOE. If k is even, a similar argument gives the type as OEO.

If the next key after $X_{(k)}$ is $Z_{(l)}$ for some l, a similar argument applies. Now suppose the next key is also an X. Again assume that k is odd and that the next key after $X_{(k)}$ is $X_{(k-1)}$. [Note that this can happen because of the 2-sorting, even though the X's were originally put in order.] In this case the total number of predecessors of $X_{(k)}$ is again odd, so the type must be EOE or EEO. The number of predecessors of $X_{(k)}$ in the ordered version of (3.2) is one greater than the number from (3.1). Suppose there are P_1 Y-predecessors coming from (3.1), and P_2 coming from (3.2). If $P_1 \leq P_2$, then the largest Z-predecessor must come from (3.2), and $X_{(k)}$ has type EOE. If $P_1 > P_2$, the largest Y-predecessor comes from (3.1), and again $X_{(k)}$ has type EOE. If k is even, a similar argument gives the type as OEO, and the cases where $Y_{(j)}$ or $Z_{(l)}$ cause an inversion are treated by the same argument.

B. Now suppose, for example, that a key has type OEO, and suppose it is $X_{(k)}$ for k even. Just before merging the sorted versions of (3.1) and (3.2), consider the position of $X_{(k)}$ with respect to its two “neighbors” in list (3.1). If $X_{(k)}$ is in the “correct” position, it would have an odd number of predecessors [one more from (3.1) than from (3.2)]. So either $X_{(k)}$ is greater than the next key in (3.1) or less than the previous key in (3.1). But in the latter case, the previous key would cause an inversion, and must have type EEO, OOO, or OOE; by part A of the proof, this is impossible. Hence $X_{(k)}$ causes an inversion. Similar arguments give the EOE case. ∎

We sum up the results of this section in the following theorem, to be proved in the next section.

Theorem 1. $(I_{3n} - 3n/4)/\sqrt{n} \xrightarrow{\mathcal{D}} N(0, 9/32)$.

4. PROOF OF THEOREM 1

Our analysis in Section 3 shows that the distribution of N_4 depends only on the relative order of appearance of the X's, Y's, and Z's as one traverses the interval (0, 1) from left to right. Thus a simple urn model describes the process of interest. Let the state space Ω_0 of our process be defined by

$$\Omega_0 = \{OOE, EEO, OEE, EOO, OOO, EEE, EOE, OEO\}. \tag{4.1}$$

Initially the urn contains n balls of type X, n of type Y, and n of type Z. At each stage, a ball is drawn at random from the urn and its type recorded, and the process continues until all balls are gone. At stage k, the process is in state, say, OEO, if the first k balls drawn include an odd number of X's, an even number of Y's, and an odd number of Z's.

Thus N_4, which simply counts the number of times the process is in states EOE and OEO, has a direct combinatorial interpretation, and one might hope for a simple proof of its asymptotic normality. As noted in Section 3, $N_i + N_4$, $i = 1, 2, 3$, has an asymptotically normal distribution; but this does not seem of much help in establishing the result for N_4 alone. The difficulty is that the probabilities of drawing the different types of ball at a given stage depend not just on the previous state ω, but on the actual numbers of each type previously drawn, so that the evolution of the process is complicated. We therefore proceed by a kind of Poissonization argument to define a Markov chain on a state space formed from Ω_0, and identify the distribution of N_4 with a conditional distribution resulting from the "Poissonized" problem.

We construct three independent Poisson processes with intensity n, as follows. Suppose that $T_1, T_2, \ldots$ are i.i.d. Exponential(1) random variables, and set $T_{n,j} \equiv T_j/3n$ for $j = 1, 2, \ldots$ and $n = 1, 2, \ldots$. Thus $T_{n,j} \overset{\mathcal{D}}{=}$ Exponential($3n$), and the stochastic process $\mathcal{N}_n$ defined by

$$\mathcal{N}_n(t) = \sum_{j=1}^{\infty} 1_{[T_{n,j} \leq t]}$$

is a Poisson process with intensity $3n$. Now let $\Delta_1, \Delta_2, \ldots$ be i.i.d. with uniform distribution on $\{1, 2, 3\}$: $P(\Delta_j = k) = \frac{1}{3}$ for $k \in \{1, 2, 3\}$ and all $j = 1, 2, \ldots$, and define three thinned versions, $\mathcal{N}_n^X$, $\mathcal{N}_n^Y$, and $\mathcal{N}_n^Z$ of the process $\mathcal{N}_n$ as follows:

$$\mathcal{N}_n^X(t) = \sum_{j=1}^{\infty} 1_{[T_{n,j} \leq t]} 1_{[\Delta_j = 1]}, \quad \mathcal{N}_n^Y(t) = \sum_{j=1}^{\infty} 1_{[T_{n,j} \leq t]} 1_{[\Delta_j = 2]},$$

$$\mathcal{N}_n^Z(t) = \sum_{j=1}^{\infty} 1_{[T_{n,j} \leq t]} 1_{[\Delta_j = 3]}.$$

The processes $\mathcal{N}_n^X$, $\mathcal{N}_n^Y$, and $\mathcal{N}_n^Z$ are then independent Poisson processes with intensity n (cf. Çinlar [4], p. 89).

Each point in the Poisson process $\mathcal{N}_n$ can be associated in an obvious way with one of the states in Ω_0. For calculation purposes, however, it is simpler to group the elements of Ω_0, and define a new state space

$$\Omega_1 \equiv \{\{OOE, EEO\}, \{OEE, EOO\}, \{OOO, EEE\}, \{EOE, OEO\}\}. \tag{4.2}$$

Denote the four states of Ω_1 by e_1, e_2, e_3, e_4, respectively. It is easily seen that $\mathcal{N}$ defines a Markov chain $V(n)$ on Ω_1, with $V(0) = e_3$ and with evolution governed by the following transition matrix P_1:

$$\begin{pmatrix} 0 & 1/3 & 1/3 & 1/3 \\ 1/3 & 0 & 1/3 & 1/3 \\ 1/3 & 1/3 & 0 & 1/3 \\ 1/3 & 1/3 & 1/3 & 0 \end{pmatrix}.$$

After $3n$ events of the chain V have been observed, we have observed

$$M_n^X \equiv \sum_{j=1}^{3n} 1_{\{\Delta_j=1\}} \quad \text{points labeled as } X,$$

$$M_n^Y \equiv \sum_{j=1}^{3n} 1_{\{\Delta_j=2\}} \quad \text{points labeled as } Y,$$

$$M_n^Z \equiv \sum_{j=1}^{3n} 1_{\{\Delta_j=3\}} \quad \text{points labeled as } Z.$$

Note that $(M_n^X, M_n^Y, M_n^Z) \sim \text{Mult}_3(3n, 1/3, 1/3, 1/3))$. Define

$$\mathcal{N}_4(k) \equiv \sum_{j=1}^{k} 1_{\{V(j)\in e_4\}}, \tag{4.3}$$

so that $\mathcal{N}_4(3n)$ is just the "Poissonized version" of the random variable N_4 whose distribution we seek. The asymptotic mean and variance of $\mathcal{N}_4$ may be easily computed by renewal process arguments.

Lemma 1. *Let T denote the first return time to e_4 of the Markov chain V on Ω_1. Then $E(T) = 4$, $Var(T) = 6$.*

Proof. It is not difficult to see that

$$P(T = k) = \frac{2^{k-2}}{3^{k-1}} \quad \text{for} \quad k \geq 2. \tag{4.4}$$

For $k = 2$, the chain goes to another state, then returns to e_4; since there are 3 other states, the probability of return in 2 steps is just 3(1/9), or 1/3. For $k > 2$, note that after leaving e_4, at step $j < k$ the process has only 2 possibilities: it must leave its current state, and it can't go back to e_4. We get $3[2^{k-2}]$ possible routes, out of a total of 3^k routes, that give a first return to e_4 in k steps. This gives (4.4). The calculation of the mean and variance follows easily. ∎

Lemma 2. *$\mathcal{N}_4(3n)$ has an asymptotically normal distribution with*

$$E(\mathcal{N}_4(3n)) \sim 3n/4, \quad \text{Var}(\mathcal{N}_4(3n)) \sim 9n/32.$$

Proof. By standard arguments for renewal theory (cf. Ross [13], p. 62), $\mathcal{N}_4(3n)$ is asymptotically normal with mean and variance given by

$$E(\mathcal{N}_4(3n)) \sim \frac{3n}{E(T)} \quad \text{and} \quad \text{Var}(\mathcal{N}_4(3n)) \sim \frac{\text{Var}(T)(3n)}{E(T)^3}.$$

The result now follows from Lemma 1. ∎

We see from Lemma 2 that the "Poissonized" $\mathcal{N}_4$ has the same asymptotic variance as the "real" N_4. To make the connection between the two, we note that for a positive integer k,

$$P(N_4(3n) = k) = P(\mathcal{N}_4(3n) = k | M_n^X = n, \quad M_n^Y = n, \quad M_n^Z = n). \tag{4.5}$$

To prove Theorem 1, we will find the limit of the right-hand side of (4.5), suitably normalized, and show that the conditional distribution of $\mathcal{N}_4$ is asymptotically independent of the conditioning variables.

The state space Ω_1 for the Markov chain V is too crude for this purpose, and we now define a twelve-state Markov chain as follows. Let $e_i \cap X$, $i = 1, 2, 3, 4$, denote the event that a point of $\mathcal{N}_n$ is of type X and corresponds to one of the two configurations comprising e_i. Make a similar definition of $e_i \cap Y$ and $e_i \cap Z$. This refines the four states of Ω_1 into twelve states, denoted as follows:

$$\{1, 2, \ldots, 12\} \equiv \{e_1 \cap X, e_1 \cap Y, e_1 \cap Z, e_2 \cap X, e_2 \cap Y, e_2 \cap Z, \\ e_3 \cap X, e_3 \cap Y, e_3 \cap Z, e_4 \cap X, e_4 \cap Y, e_4 \cap Z\}.$$

The points of $\mathcal{N}$ then form a Markov chain W with transition matrix P given by

$$\begin{pmatrix}
0 & 0 & 0 & 0 & 1/3 & 0 & 0 & 0 & 1/3 & 1/3 & 0 & 0 \\
0 & 0 & 0 & 0 & 1/3 & 0 & 0 & 0 & 1/3 & 1/3 & 0 & 0 \\
0 & 0 & 0 & 0 & 1/3 & 0 & 0 & 0 & 1/3 & 1/3 & 0 & 0 \\
0 & 1/3 & 0 & 0 & 0 & 0 & 1/3 & 0 & 0 & 0 & 0 & 0 \\
0 & 1/3 & 0 & 0 & 0 & 0 & 1/3 & 0 & 0 & 0 & 0 & 0 \\
0 & 1/3 & 0 & 0 & 0 & 0 & 1/3 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1/3 & 1/3 & 0 & 0 & 0 & 0 & 0 & 0 & 1/3 & 0 \\
0 & 0 & 1/3 & 1/3 & 0 & 0 & 0 & 0 & 0 & 0 & 1/3 & 0 \\
0 & 0 & 1/3 & 1/3 & 0 & 0 & 0 & 0 & 0 & 0 & 1/3 & 0 \\
1/3 & 0 & 0 & 0 & 0 & 1/3 & 0 & 1/3 & 0 & 0 & 0 & 0 \\
1/3 & 0 & 0 & 0 & 0 & 1/3 & 0 & 1/3 & 0 & 0 & 0 & 0 \\
1/3 & 0 & 0 & 0 & 0 & 1/3 & 0 & 1/3 & 0 & 0 & 0 & 0
\end{pmatrix}.$$

The initial distribution of the chain may be defined by $\pi_0(7) = \pi_0(8) = \pi_0(9) = 1/3$. Define

$$C_{n,i} \equiv \sum_{j=1}^{3n} 1_{\{W_j = i\}}, \tag{4.6}$$

the number of visits to state i by time $3n$ for $i = 1, \ldots, 12$. Then

$$C_{n,1} + C_{n,4} + C_{n,7} + C_{n,10} = M_n^X,$$

$$C_{n,2} + C_{n,5} + C_{n,8} + C_{n,11} = M_n^Y,$$

$$C_{n,3} + C_{n,6} + C_{n,9} + C_{n,12} = M_n^Z,$$

while

$$\begin{aligned} C_{n,1} + C_{n,2} + C_{n,3} &= \mathcal{N}_1(3n), \\ C_{n,4} + C_{n,5} + C_{n,6} &= \mathcal{N}_2(3n), \\ C_{n,7} + C_{n,8} + C_{n,9} &= \mathcal{N}_3(3n), \\ C_{n,10} + C_{n,11} + C_{n,12} &= \mathcal{N}_4(3n). \end{aligned}$$

It is well known (cf., for example, Chung [3], p. 99) that functionals of a positive recurrent Markov chain satisfy a central limit theorem. In particular, for any functional of the form

$$f(W_j) = \sum_{i=1}^{12} a_i 1\{W_j = i\}, \quad \text{where } a_i \in R, \quad i = 1, \ldots, 12,$$

it follows that $\sum_{i=1}^{n} f(W_j)$, suitably normalized, has an asymptotically normal limit. Then, by the Cramér-Wold theorem (see, e.g., Billingsley [2], p. 397),

$$\underset{\sim}{C}_n \equiv (C_{n,1}, C_{n,2}, \ldots, C_{n,12})$$

has asymptotically a joint normal distribution (the covariance matrix is, of course, not of full rank, as the $C_{n,i}$ sum to $3n$.)

Computation of the asymptotic covariance matrix of the vector $\underset{\sim}{C}_n$ is nontrivial, and involves heavy use of the symmetries of the Markov chain W. The matrix is given below; the derivation is left to Appendix A.

Lemma 3. *The asymptotic covariance matrix of* $(\underset{\sim}{C}_n - (3n)/12)/\sqrt{n}$ *is given by*

$$\Sigma = \frac{1}{96}\begin{pmatrix} 19 & -5 & -5 & -5 & 1 & -5 & -5 & -5 & 1 & 7 & 1 & 1 \\ -5 & 19 & -5 & 1 & 7 & 1 & -5 & -5 & 1 & 1 & -5 & -5 \\ -5 & -5 & 19 & -5 & 1 & -5 & 1 & 1 & 7 & 1 & -5 & -5 \\ -5 & 1 & -5 & 19 & -5 & -5 & 7 & 1 & 1 & -5 & -5 & 1 \\ 1 & 7 & 1 & -5 & 19 & -5 & 1 & -5 & -5 & -5 & -5 & 1 \\ -5 & 1 & -5 & -5 & -5 & 19 & 1 & -5 & -5 & 1 & 1 & 7 \\ -5 & -5 & 1 & 7 & 1 & 1 & 19 & -5 & -5 & -5 & 1 & -5 \\ -5 & -5 & 1 & 1 & -5 & -5 & -5 & 19 & -5 & 1 & 7 & 1 \\ 1 & 1 & 7 & 1 & -5 & -5 & -5 & -5 & 19 & -5 & 1 & -5 \\ 7 & 1 & 1 & -5 & -5 & 1 & -5 & 1 & -5 & 19 & -5 & -5 \\ 1 & -5 & -5 & -5 & -5 & 1 & 1 & 7 & 1 & -5 & 19 & -5 \\ 1 & -5 & -5 & 1 & 1 & 7 & -5 & 1 & -5 & -5 & -5 & 19 \end{pmatrix}.$$

In other words,

$$\frac{\underset{\sim}{C}_n - (3n)/12}{\sqrt{n}} \xrightarrow{\mathcal{D}} N(0, \Sigma).$$

If we now take

$$A = \begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \end{pmatrix},$$

then $A\underset{\sim}{C}_n = (M_n^X, M_n^Y, M_n^Z, \mathcal{N}_1(3n), \mathcal{N}_2(3n), \mathcal{N}_3(3n), \mathcal{N}_4(3n))$, and $A\Sigma A'$ is a matrix of the form

$$\begin{pmatrix} B & F \\ F' & D \end{pmatrix}, \tag{4.7}$$

where

$$B = 3\left\{\frac{1}{3} I_{3\times 3} - (1/3)^2 \underline{1}\,\underline{1}'\right\} = I_{3\times 3} - (1/3)\underline{1}\,\underline{1}', \tag{4.8}$$

$$F = 0, \quad D = \frac{12}{32} I_{4\times 4} - \frac{3}{32} \underline{1}\,\underline{1}'. \tag{4.9}$$

From this, and the knowledge that $\underset{\sim}{C}_n$ has a limiting normal distribution, we find that $\mathcal{N}_4$ is asymptotically independent of (M_n^X, M_n^Y, M_n^Z). The final step is to show that the limiting conditional distribution of

$$\left(\frac{\mathcal{N}_4(3n) - 3n/4}{\sqrt{n}} \,\middle|\, M_n^X = n,\ M_n^Y = n,\ M_n^Z = n\right) \tag{4.10}$$

converges to the conditional distribution of the limit, which from Lemma 2 is $N(0, 9/32)$. From Eq. (4.5), this gives the desired result. In cases not involving independent summands, as in our problem, it is not always true that the limit of the conditional distribution is the conditional distribution of the limit [cf. Steck [18]). This result holds in our case, however: the following theorem restates Theorem 1 in terms of the conditional limit of the "Poissonized" process.

Theorem 1′. *Conditionally on* $B_n \equiv [M_n^X = n, M_n^Y = n, M_n^Z = n]$,

$$(\mathcal{N}_4(3n) - 3n/4)/\sqrt{n} \xrightarrow{\mathcal{D}} N(0, 9/32).$$

Proof. Note first that by Stirling's formula,

$$P(B_n) = \frac{(3n)!}{(n!)^3}\left(\frac{1}{3}\right)^{3n} \sim \frac{\sqrt{3}}{2\pi n}.$$

Thus if we can find the limit of

$$\begin{aligned} &P\{((\mathcal{N}_4(3n) - 3n/4)/\sqrt{n}, (M_{3n}^X - n)/\sqrt{n}, (M_{3n}^Y - n)/\sqrt{n}) \\ &\quad = ((k - 3n/4)/\sqrt{n}, (j_1 - n)/\sqrt{n}, (j_2 - n)/\sqrt{n})\}, \end{aligned}$$

we can find the desired conditional limit. This last step is accomplished via a local limit theorem for Markov chains, due to Kolmogorov [7]; the details are left to Appendix B.

5. POSSIBLE EXTENSION TO (h, 2, 1)-SHELL SORT

The outlines of a possible extension can be seen in the arguments for (3, 2, 1)-Shell Sort, but the combinatorial difficulties seem to become formidable as h grows. For any h, we can formulate a simple urn model, as at the beginning of Section 4, and presumably characterize the number of inversions as a subset of the "parity set" corresponding to our Ω_0. For $h = 5$, for example, the "parity set" has 32 members, and the number of inversions corresponds to a 12-member subset which can be identified with a bit of work. The problem can then be "Poissonized" as in Section 4, and there appears to be no reason why the conditional limit argument via a local limit theorem should not work to give asymptotic normality of the number of inversions. The difficulties appear to lie in identifying the appropriate "parity set" and especially in finding the variance of the number of inversions. The techniques used in our problem to find the variance will be much more difficult to apply for larger h, and we have not attempted this even for $h = 5$.

APPENDIX A: PROOF OF LEMMA 3

Here we give the derivation of the covariance matrix Σ of Lemma 3.

Let T_i be the first return time to a generic state $i \in \{1, 2, \ldots, 12\}$ of the Markov chain W. First note that since the transition matrix P of W (given in Section 4) is doubly stochastic, the stationary distribution is uniform on $\{1, 2, \ldots, 12\}$, and renewal theory results give that $E(T_i) = 12$. To find the variance of T_i, recall that the "aggregated" version of the state space, which we denoted Ω_0, consists of $\{e_1, e_2, e_3, e_4\}$, where $e_1 = \{1, 2, 3\}$, etc. The return time to each of the states e_i has mean 4 and variance 6, as given in Lemma 1. For $W(j)$ to return to a state $i \in e_r$, the chain must return to e_r. On each such return, it has probability 1/3 of being in state i. Thus the sojourn from state i can be represented as

$$T_i = \sum_{j=1}^{K} T_{i,j},$$

where each $T_{i,j}$ represents a return time from e_r to e_r, K represents the number of returns to e_r until i is reached, and K is independent of the $T_{i,j}$. Then K is a geometric random variable with parameter 1/3, with mean 3 and variance 6. Thus

$$\begin{aligned} \mathrm{Var}(T_i) &= E[\mathrm{Var}(T_i)|K] + \mathrm{Var}[E(T_i|K)] \\ &= E[K\ \mathrm{Var}(T_{i,j}) + \mathrm{Var}[KE(T_{i,j})] \\ &= 6E(K) + 16\ \mathrm{Var}(K) = 18 + 96 = 114. \end{aligned}$$

Using a renewal theory result as in Lemma 2, this gives the asymptotic variance for the number of visits to state i in the first $3n$ steps of the Markov chain W as

$$114(3n)/(12)^3 = 19n/96. \tag{A.1}$$

This gives us the diagonal elements of Σ. The off-diagonal elements result from knowing the variance of $\mathcal{N}_i(3n)$ and using symmetry. For example, knowing that $\mathrm{Var}(\mathcal{N}_4(3n)) \sim 9n/32$, then asymptotic equality of the covariances of the pairs $(C_{n,10}, C_{n,11})$, $(C_{n,10}, C_{n,12})$, and $(C_{n,11}, C_{n,12})$, together with (A.1), gives $-5n/96$ for their common value, and the same analysis can be done for e_1, e_2, e_3. Further symmetries are identified by counting the minimal number of steps needed to go from one state to another, and symmetrizing the resulting matrix. This leads to

$$\Sigma = \frac{1}{96}\begin{pmatrix} 19 & -5 & -5 & x & y & x & x & x & y & z & y & y \\ -5 & 19 & -5 & y & z & y & x & x & y & y & x & x \\ -5 & -5 & 19 & x & y & x & y & y & z & y & x & x \\ x & y & x & 19 & -5 & -5 & z & y & y & x & x & y \\ y & z & y & -5 & 19 & -5 & y & x & x & x & x & y \\ x & y & x & -5 & -5 & 19 & y & x & x & y & y & z \\ x & x & y & z & y & y & 19 & -5 & -5 & x & y & x \\ x & x & y & y & x & x & -5 & 19 & -5 & y & z & y \\ y & y & z & y & x & x & -5 & -5 & 19 & x & y & x \\ z & y & y & x & x & y & x & y & x & 19 & -5 & -5 \\ y & x & x & x & x & y & y & z & y & -5 & 19 & -5 \\ y & x & x & y & y & z & x & y & x & -5 & -5 & 19 \end{pmatrix}.$$

Using the relation

$$C_{n,10} + C_{n,11} + C_{n,12} = C_{n,1} + C_{n,6} + C_{n,8},$$

we have $\mathrm{Var}(C_{n,1} + C_{n,6} + C_{n,8}) \sim 9n/32$, which gives $x = -5$. The relation $M_n^X = C_{n,1} + C_{n,4} + C_{n,7} + C_{n,10}$ and the knowledge that $\mathrm{Var}(M_n^X) \sim 2n/3$ then allow the conclusion that $z = 7$. Finally, the fact that the rows sums of Σ are zero gives $y = 1$, and we have the matrix of Lemma 3.

APPENDIX B: PROOF OF THEOREM 1′

To prove Theorem 1′, we extend the result of Kolmogorov [7] to establish a local limit theorem for

$$(\mathcal{N}_4(3n) - 3n/4)/\sqrt{n}.$$

In turn, Kolmogorov's result rests on a local limit theorem for i.i.d. random variables given by Meizler, Parasyuk, and Rvačeva [11]; similar results can be found in, for example, Bhattacharya and Rao [1], p. 237. We follow Kolmogorov's treatment closely and will make frequent reference to his paper; for convenience we will simply refer to it as K(1962). The main idea of the proof uses the fact (first noted by Doeblin) that excursions from a fixed state of a positive recurrent Markov chain are i.i.d. random variables.

Note that in our 12-state Markov chain W, the exit probabilities for states $\{10, 11, 12\}$, i.e., the states comprising e_4, are all the same. This implies that the excursions of the chain from e_4 are i.i.d. For $j = 1, 2, \ldots,$ consider the vector

$$G_j \equiv (\mathcal{N}_4, M^X(j), M^Y(j), M^Z(j)),$$

where each component measures the count in the j^{th} excursion from e_4 by the 12-state Markov chain W. (Thus, the first component in G_j is by definition equal to 1, $M^X(j)$ is the number of X's in the jth excursion, etc.) Then

$$H_k = \sum_{i=1}^{k} G_j$$

is a sum of i.i.d. random vectors, and gives the cumulative count for the first k excursions, with the first component of H_k being identically equal to k. The minimal lattice for the G_j is isomorphic to Z^3, the 3-dimensional integer lattice.

Let $\underline{q} = (1/4, 1/3, 1/3, 1/3)$; the q_i represent the proportion of time the chain spends in $(\mathcal{N}_4, M^X(j), M^Y(j), M^Z(j))$, respectively. Let

$$\underline{y} = \frac{\underline{m} - 4(m_1)\underline{q}}{\sqrt{m_1}},$$

where $\underline{m} = (m_1, m_2, m_3, m_4)$ is a possible value for H_{m_1}, i.e., it has nonnegative integer components and

$$m_1 \leq (m_2 + m_3 + m_4)/2. \tag{B.1}$$

The vector $\underline{y}$ is analogous to (5.6) in K(1962); its first component is identically zero. The local limit theorem of Meizler, Parasyuk, and Rvačeva [11] is now applied to the i.i.d. summands G_j. As in K(1962), Eq. (5.5), we use m_1 as the summation index, as it augments by one for each completed excursion. We have, analogously to (5.5),

$$m_1^{3/2} P(H_{m_1} = \underline{m}) = p^V(\underline{y}) + o(1) \quad \text{as} \quad m_1 \to \infty,$$

where again $\underline{m}$ is a possible value for H_{m_1} and p^V is the Gaussian density corresponding to mean zero and covariance matrix V given by the covariances of the components of G_j (i.e., these are the covariances for a single excursion from e_4; since the first component of G_j is degenerate, we can write this as a 3×3 matrix.) The matrix V may be calculated with modest effort:

$$V = \begin{pmatrix} 14/9 & 2/9 & 2/9 \\ 2/9 & 14/9 & 2/9 \\ 2/9 & 2/9 & 14/9 \end{pmatrix}. \tag{B.2}$$

Then using the fact that m_1 is asymptotically equivalent to $\bar{m}/4$, where $\bar{m} \equiv m_2 + m_3 + m_4$, we get the analogue of (5.8) in K(1962):

$$(\bar{m})^{3/2} P(H_{m_1} = \underline{m}) = 4^{3/2} p^V(\underline{y}) + o(1). \tag{B.3}$$

Now let $\underline{m}$ be a vector (m_1, m_2, m_3, m_4) satisfying $\bar{m} = m_2 + m_3 + m_4 = 3n$, with first component unspecified, subject to (B.1). Following Kolmogorov's notation, let

$$W_1(\underline{m}) \equiv P(\text{after } 3n \text{ stages of the Markov chain, } (\mathcal{N}_4(3n), M_{3n}^X, M_{3n}^Y, M_{3n}^Z) = \underline{m}),$$

and let

$$W_1^1(\underline{m}) \equiv P(\text{after } 3n \text{ stages of the Markov chain, } (\mathcal{N}_4(3n), M_{3n}^X, M_{3n}^Y, M_{3n}^Z) = \underline{m}, \text{ and the chain is in } \varepsilon_4).$$

Thus $W_1^1(\underline{m})$ corresponds to $P(H_{m_1} = \underline{m})$ in our previous notation.

Now we apply (5.1) of K(1962) to extend (7.3) to the case when the chain is not necessarily in e_4 after $3n$ steps, and get

$$\bar{m}^{3/2} W_1(\underline{m}) = 4^{5/2} p^V(\underline{y}) + o(1),$$

with $\underline{y}$ defined as before. This is essentially (5.9) of K(1962).

The remaining task is to express the local limit in terms of the vector

$$\underline{x} = ([m_1 - 3n/4]/\sqrt{3n}, [m_2 - n]/\sqrt{3n}, [m_3 - n]/\sqrt{3n}, [m_4 - n]/\sqrt{3n})$$

and the covariance matrix $n\Lambda$ of $(\mathcal{N}_4(3n), M_{3n}^X, M_{3n}^Y, M_{3n}^Z)$. From previous work, the principal minor of Λ is

$$\begin{pmatrix} 3/32 & 0 & 0 \\ 0 & 2/9 & -1/9 \\ 0 & -1/9 & 2/9 \end{pmatrix}.$$

One verifies that

$$\bar{m}^{3/2}W_1(\underline{m}) = p^{\Lambda}(x_1, x_2, x_3) + o(1),$$

where p^{Λ} is the Gaussian density with mean zero and covariance matrix $n\Lambda$. (This is a consequence of a general result of K(1962), but can easily be verified directly in this case.) This is essentially (5.15) of K(1962). This result provides a local limit theorem for

$$P(\mathcal{N}_4(3n) = k,\, M^X_{3n} = n,\, M^Y_{3n} = n,\, M^Z_{3n} = n),$$

and from this the distribution of $(\mathcal{N}_4(3n) - 3n/4)/\sqrt{n}$, conditional on $\{M^X_{3n} = n, M^Y_{3n} = n\}$, is seen to converge to the same (normal) limit as the unconditional distribution.

Note. For anyone attempting to follow the above analysis, it may be helpful for us to point out that formulas (5.11), (5.12), (5.13), and (5.14) of K(1962) are all stated incorrectly. In (5.11), the factor $1/\sqrt{q^{\gamma}}$ should be $\sqrt{q^{\gamma}}$; in (5.12) the $\sqrt{q^{\gamma}}$ should be $1/\sqrt{q^{\gamma}}$. In (5.13), p^{γ} should be q^{γ}; finally, (5.14) should read

$$(q^{\gamma})^{-(s+1)/2} p_{\gamma}\left[\frac{1}{\sqrt{q^{\gamma}}}(x - x^{\gamma}q/q^{\gamma})\right] = \sqrt{s}p(x).$$

ACKNOWLEDGMENT

We thank the referees for useful remarks, in particular a conjecture leading to Proposition 1.

REFERENCES

[1] R. N. Bhattacharya and R. Ranga Rao, Normal approximation and asymptotic expansions, Wiley, New York, 1976.

[2] P. Billingsley, Probability and measure (2nd edition), Wiley, New York, 1986.

[3] K. L. Chung, Markov chains (2nd edition), Springer-Verlag, New York, 1967.

[4] E. Çinlar, Introduction to stochastic processes, Prentice-Hall, Englewood Cliffs, NJ, 1975.

[5] S. Janson and D. Knuth, Shellsort with three increments, Random Struct Alg 10 (1997), 125–142.

[6] D. Knuth, The art of computer programming 3: Sorting and searching, Addison-Wesley, Reading, MA, 1973.

[7] A. N. Kolmogorov, A local limit theorem for Markov chains, Select Transl Math Stat Probab 2 (1962), 109–129 [transl of Izv Akad Nauk SSSR Ser Mat 13 (1949), 281–300].

[8] J. Lent and H. Mahmoud, On tree-growing search strategies, Ann Appl Probab 6 (1996), 1284–1302.

[9] G. Louchard, Brownian motion and algorithm complexity, BIT 26 (1986), 17–34.

[10] H. M. Mahmoud, Sorting: A distribution theory, Wiley, New York, 2000.

[11] D. G. Meizler, O. S. Parasyuk, and E. L. Rvačeva, A multidimensional local limit theorem of the theory of probability, Dokl Akad Nauk SSSR 60 (1948), 1127–1128. (in Russian)

[12] R. Pyke, Spacings (with discussion), J Roy Stat Soc Ser B 7 (1965), 395–449.

[13] S. M. Ross, Stochastic processes, Wiley, New York, 1983.

[14] R. Sedgewick, Analysis of Shellsort and related algorithms, Algorithms—ESA '96: Fourth Annu Eur Symp Alg, Barcelona, Spain, September 25–27, 1996, J. Diaz, M. Serna, (Editors), Springer-Verlag, Berlin, New York, pp. 1–11.

[15] R. Sedgewick and P. Flajolet, An introduction to the analysis of algorithms, Addison-Wesley, Reading, MA, 1996.

[16] D. Shell, A high-speed sorting procedure, Commun ACM 2 (1959), 30–32.

[17] R. T. Smythe and J. A. Wellner, Stochastic analysis of shell sort, Algorithmica 31 (2001), 442–457.

[18] G. P. Steck, Limit theorems for conditional distributions, Univ Calif Publ Stat 2 (1957), 235–284.