

Research Article

The Distribution of the Size of the Union of Cycles for Two Types of Random Permutations

Tamás Lengyel

Mathematics Department, Occidental College, 1600 Campus Road, Los Angeles, CA 90041, USA

Correspondence should be addressed to Tamás Lengyel, lengyel@oxy.edu

Received 28 August 2010; Accepted 5 October 2010

Academic Editor: Johannes Hattingh

Copyright © 2010 Tamás Lengyel. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We discuss some problems and permutation statistics involving two different types of random permutations. Under the usual model of random permutations, we prove that the shifted coverage of the elements of $\{1, 2, \dots, k\}$ of a random permutation over $\{1, 2, \dots, n\}$; that is, the size of the union of the cycles containing these elements, excluding these elements themselves, follows a negative hypergeometric distribution. This fact gives a probabilistic model for the coverage via the canonical cycle representation. For a different random model, we determine some random permutation statistics regarding the problem of the lost boarding pass and its variations.

1. Introduction

Permutation statistics play an important role in exploring structural properties of random phenomena from a statistical point of view and often do in the context of the analysis of algorithmic applications. We select a permutation $\pi = (\pi(1), \pi(2), \dots, \pi(n))$ of $\{1, 2, \dots, n\}$ at random with $\pi(i)$ denoting the element to which i gets mapped in the permutation. Sometimes, the mapping is also denoted by $i \rightarrow \pi(i)$. Each permutation can be uniquely expressed as the product of disjoint cycles. A cycle is an ordered subset of the permutation whose elements cyclically trade places with one another; for example, the permutation 3241 is the product of the (134) 3-cycle and the (2) 1-cycle. Here, the notation (134) means that starting from the original ordering 1234, the first element is mapped into the third, the third into the fourth, and the fourth into the first, while the second stays unchanged.

Let π_i denote the set of elements in the cycle that contains element i of a random permutation π of $\{1, 2, \dots, n\}$; that is, $\pi_i = \{i, \pi(i), \pi(\pi(i)), \dots\}$. We also use the notation π_S for the (set) union of cycles containing all of the elements in $S \subseteq \{1, 2, \dots, n\}$ (thus, in fact, π_i is a shortcut for $\pi_{\{i\}}$). Sometimes, we write π_i in its cycle form $(i, \pi(i), \pi(\pi(i)), \dots)$ when we want to emphasize the actual cyclical order in the cycle and not only its membership.

In the standard model, all permutations are assumed to be equally likely among the permitted permutations. If all the $n!$ permutations are permitted, then one of the most useful tools in analyzing these permutations is a correspondence known as the canonical cycle representation (although often an alternative version is referred to as canonical cycle form): in the cycle product form of the permutation, write each cycle such that it ends with its least element and order these cycles so that these last elements of the cycles are in increasing order. This correspondence does not seem to match any real life situation. (Had we ordered the cycles in decreasing order or in increasing order but according to their largest elements, then this process models how new records are achieved.) Note that one might benefit from the canonical cycle representation even if the permutations are not equally likely (cf. proof of Theorem 3.9).

Except for Corollary 2.7, membership problems are well known for the standard model (cf. [1], Section 3). We prove that the size of the union of the cycles containing any k prescribed elements excluding these elements themselves follows a negative hypergeometric distribution. For completeness, we also included the well-known results and their short standard proofs in Section 2.

Note that in the second random permutation model, the cycle structure is simple with one nontrivial cycle. The distribution of its size is determined in Remark 3.4 and Theorem 3.6. The number of valid permutations for a generalized version is given in Theorem 3.9.

Theorems 2.3, 2.4, 3.5, 3.6, 3.9, and Corollary 2.7, as well as the distributional observation given in identity (3.5) and its consequences in the second solution of the lost boarding pass problem seem to be new. We also provide a straightforward proof of Theorem 3.2. Identity (3.3) also appeared in [2].

We note that [3] contains a great selection of results regarding random permutation statistics by using tools of analytic combinatorics mainly applied to bivariate generating functions.

Apparently, the second model has the flavor of a statistical resampling method whose purpose is to exchange labels on data points when performing significance tests.

2. The Union of Cycles with Given Membership and Its Cardinality

Now, we discuss membership issues by using the canonical cycle form.

Theorem 2.1. *We have that $P(k \in \pi_1) = P(2 \in \pi_1) = 1/2$ with $2 \leq k \leq n$.*

Proof. Without loss of generality, we answer the question for $2 \in \pi_1$. We use the canonical cycle form defined in Section 1: $P(2 \in \pi_1)$ is related to the position of 2 relative to that of 1. In fact, 2 belongs to the cycle π_1 of element 1 exactly if 2 comes before 1 whose probability is $1/2$. \square

The previous question on $P(2 \in \pi_1)$ has an equivalent form $P(\pi_1 = \pi_2)$. We can generalize this to k cycles about shared membership.

Theorem 2.2. *For $k \geq 2$, we have that*

$$m_k = P(\pi_1 = \pi_2 = \cdots = \pi_k) = \frac{1}{k}, \quad (2.1)$$

and the probability remains the same for any k distinct indices.

Proof. Similarly to the proof of Theorem 2.1, the elements $2, 3, \dots$, and k belong to the cycle of element 1 exactly if $2, 3, \dots$, and k each comes before 1; that is, m_k is the probability that in a random permutation 1 is the last of these numbers which is $1/k$. The probability is not affected by the choice of the indices by symmetry. \square

Now we find, in probabilistic terms, how much of $\{1, 2, \dots, n\}$ is covered by taking the union of the cycles that include any given k elements, for example, $1, 2, \dots, k$.

Theorem 2.3. *For all $m, 1 \leq k \leq m \leq n$, we get that*

$$P_k(m, n) = P(|\pi_{\{1,2,\dots,k\}}(n)| = m) = \frac{k}{m} \frac{\binom{n-k}{n-m}}{\binom{n}{n-m}} = \frac{\binom{m-1}{k-1}}{\binom{n}{k}}. \tag{2.2}$$

With some calculations, we get the expected size and the variance of the size of the coverage in

Theorem 2.4. *We have that*

$$E|\pi_{\{1,2,\dots,k\}}(n)| = \frac{k}{k+1}(n+1), \quad \text{var}(|\pi_{\{1,2,\dots,k\}}(n)|) = \frac{k(n-k)(n+1)}{(k+1)^2(k+2)}. \tag{2.3}$$

For any fixed $k, 1 \leq k \leq n$, as $n \rightarrow \infty$, we get that

$$\text{var}(|\pi_{\{1,2,\dots,k\}}(n)|) \sim \frac{k}{(k+1)^2(k+2)}n^2. \tag{2.4}$$

Remark 2.5. The case $k = 1$ with $P_1(m, n) = 1/n, 1 \leq m \leq n$ is trivial since 1 can go to n equally likely positions in the canonical representation. With the indicator variables $X_i = 1, i = 1, 2, \dots, n$, exactly if i is in the cycle π_1 , we have $X_1 = 1$, since $1 \in \pi_1$ and $P(i \in \pi_1) = 1/2$ if $2 \leq i \leq n$ by Theorem 2.1, and hence $E|\pi_1| = (n+1)/2$. Note that the X_i s are not independent since $m_j \neq m_2^{j-1}$, for $j \geq 3$, by Theorem 2.2. Indeed, as we have just mentioned $|\pi_1| = 1 + \sum_{i=2}^n X_i$ has uniform distribution rather than binomial.

Remark 2.6. Another interesting special case deals with $m = n$, that is, determining the probability $P_k(n, n) = k/n, 1 \leq k \leq n$ that the combined cycles covering k preselected elements cover the whole set $\{1, 2, \dots, n\}$. Surprisingly, the most likely number N of elements covered is n for $k \geq 2$ as we will see it in (2.5) of the proof below.

Proof of Theorem 2.3. We use the canonical cycle form again: we can characterize all permutations with the required property. We start with the case in which $m = n$ (cf. Remark 2.6). The k elements cover the whole set $\{1, 2, \dots, n\}$ exactly if the last element of the permutation in the correspondence is not greater than k which has a chance of k/n .

In the general case, if the union consists of m elements, then the probability of this happening is k/m by the previous case, and the remaining $n - m$ positions can be filled $\binom{n}{n-m}(n - m)!$ ways, out of which $\binom{n-k}{n-m}(n - m)!$ will have the smallest k elements in the union, and this completes the proof of the first part of the identity (2.2).

In order to prove the second part of the theorem, we note that, for $k \geq 2$, the distribution of the random variable $|\pi_{\{1,2,\dots,k\}}(n)|$ is, perhaps somewhat surprisingly, tail

heavy on the right; that is, $P_k(m, n) = P(|\pi_{\{1,2,\dots,k\}}(n)| = m)$ is getting larger as m is getting closer to n . We observe that

$$\frac{P_k(m+1, n)}{P_k(m, n)} = \frac{m}{m+1-k}, \quad k \leq m \leq n-1. \quad (2.5)$$

For $k = 1$, we get that the ratio is 1, and the coverage follows the uniform distribution over $\{1, 2, \dots, n\}$. For $k \geq 2$, the ratio exceeds 1. It is easy to see that

$$P_2(m, n) = \frac{2(m-1)}{n(n-1)}, \quad k \leq m \leq n \quad (2.6)$$

is a linear function in m , and in general, $P_k(m, n)$ as a function of m over the integers $\{k, k+1, \dots, n\}$ is a polynomial of degree $k-1$ by (2.5). More precisely, it follows by (2.5) and $P_k(k, n) = k!(n-k)!/n! = 1/\binom{n}{k}$ that

$$P_k(m, n) = \frac{\binom{m-1}{k-1}}{\binom{n}{k}}, \quad k \leq m \leq n. \quad (2.7)$$

(This fact also guarantees that the sequence $P_k(m, n), m = k, k+1, \dots, n$, is strictly concave up for $k \geq 3$.) \square

It follows that the distribution of the shifted coverage $|\pi_{\{1,2,\dots,k\}}(n)| - k$ belongs to the family of negative hypergeometric distributions. This observation gives a probabilistic model for the coverage via the canonical cycle representation of permutations.

Corollary 2.7. *For the shifted coverage, one has that $|\pi_{\{1,2,\dots,k\}}(n)| - k \sim \text{NegativeHypergeometric}[k, n-k; k]$.*

Now, we present an alternative direct proof of the corollary without using (2.7).

Proof of Corollary 2.7. We have k black and $n-k$ white balls. We select balls without replacement until all black balls have been selected, and, at this point, let Y_k denote the number of white balls selected. Clearly, $Y_k \sim \text{NegativeHypergeometric}[k, n-k; k]$; that is, $P(Y_k = x) = \binom{x+k-1}{k-1} / \binom{n}{k}, 0 \leq x \leq n-k$. With $m = x+k$, we get that $P(Y_k = m-k) = P(Y_k + k = m) = \binom{m-1}{k-1} / \binom{n}{k}, k \leq m \leq n$, for the distribution of the total number of selected balls (including the k black ones). Note that this probability remains the same if we distinguish the balls of the same color; for example, individually label the black and white balls in two separate groups, since it introduces an extra factor $k!(n-k)!$ in both the numerator and denominator. We can find a correspondence between the k black and $n-k$ white balls, and the sets $S_k = \{1, 2, \dots, k\}$ and $\{1, 2, \dots, n\} \setminus S_k$, respectively. First, we index the black balls from 1 to k and the white balls from $k+1$ to n . We set $\pi(1)$ to be the index of the first ball. We also record the color of the ball in every step. If the index is different from 1, then we go on and pick the next ball whose index is to be assigned to $\pi^{(2)}(1) = \pi(\pi(1))$. We keep picking elements until $\pi^{(j)}(1) = 1$ for some j . Let $s = s_1$ be the smallest index in S_k not in $\pi_1 = (\pi(1), \pi^{(2)}(1), \dots, \pi^{(j)}(1) = 1)$. If such an s exists, then we set $\pi(s)$ to be the index of the next ball and keep continuing until the cycle π_s is completed. We set $s = s_2$ to be the smallest index in S_k not in $\pi_1 \cup \pi_{s_1}$. If such an s exists, then we set $\pi(s)$ to be the index of the next ball and keep continuing until π_s is completed, and so forth. The process goes on until no such

s exists. Note that the coverage of $\pi_{\{1,2,\dots,k\}}(n)$ consists of k black balls and $|\pi_{\{1,2,\dots,k\}}(n)| - k$ white balls.

For $k = 1$, $|\pi_1(n)| - 1$ is the number of elements of the first cycle π_1 in the canonical cycle representation before 1 is included; that is, it is the number of white balls before the only black ball is selected. For a general k , $|\pi_{\{1,2,\dots,k\}}(n)| - k$ counts the white balls before the set $\{1, 2, \dots, k\}$ is completely included. \square

3. Some New Results Regarding the Lost Boarding Pass

Now, we discuss a popular problem and some of its variations involving random permutations of a different type. We answer questions similar to those addressed in Section 2 and derive some other properties regarding certain generalizations.

The problem of the lost boarding pass (see [4, page 35]). One hundred people line up to board an airplane, but the first has lost their boarding pass and takes a random seat instead. Each subsequent passenger takes their assigned seat if available, otherwise a random unoccupied seat. What is the probability that the last passenger to board finds their seat occupied?

We assume that the passengers and the seats are ordered and the i th passenger takes the seat $\pi(i)$, $1 \leq i \leq n$. In our original standard model of random permutations, each passenger would randomly pick their seat while here the i th passenger's assigned seat is i but then the first one picks their seat at random, out of necessity. The other passengers make random choices only if their seats are occupied. In a more general context, if r stands for the number of passengers who have lost their boarding passes, then the original model corresponds to $r = n$, while here $r = 1$ with the understanding that the first passenger lost their pass. Remarks 3.3 and 3.7 below make it clear that it is irrelevant which of the first $n - 1$ passengers lost their pass. Of course, if only the last passenger lost their boarding pass, then even they will be properly seated. On the other hand, for larger r , the actual r -element set L of passengers without boarding passes does matter. Note that if the last passenger finds their seat occupied, then they end up at a seat assigned to one of the passengers in L . In summary, this model can be viewed as a generalization of the original one with $r < n$. We focus on the case with $L = \{1\}$, and thus $r = 1$.

Remark 3.1. By (2.1), we would get that $P(\pi_1 = \pi_n) = P(\pi_1 = \pi_2) = m_2 = 1/2$. However, it turns out that the random permutation model used for establishing (2.1) does not work here. In fact, $P(i \in \pi_1)$ changes with i , for example, $P(2 \in \pi_1) = 1/n$ and $P(3 \in \pi_1) = 1/(n - 1)$ as we will see below in the solution (when we calculate q_k). The structure of the cycles is also different from the above random model, and the resulting permutation is far from being typical: it has a long cycle π_1 (of asymptotic average length $\log n$, cf. Remark 3.4), and otherwise, only one-element cycles (fixed points). It is easy to see that there are 2^{n-1} such permutations for $n \geq 1$, since we have $\binom{n-1}{|\pi_1|-1}$ ways to construct the nontrivial cycle with size $|\pi_1| = 1, 2, \dots, n$. In a similar fashion, half of these permutations have the last passenger finding their seat occupied. Note, however, that the permutations occur with very different probabilities depending on not only the size of π_1 but its membership according to Theorem 3.5 (notwithstanding, it also gives the answer 0.50 to the problem of the lost boarding pass by the fact that (3.3) has a last factor of one exactly if the last passenger finds their seat already occupied, and thus, making $P(n \in \pi_1) = P(n \notin \pi_1) = 1/2$).

We find the answer by presenting two approaches. There are other solutions using recurrence relations or other techniques. The "definitive" solution is given in [4, pages 36-37]

although it is of no help with respect to distributional concerns. (Winkler's solution requires no calculations and is based on conditionalizing with two competing targets. It also shows that references to the actual permutation can be completely dropped.) Regarding the case $L = \{1, 2, \dots, r\}$, we also mention an interesting result by Misra [2] and determine the number of valid permutations.

The first approach is based on finding the probability q_k that passenger k finds their seat occupied. We note that the cycle of 1 is quite special: it looks like $\pi_1 = (a_1 = 1, a_2, \dots, a_l)$ with $l = |\pi_1|$, $a_1 = 1 < a_2 < a_3 < \dots < a_l$, $a_{i+1} = \pi(a_i)$, $i = 1, 2, \dots, l - 1$, and $\pi(a_l) = 1$; thus, the elements of the only nontrivial cycle are in a monotone increasing order. (We note that by assuming that a little argument erupts whenever a passenger with a boarding pass finds their seat occupied, then the first passenger can figure out their original seat assignment since their seat is taken by the last upset person. The distribution of the waiting time T for obtaining this information is implicitly discussed in the second solution since $P(T = l) = o_{n-l}$, $2 \leq l \leq n$, and the probability that none of the n passengers get upset is o_n .) One might ask about the chances that the second, the third, and so forth, passengers will find their own seats occupied rather than simply the last one. With $q_k = P(\text{passenger } k \text{ finds their own seat occupied}) = P(k \in \pi_1)$, we get that $q_1 = 0$, $q_2 = P(\text{passenger 1 takes the seat of passenger 2}) = 1/n$, and $q_n = P(a_{|\pi_1|} = n)$. With the notation $i \rightarrow j$ meaning that passenger i takes the seat of passenger j : $q_3 = P((1 \rightarrow 3) \cup (1 \rightarrow 2 \rightarrow 3)) = (1/n) + 1/n(1/n - 1) = 1/n - 1$ and

$$\begin{aligned} q_4 &= P((1 \rightarrow 4) \cup (1 \rightarrow 2 \rightarrow 4) \cup (1 \rightarrow 2 \rightarrow 3 \rightarrow 4) \cup ((1 \rightarrow 3 \rightarrow 4) \cap (2 \rightarrow 2))) \\ &= \frac{1}{n} + \frac{1}{n} \frac{1}{n-1} + \frac{1}{n} \frac{1}{n-1} \frac{1}{n-2} + \frac{1}{n} \frac{1}{n-2} = \frac{1}{n-2}. \end{aligned} \quad (3.1)$$

In a similar way, $q_5 = 1/(n-3)$. More complicated calculations lead to $q_k = 1/(n-k+2)$. Indeed, in general, we encounter exactly the various terms in the expansion of $q_k = (1/n) \prod_{i=1}^{k-2} (1 + (1/n - i)) = 1/(n-k+2)$ by properly conditionalizing with respect to all seat selections leading to $k \in \pi_1$. We have just proved

Theorem 3.2.

$$q_k = P(\text{passenger } k \text{ finds their own seat occupied}) = \frac{1}{n-k+2}, \quad 2 \leq k \leq n. \quad (3.2)$$

In particular, we get $q_n = 1/2$, which provides the solution to the problem.

Remark 3.3. We note that q_k depends on the actual set L of passengers who lost their boarding passes if $r = |L| > 1$. The above derivation assumes that $r = 1$ and the first passenger lost their pass. However, it is easy to see that if it is the j th rather than the first passenger who is without a boarding pass, then we can simply ignore the first $j - 1$ passengers since they are properly seated. With this reduction in the size of the actual problem, the original answer $q_k = 1/(n - k + 2)$ is preserved in Theorem 3.2 for every $j + 1 \leq k \leq n$, since we simply shift both n and k by $j - 1$ to the left. Clearly, $q_1 = q_2 = \dots = q_{j-1} = q_j = 0$. In contrast, if $r > 1$, then things get more complicated. In general, we consider the condition C_L which represents the event that exactly the passengers in L lost their boarding passes. In this case, the calculation of the conditional probability that the k th passenger finds their seat occupied may depend on the choice of L . For example, if $r = n - 1 \geq 1$, then the corresponding conditional probability $q_k(L) = P(\text{passenger } k \text{ finds their own seat occupied} \mid C_L)$ for the last passenger with $k = n$

is $1 - 1/n$ or $1 - 1/(n - 1)$ depending on whether $L = \{1, 2, \dots, n - 1\}$, that is, the first $n - 1$, or $L = \{2, 3, \dots, n\}$, that is, the last $n - 1$ passengers lost their boarding passes. Of course, $n \in L$ makes no sense since the last passenger has no choice and takes the only empty seat left; therefore, we assume that $n \notin L$. Another example with a small r : if $r = |L| = 2$, then $q_3(\{1, a\}) = 2/n, q_3(\{2, b\}) = 1/(n - 1), q_3(\{c, d\}) = 0$ with $a > 1, b > 2, c \geq 3, d > c$, and $n \geq 3$. One might speculate that $q_k(L)$ depends only on an ordered subset of L .

Remark 3.4. Now, we can find the average number of people who pick their seats at random. The average is the harmonic number H_n . In fact, we set the indicator variables $X_k = 1, k = 1, 2, \dots, n$ exactly if k is in the cycle π_1 . We get that $|\pi_1| = 1 + \sum_{k=2}^n X_k$ since $X_1 = 1$, and thus, $E|\pi_1| = 1 + E(\sum_{k=2}^n X_k) = 1 + \sum_{k=2}^n q_k = H_n \sim \log n$.

We can also derive the following theorem on the probability of membership and shared membership, similar to Theorems 2.1 and 2.2 of the original standard random permutation model.

Theorem 3.5. *We have that $P(\pi_1 = \pi_k) = P(k \in \pi_1) = 1/(n - k + 2)$ with $2 \leq k \leq n$. Also, for the full nontrivial cycle, with arbitrary $1 < a_2 < a_3 < \dots < a_l \leq n$, we get that*

$$P(\pi_1 = (a_1 = 1, a_2, \dots, a_l)) = \frac{1}{\prod_{i=1}^l (n - a_i + 1)}. \tag{3.3}$$

It is a trivial lower bound on $P(\pi_1 = \pi_{a_2} = \dots = \pi_{a_l})$, as we might skip some indices of the cycle and l can be less than $|\pi_1|$, and an equality if $a_i = i, 2 \leq i \leq l = |\pi_1|$, that is, $P(\pi_1 = \pi_2 = \dots = \pi_{|\pi_1|}) = 1/n(n - 1) \dots (n - |\pi_1| + 1)$.

In general, for $1 < i_1 < i_2 < \dots < i_j \leq n$ with $j \geq 2$, we have

$$\begin{aligned} P(\pi_1 = \pi_2 = \dots = \pi_{i_j}) &= P(\pi_1 = \pi_{i_1} = \pi_{i_2} = \dots = \pi_{i_j}) \\ &= \frac{1}{\prod_{i=1}^j (n - i_i + 2)}, \end{aligned} \tag{3.4}$$

and thus, the probability depends on the choice of indices.

Proof. The proof is based on the way random selections are made. Identity (3.4) follows by the repeated applications of Theorem 3.2. Indeed, $P(\pi_1 = \pi_{i_1} = \pi_{i_2} = \dots = \pi_{i_j}) = P(\pi_1 = \pi_{i_1})P(\pi_1 = \pi_{i_2-i_1+1} = \pi_{i_3-i_1+1} = \dots = \pi_{i_j-i_1+1}), i_j - i_1 + 1 \leq n - i_1 + 1$, since we can recast the problem as if the nontrivial cycle started with the i_1 th passenger rather than the first one. \square

As a consequence of Remark 3.4 and Theorem 3.5, now we can determine the exact distribution of $|\pi_1|$ since the construction of π_1 corresponds to that of consecutive records (cf. [5]).

Theorem 3.6. *For the distribution of the size of the nontrivial cycle, one gets that $P(|\pi_1| = j) = s(n, j)/n!, 1 \leq j \leq n$, with $s(n, j)$ being the unsigned Stirling number of the first kind indicating the number of permutations over an n -element set with exactly j cycles.*

Proof. We observe that the variables X_i introduced in Remark 3.4 are independent random variables according to (3.4). The proof can be finished by the usual argument used for proving that the positions of the records form an independent sequence of random variables (cf. [6]). Note that for the usual records, position j is that of a record with probability $1/j$. But here, for the X_i s, the order of the indices is reversed between 2 and n . We add that the standard proof that the distribution of the number of records and cycles is the same uses the alternative canonical cycle representation mentioned in the introduction. \square

The second approach is based on calculating the probability o_k that exactly the last k passengers will get their own seats. We can rephrase Remark 3.4 by saying that the number of people getting their own seats is $n - H_n$ on the average. A portion of this number is formed by people with consecutive indices immediately following the largest index $m = a_{|\pi_1|}$ in π_1 . This is the point after which people take their own seats one-by-one and follow the proper seating from then on. In fact, if $m = n - k$, then exactly the last k passengers (without any gap) will get their own seats. Depending on k , we get that

$$\begin{aligned} o_k &= P(\text{exactly the last } k \text{ passengers will get their own seats}) \\ &= P(\text{at least the last } k \text{ passengers will get their own seats}) \\ &\quad - P(\text{at least the last } k + 1 \text{ passengers will get their own seats}) \\ &= \frac{1}{k+1} - \frac{1}{k+2} = \frac{1}{(k+1)(k+2)}, \quad k = 0, 1, \dots, n-2. \end{aligned} \tag{3.5}$$

The reason for $P(\text{at least the last } k \text{ passengers will get their own seats}) = 1/(k+1)$ is that it suffices to focus on π_1 . As this cycle develops, at any stage of the selection, the current passenger faces two competing targets: the seat assigned to the first passenger and the set of seats assigned to the last k passengers. The cycle is either finished by the first choice with probability $1/(k+1)$ or it will move to a seat assigned to one of the last k passengers with probability $k/(k+1)$. We note that a similar argument shows up in the proof in [4] with $k = 1$; however, here, we are interested in the complete distribution $o_k, 0 \leq k \leq n$.

We have that $o_0 = P(n \in \pi_1) = 1/2$ by (3.5) which solves the problem, in agreement with the above result.

We note that clearly, $o_n = P(\pi_1 = (1)) = 1/n$ and $o_0 \neq (n-1)!/n! = 1/n$, for not just the permutation of the single cycle but all permutations which include 1 and n in the same cycle π_1 should be accounted for. Of course, the case with $k = n - 1$ is impossible; thus, $o_{n-1} = 0$. It is worth mentioning that the question regarding another permutation statistic, the maximum element of a cycle is answered by o_k : $P(\max\{j \mid j \in \pi_1\} = k) = P(a_{|\pi_1|} = k) = o_{n-k}$ for $k = 2, \dots, n$ and $P(\max\{j \mid j \in \pi_1\} = 1) = P(a_{|\pi_1|} = 1) = o_n = 1/n$.

The expected number of passengers properly seated beyond the m th passenger defined above is $H_n - 1 + 1/n$. In fact,

$$\begin{aligned} \sum_{k=0}^n k o_k &= n \frac{1}{n} + \sum_{k=1}^{n-2} \frac{k}{(k+1)(k+2)} = 1 + \sum_{k=1}^{n-2} \left(\frac{k+1}{(k+1)(k+2)} - \frac{1}{(k+1)(k+2)} \right) \\ &= 1 + \left(H_n - 1 - \frac{1}{2} \right) - \left(\frac{1}{2} - \frac{1}{n} \right) = H_n - 1 + \frac{1}{n}. \end{aligned} \tag{3.6}$$

Thus, on the average, there are $H_n \sim \log n$ passengers who are seated at random, and there are asymptotically $n - 2 \log n$ passengers who happened to take their seats before the procedure settles to proper seating with asymptotically $\log n$ people involved in it.

Remark 3.7. Note that the answer does not depend on n . Also, we can start anywhere: it is irrelevant which of the first $n - 1$ passengers lost their boarding pass; the answer is simply $P(n \in \pi_i) = 1/2, i < n$, if the i th passenger lost their pass.

With a proof based on equivalences due to exchanging pairs of seats and (3.3), Misra [2] obtained the following result.

Theorem 3.8 (see [2, Theorem 4.1]). *The probability that the last passenger finds their seat when the first r passengers seat themselves at random is $1/(r + 1)$, that is,*

$$1 - q_n(\{1, 2, \dots, r\}) = \frac{1}{r + 1}, \quad 0 \leq r \leq n - 1. \quad (3.7)$$

In other words, we can create the ultimate chaos by letting the r passengers without their boarding passes enter the plane first; thus, $L = \{1, 2, \dots, r\}$ and $q_n(L) = 1 - (1/(r + 1))$ which agrees with our answers based on Theorem 3.2 if $r = 1$ and the ones in Remark 3.3.

Of course, in real life, these passengers should enter last, since this way they will have a better chance to guess their seats correctly. In fact, this group of people will never know what was their original seat assignment. They might not even care about it and simply choose one of the remaining empty seats as it was theirs in the first place; thus, all of them are happily seated at the end. Not to mention that this way the passengers with boarding passes will not be aggravated by finding their seats occupied.

We note that this kind of seating arrangement seem to be easier to analyze. For instance, an even better explanation comes for the permutation count 2^{n-1} , mentioned in Remark 3.1, from applying the canonical cycle representation: for $i > 1$, there are two ways to put i , either before or after 1, corresponding to the options $i \in \pi_1$ and $i \notin \pi_1$, respectively; otherwise, the orders are determined. In a similar fashion, relative to the position of 1 and 2, we find that there are $2 \cdot 3^{n-2}, n \geq 2$, valid permutations if $L = \{1, 2\}$. In general, we get

Theorem 3.9. *For $L = \{1, 2, \dots, r\}$, the number of valid permutations is $r!(r + 1)^{n-r}$ if $n \geq r$.*

Proof. The proof uses the canonical cycle representation. Any ordering of the elements of L uniquely determines the ordering of the other $n - r$ elements once their relative positions with respect to L are chosen. In fact, for each number $i \notin L, 2 \leq i \leq n$, there are $r + 1$ segments relative to the r elements of L to hold i . The segments ending with the elements of L correspond to the cycles, and thus, within the segments, the order of these numbers is set by the nature of the seat selections. \square

Note that this argument does not directly apply to a set L with a gap. For example, if $2 \notin L$, then the cycle $(1x2\dots), x \geq 3$; hence, the ordering $x, 2, \dots, 1$ in the first segment cannot occur.

Acknowledgments

The author wishes to thank Gregory P. Tollisen and the referees for helpful suggestions and comments.

References

- [1] L. Lovász, *Combinatorial Problems and Exercises*, North-Holland, Amsterdam, The Netherlands, 1979.
- [2] N. Misra, "The missing boarding pass," *Resonance*, vol. 13, pp. 662–679, 2008.
- [3] P. Flajolet and R. Sedgewick, *Analytic Combinatorics*, Cambridge University Press, Cambridge, UK, 2009.
- [4] P. Winkler, *Mathematical Puzzles: A Connoisseur's Collection*, A K Peters Ltd., Natick, Mass, USA, 2004.
- [5] D. Collins, T. Lengyel, and G. P. Tollisen, "Keeping scores," *Journal of Statistical Planning and Inference*, vol. 137, no. 7, pp. 2208–2213, 2007.
- [6] G. Blom, L. Holst, and Dennis Sandell, *Problems and Snapshots from the World of Probability*, Springer, New York, NY, USA, 1994.