UNFAIR PERMUTATIONS

HELMUT PRODINGER, CARSTEN SCHNEIDER, AND STEPHAN WAGNER

Abstract. We study *unfair permutations*, which are generated by letting n players draw numbers and assuming that player i draws i times from the unit interval and records her largest value. This model is natural in the context of partitions: the score of the i-th player corresponds to the multiplicity of the summand i in a random partition, with the roles of minimum and maximum interchanged. We study the distribution of several parameters, namely the position of player i, the number of inversions, and the number of ascents. To perform some of the heavy computations, we use the computer algebra package Sigma.

1. Introduction

It is well known that if one draws n times from the uniform distribution (we will say that players $1, 2, \dots, n$ draw numbers) and orders the players according to the numbers they have drawn, from smallest to largest, one gets a random permutation, i. e., every permutation is equally likely. This, of course, remains true if numbers are drawn from an arbitrary continuous probability distribution.

But what happens if the chances are no longer fair? If some players have a better chance to draw a higher number? This will be our point of view in the present note. To be precise, we allow player i to draw i random numbers, and take the best (largest) as her result. Now it is intuitive that higher numbered players tend to have higher results, and thus tend to appear later in the list that goes from smallest to largest. The distribution of permutations is no longer uniform, and $12\dots n$ should be much more likely than $n\,n-1\,\dots 1$.

A simple combinatorial motivation for this particular probabilistic model comes from the theory of partitions: let $p(n)$ be the number of partitions of n. There is an obvious bijection between partitions of $n-rk$ and partitions of n that contain the summand r at least k times. Therefore, if $X_{n,r}$ denotes the multiplicity of the summand r in a random partition of n, we have

$$\mathbb{P}(X_{n,r} \geq k) = \frac{p(n-rk)}{p(n)}.$$

Date: March 22, 2011.

This material is based upon work supported by the National Research Foundation of South Africa under grant number 2053748.

The second author is supported by the Austrian Science Fund (FWF) grants P20162-N18 and P20347-N18.

This material is based upon work supported by the National Research Foundation of South Africa under grant number 70560.

From Rademacher's asymptotic formula for $p(n)$, one obtains

$$\mathbb{P}(X_{n,r} \geq k) = \exp\Big(-\frac{\pi k r}{\sqrt{6n}} + O\Big(\frac{k^2}{n^{3/2}} + \frac{1}{\sqrt{n}}\Big)\Big),$$

uniformly for $k = o(n^{3/4})$ if r is fixed. This shows that the sequence of renormalised random variables $(\pi/\sqrt{6n}\,)X_{n,r}$ converges weakly to an exponential distribution $\mathsf{Exp}(r)$. Since

$$\min(Y_1, Y_2, \dots, Y_r) \sim \mathsf{Exp}(r)$$

if $Y_1, Y_2, \dots, Y_r$ are i.i.d. random variables following an $\mathsf{Exp}(1)$-distribution, we have an immediate correspondence to the model described above (with a minimum in place of a maximum). In the following, we consider several statistics of "unfair" permutations which thus also have natural interpretations for partitions.

To begin with, let us note that the probability for player i to draw the highest number of all players is precisely $i/(n(n+1)/2)$, since she draws i of the $n(n+1)/2$ numbers. More generally, if we want to compute the probability of a specific permutation $\sigma = a_1\, a_2 \dots a_n$ to arise, we find that $a_n/(n(n+1)/2) = a_n/(a_1+a_2+\cdots+a_n)$ is the probability that player a_n draws the highest number, $a_{n-1}/(a_1 + a_2 + \cdots + a_{n-1})$ is the (conditional) probability that player a_{n-1} draws the highest number of the remaining players, etc., so that the probability of the permutation $a_1\, a_2 \dots a_n$ is found to be

$$\prod_{j=1}^{n} \frac{a_j}{\sum_{i=1}^{j} a_i} = \frac{n!}{a_1(a_1 + a_2) \dots (a_1 + \cdots + a_n)}. \tag{1}$$

As an example, we consider the identical permutation with $a_i = i$. Then $a_1 + \cdots + a_i = i(i + 1)/2$, and we get the probability

$$\frac{n!2^n}{n!(n + 1)!} = \frac{2^n}{(n + 1)!}.$$

Likewise, for the reversed permutation with $a_i = n + 1 - i$, the probability is easily found to be

$$\frac{1}{(2n - 1)!!}.$$

The following section deals with the distribution of the position of a given player, for which we determine mean, variance and limit distributions in the cases that i is fixed and that $i \sim \alpha n$ for some constant α. In Sections 3 to 5, we study two classical permutation parameters, namely the number of inversions (or, to be precise, anti-inversions, which is just the number of pairs that are *not* inversions), and the number of ascents. These are but two examples of interesting permutation statistics, many more have been studied in the literature, see for instance the books of Bóna [3] and Stanley [18]; the interested reader is also referred to the recent papers by Dukes [5] on permutation statistics on involutions and by Regev and Roichman [15] on permutation statistics on the alternating group. Variations of our two statistics (ascents/descents and inversions) were studied in yet another recent article of Chebikin [4]. As one can imagine, there are many more parameters to be studied, and we hope that others will find this paper interesting and continue our research.

2. The position of a given player

In the introduction, we started with the probability that i is the highest scoring player. The opposite, i. e., i being the lowest scoring player, is a bit more difficult. If we let the players draw from a uniform distribution on $[0,1]$, then the i-th player's result has distribution function x^i and thus density ix^{i-1}. Hence we obtain

$$\begin{aligned}\mathbb{P}(i \text{ is lowest scoring player}) &= \int_0^1 ix^{i-1}\cdot(1-x^1)\cdots(1-x^{i-1})\cdot(1-x^{i+1})\cdots(1-x^n)\,dx \\ &= \int_0^1 \frac{ix^{i-1}}{1-x^i}\prod_{k=1}^n(1-x^k)dx\end{aligned}$$

for the probability that all other players score more than player i. The fact that the probabilities sum to 1 can be easily seen by noting that

$$\frac{d}{dx}\prod_{k=1}^n(1-x^k) = -\sum_{i=1}^n \frac{ix^{i-1}}{1-x^i}\cdot\prod_{k=1}^n(1-x^k).$$

These probabilities do not depend "too much" on n and indeed tend to limits as $n\to\infty$, which is intuitive. Let us give some numerical values of the limit probabilities

$$p_i = \int_0^1 \frac{ix^{i-1}}{1-x^i}\prod_{k=1}^\infty(1-x^k)dx$$

for $i = 1,2,\ldots,10$:

.51609, .21321, .10731, .059750, .035489, .022072, .014217, .0094162, .0063812, .0044086.

It was found (by means of a different approach) in the recent paper [7] that p_i is precisely the limit probability of the event that i is the summand in a random partition that occurs with largest multiplicity, which agrees with the heuristics given in the introduction (it is not difficult to make the argument precise). It was also shown there that

$$p_i \sim \pi\sqrt{2i}e^{-\pi\sqrt{2i/3}}$$

as $i\to\infty$.

Let us now go one step further and consider the distribution of the i-th player's final rank. Fix i first; the probability generating function for the random variable $R_{n,i}$ defined as the number of players ranked behind player i is then given by

$$r_{n,i}(u) = \sum_{k=0}^{n-1} r_{n,i,k}u^k = \int_0^1 ix^{i-1}\prod_{\substack{j=1\\ j\neq i}}^n \left(1-x^j+ux^j\right)\,dx.$$

Here, $r_{n,i,k}$ is the probability that there are precisely k players ranked behind player i (in particular, $r_{n,i,0}$ is the probability that i is the lowest scoring player). The mean of $R_{n,i}$

can be determined directly: by independence of the random variables in the draw, player i beats player j with probability $i/(i+j)$. Hence, we obtain the expected value

$$\mathbb{E}(R_{n,i}) = \sum_{\substack{j=1\\ j\neq i}}^{n} \frac{i}{i+j} = i(H_{n+i} - H_i) - \frac{1}{2},$$

where $H_k = \sum_{j=1}^{k} 1/j$ denotes the k-th harmonic number. In the following, we will frequently make use of the asymptotic expansion

$$H_k = \log k + \gamma + \frac{1}{2k} + O(k^{-2}),$$

where γ is the Euler-Mascheroni constant, see for instance [6, B.28]. For the variance, a different approach is necessary since the events "player i beats player j", $j = 1, 2, \dots, n$, are not independent. Hence we differentiate the probability generating function with respect to u and obtain

$$\mathbb{V}(R_{n,i}) = r''_{n,i}(1) + r'_{n,i}(1) - \mathbb{E}(R_{n,i})^2 = r''_{n,i}(1) + \mathbb{E}(R_{n,i}) - \mathbb{E}(R_{n,i})^2.$$

For the first term, we get

$$\begin{aligned}
r''_{n,i}(1) &= \int_0^1 ix^{i-1}\Bigg(\Big(\sum_{\substack{j=1\\ j\neq i}}^{n} x^j\Big)^2 - \sum_{\substack{j=1\\ j\neq i}}^{n} x^{2j}\Bigg)\,dx \\
&= \int_0^1 ix^{i-1}\Bigg(\sum_{j=2}^{2n}(n-|n+1-j|)x^j - \sum_{j=1}^{n} 2x^{i+j} - \sum_{j=1}^{n} x^{2j} + 2x^{2i}\Bigg)\,dx \\
&= \sum_{j=2}^{2n} \frac{i(n-|n+1-j|)}{i+j} - \sum_{j=1}^{n} \frac{2i}{2i+j} - \sum_{j=1}^{n} \frac{i}{i+2j} + \frac{2}{3} \\
&= i(i+1)H_i + 2iH_{2i} - 2i(n+i+1)H_{n+i} - 2iH_{n+2i} + i(2n+i+1)H_{2n+i} \\
&\quad + \frac{2}{3} - i\sum_{j=1}^{n} \frac{1}{i+2j}.
\end{aligned}$$

Hence we finally have the following theorem:

Theorem 1. *The mean and variance of the number of players ranked behind player i are*

$$\begin{aligned}
\mathbb{E}(R_{n,i}) &= i\,(H_{n+i} - H_i) - \frac{1}{2}, \\
\mathbb{V}(R_{n,i}) &= i(i-1)H_i + 2iH_{2i} - 2i(n+i)H_{n+i} - 2iH_{n+2i} + i(2n+i+1)H_{2n+i} \\
&\qquad - i^2(H_{n+i} - H_i)^2 - \frac{1}{12} - i\sum_{j=1}^{n} \frac{1}{i+2j}.
\end{aligned}$$

For fixed i, the probability generating function tends to a limit:

$$r_i(u) = \lim_{n\to\infty} r_{n,i}(u) = \int_0^1 \frac{ix^{i-1}}{1-x^i+ux^i} \prod_{j=1}^{\infty} \left(1 - x^j + ux^j\right)\, dx,$$

which is again the probability generating function of a discrete distribution (in particular, the probability that i is the lowest scoring player is $p_i = r_i(0)$). Hence the sequence $R_{n,i}$ of random variables converges weakly to a discrete limit ([6, Theorem IX.1]). This can be interpreted in terms of partitions as well: if we rank the summands of a random partition by their multiplicity, then the number of summands with larger multiplicity than i follows the distribution given by the probability generating function $r_i(u)$ (in the limit).

The situation becomes more interesting if we let i grow with n, i. e., $i \sim \alpha n$ for some $\alpha \in (0,1]$. It is then easy to see that mean and variance are of order n and n^2 respectively, and so we study the normalised random variable $N_{n,i} = n^{-1}R_{n,i}$, which turns out to converge to a limit distribution:

Theorem 2. *Let $R_{n,i}$ be the number of players ranked behind player i, and $N_{n,i}$ the normalised random variable $n^{-1}R_{n,i}$. If $i \sim \alpha n$, then $N_{n,i}$ converges weakly to a random variable with density*

$$f(z) = -\frac{y \log y}{z - y^{1/\alpha}}, \qquad 0 \le z \le 1,$$

where y is given implicitly by

$$z = -\frac{\alpha(1-y^{1/\alpha})}{\log y}.$$

Proof. The moment generating function of the random variable $N_{n,i}$ is given by

$$\begin{aligned} G_{n,i}(t) = \mathbb{E}(e^{tN_{n,i}}) = r_{n,i}(e^{t/n}) &= \int_0^1 ix^{i-1} \prod_{\substack{j=1\\ j\ne i}}^{n} \left(1 - x^j + e^{t/n}x^j\right)\, dx \\ &= \int_0^1 \prod_{\substack{j=1\\ j\ne i}}^{n} \left(1 - y^{j/i} + e^{t/n}y^{j/i}\right)\, dy \\ &= \int_0^1 \exp\Bigg(\sum_{\substack{j=1\\ j\ne i}}^{n} \log\left(1 - y^{j/i} + e^{t/n}y^{j/i}\right)\Bigg)\, dy. \end{aligned}$$

We fix t and let n tend to infinity. Then

$$\begin{aligned} G_{n,i}(t) &= \int_0^1 \exp\Bigg(\sum_{\substack{j=1\\ j\ne i}}^{n} \log\Big(1 + \frac{t}{n}y^{j/i} + O(n^{-2})\Big)\Bigg)\, dy \\ &= \int_0^1 \exp\Bigg(\frac{t}{n}\sum_{j=1}^{n} y^{j/i} + O(n^{-1})\Bigg)\, dy, \end{aligned}$$

where the O-term is uniform in y. Recall that $i \sim \alpha n$, which implies

$$\frac{t}{n}\sum_{j=1}^{n} y^{j/i} = \frac{t}{n}\cdot\frac{1-y^{n/i}}{y^{-1/i}-1} \overset{n\to\infty}{\longrightarrow} -\frac{\alpha t(1-y^{1/\alpha})}{\log y}.$$

However, we need uniformity of this limit in order to interchange limit and integral. Suppose that n is large enough so that $|\alpha n/i - 1| \le \epsilon$. Then

$$\begin{aligned}\frac{1}{n}\sum_{j=1}^{n}\left(y^{j/i} - y^{j/(\alpha n)}\right) &= \frac{1}{n}\sum_{j=1}^{n} y^{j/(\alpha n)}\left(y^{j(\alpha n/i-1)/(\alpha n)}-1\right)\\ &\ge \frac{1}{n}\sum_{j=1}^{n} y^{j/(\alpha n)}\left(y^{j\epsilon/(\alpha n)}-1\right) \ge -\epsilon(1+\epsilon)^{-1-1/\epsilon},\end{aligned}$$

since $-\epsilon(1+\epsilon)^{-1-1/\epsilon}$ is the minimum of the function $u \mapsto u(u^\epsilon - 1)$. An analogous upper estimate holds as well. Hence, if we replace the sum over $y^{j/i}$ by the sum over $y^{j/(\alpha n)}$, the difference tends to 0 uniformly in y. Now we recognize the latter sum as a Riemann sum and obtain

$$\int_0^1 y^{u/\alpha}\,du \ge \frac{1}{n}\sum_{j=1}^{n} y^{j/(\alpha n)} \ge \int_0^1 y^{u/\alpha}\,du - \frac{1}{n}.$$

So we have uniform convergence to

$$t\cdot\int_0^1 y^{u/\alpha}\,du = -\frac{\alpha t(1-y^{1/\alpha})}{\log y}$$

and thus

$$\lim_{n\to\infty} G_{n,i}(t) = \int_0^1 \exp\Big(-\frac{\alpha t(1-y^{1/\alpha})}{\log y}\Big)\,dy.$$

If we finally perform the substitution

$$z = -\frac{\alpha(1-y^{1/\alpha})}{\log y},$$

then we end up with

$$\lim_{n\to\infty} G_{n,i}(t) = \int_0^1 e^{tz}\cdot\frac{y\log y}{y^{1/\alpha}-z}\,dz,$$

and the convergence is uniform if t is restricted to a compact interval. The above integral is precisely the moment generating function of a random variable with density

$$f(z) = -\frac{y\log y}{z-y^{1/\alpha}},$$

which finally proves that the sequence of normalised random variables $N_{n,i} = n^{-1}R_{n,i}$ converges weakly to a random variable with the rather unusual density $f(z)$ (see [6, Theorem IX.4]) and therefore completes the proof of our theorem. □

3. The number of anti-inversions

An inversion in a permutation is a pair $i < j$ such that $a_i > a_j$ (player i beats player j). Equivalently, one can consider anti-inversions, i. e., pairs $i < j$ such that player j beats player i. We expect that in our setting the number of anti-inversions should be higher than the number of inversions. For ordinary permutations, these numbers are obviously the same on average; inversions are a very classical permutation statistic, see e. g. [10]. The number of inversions is extremely important in the study of sorting algorithms; for its distribution compare also [12]. There is also an interesting relation between edge weights in recursive trees and inversions of permutations, as shown by Kuba and Panholzer [11]. Our aim in this section is to prove the following theorem:

Theorem 3. *Let A_n be the number of anti-inversions in a random unfair permutation of $\{1, 2, \ldots, n\}$. Then the mean and variance of A_n are given by*

$$\mathbb{E}(A_n) = \frac{(2n+1)^2}{8}(H_{2n} - H_n) + \frac{1}{16}H_n - \frac{5n}{8} \tag{2}$$

and

$$\begin{aligned}\mathbb{V}(A_n) = {} & \frac{n(29+126n+72n^2)}{216} + \frac{35+108n+81n^2-27n^3}{162}H_n \\ & + \frac{-3-16n-10n^2+8n^3}{12}H_{2n} + \frac{-16+27n-54n^3}{108}H_{3n} \\ & + \frac{n(1+3n+2n^2)}{6}\Big(3H_{2n}^{(2)} - 2H_n^{(2)} + 4\sum_{1\le i\le 2n}\frac{(-1)^i H_i}{i}\Big) \\ & + \frac{8}{27}\sum_{i=1}^{n}\frac{1}{3i-2} + \frac{(-1)^n n}{4}\Big(\sum_{i=1}^{n}\frac{(-1)^i}{i} - \sum_{i=1}^{3n}\frac{(-1)^i}{i}\Big),\end{aligned} \tag{3}$$

where $H_k = \sum_{j=1}^{k} 1/j$ and $H_k^{(2)} = \sum_{j=1}^{k} 1/j^2$ denote harmonic numbers and second-order harmonic numbers respectively. The normalised random variable $\frac{A_n - \mathbb{E}(A_n)}{\sqrt{\mathbb{V}(A_n)}}$ converges weakly to a standard normal distribution.

Remark. Asymptotically, the expected value is

$$\mathbb{E}(A_n) = \frac{\log 2}{2}n^2 + \Big(\frac{\log 2}{2} - \frac{3}{4}\Big)n + O(\log n) = 0.3465735903n^2 - 0.4034264097n + O(\log n),$$

which should be compared to the fair case, in which the average number of anti-inversions is smaller (as expected):

$$\frac{n(n-1)}{4} = 0.25n^2 - 0.25n.$$

The variance, on the other hand, is smaller for unfair permutations (which is also intuitive, since fair permutations are "more random" in a certain sense): making use of the identities

$\sum_{j=1}^{\infty} 1/j^2 = \pi^2/6$ and

$$\begin{aligned}\sum_{j=1}^{\infty} \frac{(-1)^j H_j}{j} &= \int_0^1 \sum_{j=1}^{\infty} (-1)^j H_j x^{j-1}\,dx = -\int_0^1 \frac{\log(1+x)}{x(1+x)}\,dx \\ &= \int_0^1 \left(\frac{\log(1+x)}{1+x} - \frac{\log(1+x)}{x}\right) dx = \frac{(\log 2)^2}{2} - \frac{\pi^2}{12},\end{aligned}$$

(the last identity is covered, e.g., by the general formula [13, p.50,Eq.6] up to some synchronization) we obtain

$$\mathbb{V}(A_n) \sim \left(\frac{1}{3} - \frac{\pi^2}{18} + \frac{2\log 2}{3} - \frac{\log 3}{2} + \frac{2\log^2 2}{3}\right) n^3 \sim 0.01811 n^3,$$

as compared to the classical case, where the variance is

$$\frac{n(2n+5)(n-1)}{72} \sim 0.02777 n^3.$$

Note that mean and variance (and the central limit theorem) are somewhat easier to obtain in the classical case, since the random number of (anti-)inversions can be written as a sum of n independent random variables, distributed uniformly on $\{0, 1, \dots, i-1\}$ for $i = 1, 2, \dots, n$. See [12, 16] for details.

Proof. Let us use an indicator variable $X_{i,j}$, which is 1 if j beats i, and 0 otherwise. As noted earlier, $\mathbb{E}(X_{i,j}) = j/(i+j)$ is the probability that j beats i.

Now the expected value of anti-inversions is

$$\begin{aligned}\mathbb{E}(A_n) &= \sum_{1\le i<j\le n} \frac{j}{i+j} \\ &= \sum_{1\le j\le n} j(H_{2j-1} - H_j) \\ &= \sum_{1\le j\le n} jH_{2j} - \frac{n}{2} - \sum_{1\le j\le n} jH_j.\end{aligned}$$

The formula

$$\sum_{1\le j\le n} jH_j = \binom{n+1}{2} H_n - \frac{n(n-1)}{4}$$

is classical [9], and

$$\begin{aligned}\sum_{1\le j\le n} jH_{2j} &= \sum_{1\le j\le n} j \sum_{1\le i\le j} \left[\frac{1}{2i} + \frac{1}{2i-1}\right] \\ &= \sum_{1\le i\le n} \left[\frac{1}{2i} + \frac{1}{2i-1}\right] \sum_{i\le j\le n} j\end{aligned}$$

$$
\begin{aligned}
&= \sum_{1\le i\le n} \Big[\frac{1}{2i}+\frac{1}{2i-1}\Big]\Big[\binom{n+1}{2}-\binom{i}{2}\Big]\\
&= \binom{n+1}{2}H_{2n} - \frac14 \sum_{1\le i\le n}(i-1) + \sum_{1\le i\le n}\Big[-\frac{i}{4}+\frac18+\frac{1}{8(2i-1)}\Big]\\
&= \binom{n+1}{2}H_{2n} - \frac{n(2n-1)}{8} + \frac18\Big(H_{2n}-\frac12 H_n\Big).
\end{aligned}
$$

Summarizing, we get the desired formula (2).

The computation of the variance is much more involved. In principle, it could be done in the same way, but the calculations are very lengthy, and so we decided to use the summation toolbox Sigma.

By definition, the variance is given by

$$\mathbb{V}(A_n) = \mathbb{E}\Big(\sum_{1\le i<j\le n} X_{i,j}\cdot \sum_{1\le k<l\le n} X_{k,l}\Big) - \sum_{1\le i<j\le n}\mathbb{E}(X_{i,j})\cdot \sum_{1\le k<l\le n}\mathbb{E}(X_{k,l}).$$

If all indices are distinct, then the random variables $X_{i,j}$ and $X_{k,l}$ are independent, and the corresponding terms cancel out. Therefore

$$
\begin{aligned}
\mathbb{V}(A_n) = 2 &\sum_{1\le i<j<k\le n}\mathbb{E}(X_{i,j}\cdot X_{j,k}) + 2\sum_{1\le i<j<k\le n}\mathbb{E}(X_{i,k}\cdot X_{j,k})\\
+ 2 &\sum_{1\le i<j<k\le n}\mathbb{E}(X_{i,j}\cdot X_{i,k}) + \sum_{1\le i<j\le n}\mathbb{E}(X_{i,j}\cdot X_{i,j})\\
- 2 &\sum_{1\le i<j<k\le n}\mathbb{E}(X_{i,j})\cdot\mathbb{E}(X_{j,k}) - 2\sum_{1\le i<j<k\le n}\mathbb{E}(X_{i,k})\cdot\mathbb{E}(X_{j,k})\\
- 2 &\sum_{1\le i<j<k\le n}\mathbb{E}(X_{i,j})\cdot\mathbb{E}(X_{i,k}) - \sum_{1\le i<j\le n}\mathbb{E}(X_{i,j})\cdot\mathbb{E}(X_{i,j}).
\end{aligned}
$$

Each of the terms can be easily determined combinatorially: for instance, $\mathbb{E}(X_{i,j}\cdot X_{j,k})$ is precisely the probability that the results x_i, x_j, x_k of players i, j, k satisfy $x_i < x_j < x_k$. This is independent of all other players, and so the argument that leads to (1) also shows that

$$\mathbb{E}(X_{i,j}\cdot X_{j,k}) = \frac{kj}{(i+j)(i+j+k)}.$$

Likewise, one obtains

$$
\begin{aligned}
&\mathbb{E}(X_{i,k}\cdot X_{j,k}) = \frac{k}{i+j+k},\\
&\mathbb{E}(X_{i,j}\cdot X_{i,k}) = \mathbb{E}(X_{i,j}\cdot X_{j,k}) + \mathbb{E}(X_{i,k}\cdot X_{k,j}) = \frac{kj}{(i+j)(i+j+k)} + \frac{kj}{(i+k)(i+j+k)},
\end{aligned}
$$

and $\mathbb{E}(X_{i,j} \cdot X_{i,j}) = \mathbb{E}(X_{i,j}) = j/(i+j)$. Plugging all these formulas in, we can write the variance as

$$\begin{aligned}\mathbb{V}(A_n) = 2 \sum_{1\le i<j<k\le n} \frac{kj}{(i+j)(i+j+k)} + 2 \sum_{1\le i<j<k\le n} \frac{k}{i+j+k} \\ + 2 \sum_{1\le i<j<k\le n} \frac{kj}{(i+j)(i+j+k)} + 2 \sum_{1\le i<j<k\le n} \frac{kj}{(i+k)(i+j+k)} + \sum_{1\le i<j\le n} \frac{j}{i+j} \\ - 2 \sum_{1\le i<j<k\le n} \frac{j}{i+j} \cdot \frac{k}{j+k} - 2 \sum_{1\le i<j<k\le n} \frac{k}{i+k} \cdot \frac{k}{j+k} \\ - 2 \sum_{1\le i<j<k\le n} \frac{j}{i+j} \cdot \frac{k}{i+k} - \sum_{1\le i<j\le n} \frac{j^2}{(i+j)^2}.\end{aligned}$$

The sum $\sum_{1\le i<j\le n} j/(i+j)$ has already been dealt with in the derivation of (2). The others are tedious to handle by hand, but the package Sigma is able to compute all these sums; the algorithm that is used for these purposes is outlined in the following section. We have, for instance,

$$\begin{aligned}\sum_{1\le i<j<k\le n} \frac{k}{i+j+k} = \frac{115n}{216} + \frac{-648n^3 - 486n^2 + 162n - 1}{1296} H_n \\ + \frac{216n^3 + 162n^2 - 54n - 53}{432} H_{3n} + \frac{(-1)^n(2n+1)}{16} \left(\sum_{i=1}^{n} \frac{(-1)^i}{i} - \sum_{i=1}^{3n} \frac{(-1)^i}{i} \right) \\ + \frac{(n+1)(2n+1)n}{3} \sum_{i=1}^{2n} \frac{(-1)^i}{i} + \frac{1}{27} \sum_{i=1}^{n} \frac{1}{3i-2},\end{aligned} \tag{4}$$

and similar formulas can be found for all the other sums. Putting everything together, we find the exact expression (3) for the variance of A_n.

Let us finally consider the limit distribution of A_n. It is well known (see for instance [16]) that the limit distribution in the classical case is Gaussian. The approach that is usually used to prove this result, which takes advantage of the fact that the number of inversions can be represented as a sum of independent random variables, is not applicable to our situation. However, it is possible to use a similar, yet slightly different approach: recall that the random variable $X_{i,j}$ is defined as the indicator variable of the event "j beats i". Define the shifted variable $Y_{i,j} = X_{i,j} - \mathbb{E}(X_{i,j}) = X_{i,j} - j/(i+j)$, so that

$$A_n - \mathbb{E}(A_n) = \sum_{1\le i<j\le n} Y_{i,j}.$$

The variables $Y_{i,j}$ are not all independent, but they almost are: $Y_{i,j}$ and $Y_{k,l}$ are independent whenever $\{i,j\} \cap \{k,l\} = \emptyset$. Such a set of random variables is called *dissociated*, and very general limit theorems on sums of dissociated random variables are known. Specifically, we are going to use the following result, which is a special case of [1, Theorem 2.1]:

Theorem 4. *Suppose that $Z_{i,j}$ $(1 \le i < j)$ are random variables such that $Z_{i,j}$ and $Z_{k,l}$ are independent whenever $\{i,j\} \cap \{k,l\} = \emptyset$. Furthermore, assume that $\mathbb{E}(Z_{i,j}) = 0$ and $\mathbb{E}(|Z_{i,j}|^3) < \infty$ for all i,j. Let σ_n^2 be the variance of the sum*

$$W_n = \sum_{1 \le i < j \le n} Z_{i,j}.$$

Then the sequence $\sigma_n^{-1} W_n$ of normalised random variables converges weakly to the standard normal distribution, provided that

$$\epsilon_n = \sigma_n^{-3} \sum_{1 \le i < j \le n} \mathbb{E}(|Z_{i,j}|^3)^{1/3} \Bigg(\sum_{\substack{1 \le k < l \le n \\ \{i,j\} \cap \{k,l\} \neq \emptyset}} \mathbb{E}(|Z_{k,l}|^3)^{1/3} \Bigg)^2 \to 0$$

as $n \to \infty$.

It is easy to see that all technical conditions are satisfied for the variables $Y_{i,j}$ (in particular, $\mathbb{E}(|Y_{i,j}|^3) \le 1$), and that

$$\begin{aligned} \epsilon_n &\le \sigma_n^{-3} \sum_{1 \le i < j \le n} \Bigg(\sum_{\substack{1 \le k < l \le n \\ \{i,j\} \cap \{k,l\} \neq \emptyset}} 1 \Bigg)^2 \\ &= \sigma_n^{-3} \sum_{1 \le i < j \le n} (2n-3)^2 \ll \sigma_n^{-3} n^4 \ll n^{-1/2}, \end{aligned}$$

since the variance is of order n^3, as shown above. It follows that the number of anti-inversions (and thus also the number of inversions), suitably normalised, converges weakly to a Gaussian distribution. □

4. Simplifying sums with Sigma

The Mathematica package Sigma [17] can be used to simplify multi-sums with the symbolic summation paradigms of telescoping, creative telescoping and recurrence solving. The underlying algorithms based on our refined difference field theory of Karr's ΠΣ-fields [8] do not only work for hypergeometric terms [14], but for rational expressions in terms of indefinite nested sums and products. In the following we illustrate this approach by carrying out the summation steps for the triple sum (4), i. e., for the sum

$$\sum_{1 \le i < j < k \le n} \frac{k}{i+j+k} = \sum_{k=1}^{n} \sum_{j=1}^{k-1} \sum_{i=1}^{j-1} \frac{k}{i+j+k}.$$

Here the overall tactic is to attack the sums from the innermost sum $F_1(k,j) = \sum_{i=1}^{j-1} k/(i+j+k)$ over the middle sum $F_2(k) = \sum_{j=1}^{k-1} F_1(k,j)$ to the outermost sum $F_3(n) = \sum_{k=1}^{n} F_2(k)$, and to eliminate as many summation quantifiers as possible. While processing one of these quantifiers, say $\sum_{j=1}^{k-1} F_1(k,j)$, in the setting of ΠΣ-fields, the following preparation is crucial: the occurring sums in $F_1(k,j)$ have to be represented in indefinite nested form w.r.t. the summation index j. For simplicity, we ignore indefinite nested products and we assume that the expressions under consideration are well defined for evaluations at all nonnegative integers. Then the arising building blocks can be defined recursively as follows:

(i) Any rational function $f(j) \in \mathbf{Q}(j)$ which does not introduce a pole for evaluations at nonnegative integers is an indefinite nested sum.
(ii) If $f(j)$ is an indefinite nested sum and a, b are integers with $a > 0$, then also $\sum_{i=1}^{aj+b} f(i)$ (j is replaced in f by a new variable i) is an indefinite nested sum.
(iii) If $f(j)$ and $g(j)$ are indefinite nested sums, then $f(j)+g(j)$ and $f(j)g(j)$ are indefinite nested sums. In particular, if $f(j)$ is nonzero for any evaluation at nonnegative integers, then also $1/f(j)$ is an indefinite nested sum.

Note that the inner sum $F_1(j) = \sum_{i=1}^{j-1} \frac{k}{i+j+k}$ is not indefinite nested, since j occurs inside of the summand and thus violates part (ii) of our definition. In order to transform $F_1(j)$ to a representation in indefinite nested format, we compute first a recurrence relation of $F_1(k, j)$ in j:

$$F_1(k, j+1) - F_1(k, j) = -\frac{k\big(1 + 3j + 2k + 2jk + k^2\big)}{(1 + j + k)(2j + k)(1 + 2j + k)}. \tag{5}$$

Internally, Sigma follows Zeilberger's creative telescoping paradigm [14]. For a given $d \geq 1$ and the given summand $F_0(k, j, n)$, it searches for constants $c_0(k, j), \dots, c_d(k, j)$, free of the summation index i, and for a suitable expression $g(k, j, i)$ such that the following summand recurrence holds:

$$c_0(k, j)F_0(k, j, i) + \cdots + c_d(k, j)F_0(k, j + d, i) = g(k, j, i + 1) - g(k, j, i). \tag{6}$$

In our particular example, Sigma is successful with $d = 1$ and finds $c_0(k, j) = 1$, $c_1 = -1$, and $g(k, j, i) = -k/(i+j+k)$. Finally, summing this equation (6) over i from 1 to $i-1$ yields the recurrence (5) for $F_1(k, j)$. Note that the correctness of the summand recurrence (6) can be easily verified and thus also the recurrence relation (5) is implied.

In the next step, Sigma solves the recurrence and generates the general solution

$$c + k\Bigg(\sum_{r=1}^{j} \frac{1}{k + 2r - 2} - \sum_{r=1}^{j} \frac{1}{k + r} + \sum_{r=1}^{j} \frac{1}{k + 2r - 1}\Bigg)$$

for c being a constant, i. e., not depending on j. Setting $c = -1$, the derived expression and $F_1(k, j)$ agree for $j = 1$. And since both expressions are a solution of the first order recurrence (5), they are equal for all $j \geq 1$, i. e.,

$$F_1(k, j) = -1 + k\Bigg(\sum_{r=1}^{j} \frac{1}{k + 2r - 2} - \sum_{r=1}^{j} \frac{1}{k + r} + b\sum_{r=1}^{j} \frac{1}{k + 2r - 1}\Bigg); \tag{7}$$

in particular, the identity holds for $k, j \geq 1$. Summarizing, we succeeded in transforming the input sum $F_1(k, j)$ to an expression where j does not occur inside of any summand. Moreover, the correctness of all the computations can be verified by simple polynomial arithmetic, and we obtain a rigorous proof of identity (7). In this form, we are ready to deal with the next sum $F_2(k)=\sum_{j=1}^{k-1}F_1(k,j)$. In this particular instance, Sigma finds

$$g(k,j)=-j+1-k(j+k)\sum_{r=1}^{j}\frac{1}{k+r}+\frac{k(2j+k-2)}{2}\sum_{r=1}^{j}\frac{1}{k+2r-2}+\frac{k(2j+k-1)}{2}\sum_{r=1}^{j}\frac{1}{k+2r-1}$$

as a solution for the telescoping equation $g(k,j+1)-g(k,j)=F_1(k,j)$. Hence summing this equation over j from 1 to $k-1$ gives

$$F_2(k)=\sum_{j=1}^{k-1}F_1(k,j)=-k+1-2k^2\sum_{r=1}^{k-1}\frac{1}{k+r}+\frac{3(k-2)k}{2}\sum_{r=1}^{k-1}\frac{1}{k+2r-2}+\frac{(3k-1)k}{2}\sum_{r=1}^{k-1}\frac{1}{k+2r-1}. \tag{8}$$

In order to handle the last summation quantifier, Sigma has to transform the occurring sums of (8) in terms of indefinite nested sums and products w.r.t. k. Here we follow exactly the same strategy as above. By telescoping, creative telescoping, and recurrence solving we find the following alternative representations:

$$\sum_{r=1}^{k-1}\frac{1}{k+r}=-H_k+H_{2k}-\frac{1}{2k},$$

$$\sum_{r=1}^{k-1}\frac{1}{k+2r-2}=-\frac{H_k}{2}+\frac{H_{3k}}{2}-(-1)^k\frac{1}{2}\left(\sum_{r=1}^{k}\frac{(-1)^r}{r}-\sum_{r=1}^{3k}\frac{(-1)^r}{r}\right)+\frac{3k-4}{3k(3k-2)},$$

$$\sum_{r=1}^{k-1}\frac{1}{k+2r-1}=-\frac{H_k}{2}+\frac{H_{3k}}{2}+(-1)^k\frac{1}{2}\left(\sum_{r=1}^{k}\frac{(-1)^r}{r}-\sum_{r=1}^{3k}\frac{(-1)^r}{r}\right)+\frac{1}{1-3k}.$$

Using this information we find

$$F_2(k)=-2H_{2k}k^2+\frac{k}{4}(2k+3)H_k+\frac{3(2k-1)k}{4}H_{3k}+(-1)^k\frac{k}{4}\left(\sum_{r=1}^{k}\frac{(-1)^r}{r}-\sum_{r=1}^{3k}\frac{(-1)^r}{r}\right)+\frac{1}{3},$$

and we are ready to deal with the sum $F_3(n)$. In this case, Sigma finds the solution

$$g(k)=\frac{115k-72}{216}+\frac{216k^3+162k^2-378k-1}{1296}H_k-\frac{(k-1)k(2k-1)}{3}H_{2k}+\frac{1}{27}\sum_{r=1}^{k}\frac{1}{3r-2}+\frac{216k^3-486k^2+270k-53}{432}H_{3k}$$

$$+\frac{(-1)^k(2k-1)}{16}\Big(\sum_{r=1}^{3k}\frac{(-1)^r}{r}-\sum_{r=1}^{k}\frac{(-1)^r}{r}\Big)$$

for the telescoping equation $g(k+1)-g(k)=F_2(k)$. To this end, summing this equation over k from 1 to n produces the right hand side of (4). Once more we emphasize that the computation steps for the identity we found can be verified independently by simple polynomial arithmetic. We remark further that the harmonic sums – truncated versions of multiple-zeta-values – appearing in the closed-form solution (4) also arise frequently in particle physics [2, 19].

5. The number of ascents

Let U_i be the indicator random variable that is 1 if the left neighbour of i in a random unfair permutation is smaller than i, and 0 otherwise. The sum over all U_i is the random variable "number of ascents".

Let us consider the probability that the left neighbour of i is j for fixed i and j. If player i's best result is y and player j's best result is x, then nobody else has a best result in the range $[x,y]$, and so we find the following formula for this probability:

$$\iint\limits_{0\le x<y\le 1} iy^{i-1}jx^{j-1}\frac{1}{(1-y^j+x^j)(1-y^i+x^i)}\prod_{k=1}^{n}(1-y^k+x^k)dx\,dy.$$

Summing over all pairs (i,j), we get the expected number of ascents:

$$\sum_{1\le j<i\le n}\iint\limits_{0\le x<y\le 1}\frac{iy^{i-1}jx^{j-1}}{(1-y^j+x^j)(1-y^i+x^i)}\prod_{k=1}^{n}(1-y^k+x^k)dx\,dy.$$

We make use of this integral representation to prove the following theorem:

Theorem 5. *The average number of ascents in an unfair permutation of n is*

$$\frac{n}{2}+\frac{3\log n}{8}+O(1).$$

Proof. The proof proceeds in several stages: first of all, note that by analogous reasoning,

$$\sum_{1\le j<i\le n}\iint\limits_{0\le x<y\le 1}\frac{ix^{i-1}jy^{j-1}}{(1-y^j+x^j)(1-y^i+x^i)}\prod_{k=1}^{n}(1-y^k+x^k)dx\,dy$$

is exactly the expected number of *descents*. Since the sum of the two expected values must obviously be $n-1$, it suffices to show that the difference is $3\log n/4+O(1)$, i.e.,

$$\sum_{1\le j<i\le n} ij\iint\limits_{0\le x<y\le 1}\frac{y^{i-1}x^{j-1}-x^{i-1}y^{j-1}}{(1-y^j+x^j)(1-y^i+x^i)}\prod_{k=1}^{n}(1-y^k+x^k)dx\,dy=\frac{3\log n}{4}+O(1). \quad (9)$$

Now we perform the change of variables $x = yz$ to simplify the limits of the integral, which leaves us with the expression

$$\sum_{1\le j<i\le n} ij \int_0^1 \int_0^1 \frac{y^{i+j-1}(z^{j-1}-z^{i-1})}{(1-y^j+y^jz^j)(1-y^i+y^iz^i)} \prod_{k=1}^{n}(1-y^k+y^kz^k)dy\,dz. \tag{10}$$

Let us now focus on a single summand

$$S(i,j) = \int_0^1 \int_0^1 y^{i+j-1}(z^{j-1}-z^{i-1}) \prod_k (1-y^k+y^kz^k)dy\,dz,$$

where the product is over all $k \in [1,n]$ except for i and j. It turns out that the "essential part" of the integral is the region where $1-y$ is of order $(i+j)^{-1}$, and $1-z$ of order $(i+j)^{-2}$. Write $s = i+j$ (note for later use that $s \le 2n$), let $\epsilon > 0$ be sufficiently small (for instance, $\epsilon = 1/100$), and consider the region

$$R = \{(y,z) \,:\, s^{-1-7\epsilon} \le 1-y \le s^{-1+\epsilon} \text{ and } 1-z \le s^{-2+3\epsilon}\}.$$

We first show that the integral over $[0,1]\times[0,1]\setminus R$ is negligible:

- If $1-y \ge s^{-1+\epsilon}$, then
$$y^{i+j-1} = y^{s-1} \le (1-s^{-1+\epsilon})^{s-1} \ll \exp(-s^\epsilon),$$
and so the integral over the region $\{(y,z) \,:\, 1-y \ge s^{-1+\epsilon}\}$ is $O(\exp(-s^\epsilon))$ (the remaining factors are ≤ 1).
- Suppose that $1-z \ge s^{-2+3\epsilon}$. We can also assume now that $1-y \le s^{-1+\epsilon}$. For any k with $s^{1-\epsilon} \le k \le 2s^{1-\epsilon}$, we now have
$$y^k - y^kz^k = y^k(1-z^k) \ge (1-s^{-1+\epsilon})^{2s^{1-\epsilon}}(1-(1-s^{-2+3\epsilon})^{s^{1-\epsilon}}) \gg s^{-1+2\epsilon},$$
which shows that the product in the integrand satisfies
$$\prod_k (1-y^k+y^kz^k) \le (1-Cs^{-1+2\epsilon})^{s^{1-\epsilon}} = O(\exp(-Cs^\epsilon))$$
for some $C > 0$. Hence the integral over the region $\{(y,z) \,:\, 1-z \ge s^{-2+3\epsilon}\}$ is also $O(\exp(-Cs^\epsilon))$.
- We can now already deduce that
$$z^{j-1} - z^{i-1} = z^{j-1}(1-z^{i-j}) = (1+O(s^{-1+3\epsilon}))\cdot(i-j)(1-z) \ll s^{-1+3\epsilon}. \tag{11}$$
Therefore, the whole integrand is $O(s^{-1+3\epsilon})$, and the integral over the region $\{(y,z) \,:\, 1-y \le s^{-1-7\epsilon}\}$ is
$$O\left(s^{-1+3\epsilon}\cdot s^{-1-7\epsilon}\cdot s^{-2+3\epsilon}\right) = O(s^{-4-\epsilon}).$$

Combining the three estimates, we obtain

$$S(i,j) = \iint_R y^{i+j-1}(z^{j-1}-z^{i-1}) \prod_k (1-y^k+y^kz^k)dy\,dz + O(s^{-4-\epsilon}).$$

In addition to (11), we also need to estimate the remaining two factors in the integrand; the first factor is easy:

$$y^{i+j-1} = e^{-s(1-y)}\bigl(1+O(s(1-y)^2)\bigr) = e^{-s(1-y)}(1+O(s^{-1+2\epsilon})).$$

It remains to deal with the product: we take the logarithm to obtain

$$\sum_k \log(1-y^k+y^kz^k).$$

Our first claim is that $y^k - y^kz^k = y^k(1-z^k)$ is small for all k; indeed, if $k \geq s^{1+8\epsilon}$, then

$$y^k(1-z^k) \leq y^k \leq (1-s^{-1-7\epsilon})^{s^{1+8\epsilon}} \ll \exp(-s^\epsilon).$$

If, on the other hand, $k \leq s^{1+8\epsilon}$, then

$$y^k(1-z^k) \leq 1-z^k \leq 1-(1-s^{-2+3\epsilon})^{s^{1+8\epsilon}} \ll s^{-1+11\epsilon}.$$

This holds in particular for $k=i$ and $k=j$, so that we may extend the summation to the entire interval $[1,n]$ at the expense of a small error term. Hence we have

$$\begin{aligned}
\sum_k \log(1-y^k+y^kz^k) &= \sum_{k=1}^{n} \log(1-y^k+y^kz^k) + O(s^{-1+11\epsilon}) \\
&= -\sum_{k=1}^{n} y^k(1-z^k) - \sum_{k=1}^{n}\sum_{m=2}^{\infty} \frac{1}{m}(y^k(1-z^k))^m + O(s^{-1+11\epsilon}) \\
&= -\sum_{k=1}^{n} y^k(1-z^k) + O\Bigl(\sum_{k=1}^{n} \frac{(y^k(1-z^k))^2}{1-y^k(1-z^k)}\Bigr) + O(s^{-1+11\epsilon}) \\
&= -\frac{y(1-z)}{(1-y)(1-yz)} + \frac{y^{n+1}}{1-y} - \frac{(yz)^{n+1}}{1-yz} + O\Bigl(\sum_{k=1}^{\infty} y^{2k}(1-z^k)^2\Bigr) + O(s^{-1+11\epsilon}) \\
&= -\frac{y(1-z)}{(1-y)(1-yz)} + \frac{y^{n+1}}{1-y} - \frac{(yz)^{n+1}}{1-yz} + O\Bigl(\frac{y^2(1-z)^2(1+y^2z)}{(1-y^2)(1-y^2z^2)(1-y^2z)}\Bigr) + O(s^{-1+11\epsilon}) \\
&= -\frac{y(1-z)}{(1-y)(1-yz)} + \frac{y^{n+1}}{1-y} - \frac{(yz)^{n+1}}{1-yz} + O\left(\frac{(s^{-2+3\epsilon})^2}{(s^{-1-7\epsilon})^3}\right) + O(s^{-1+11\epsilon}) \\
&= -\frac{y(1-z)}{(1-y)(1-yz)} + \frac{y^{n+1}}{1-y} - \frac{(yz)^{n+1}}{1-yz} + O(s^{-1+27\epsilon}).
\end{aligned}$$

Next we approximate y^{n+1} and z^{n+1}; first, if $1-z \geq n^{-1-\epsilon}$, then also

$$1-y \geq n^{(-1-\epsilon)(-1-7\epsilon)/(-2+3\epsilon)} \geq n^{-1+\epsilon}$$

by the definition of the region R, and thus $y^{n+1} \ll \exp(-n^\epsilon) \ll \exp(-(s/2)^\epsilon)$. Otherwise,

$$z^{n+1} = 1-(n+1)(1-z) + O(n(1-z)^2)$$

and thus

$$\frac{(yz)^{n+1}}{1-yz} = \frac{y^{n+1}}{1-yz}\,(1-(n+1)(1-z)) + O\Bigl(\frac{(1-z)^2}{1-yz}\cdot ny^{n+1}\Bigr)$$

$$
\begin{aligned}
&= \frac{y^{n+1}}{1-yz}\,(1-(n+1)(1-z)) + O\Big(\frac{(1-z)^2}{1-y}\cdot ne^{-n(1-y)}\Big)\\
&= \frac{y^{n+1}}{1-yz}\,(1-(n+1)(1-z)) + O\Big(\frac{(1-z)^2}{(1-y)^2}\cdot n(1-y)e^{-n(1-y)}\Big)\\
&= \frac{y^{n+1}}{1-yz}\,(1-(n+1)(1-z)) + O\Big(\frac{(1-z)^2}{(1-y)^2}\Big)\\
&= \frac{y^{n+1}}{1-yz}\,(1-(n+1)(1-z)) + O(s^{-2+20\varepsilon}).
\end{aligned}
$$

It follows that

$$
\begin{aligned}
&\sum_k \log(1-y^k+y^kz^k)\\
&\qquad = -\frac{y(1-z)}{(1-y)(1-yz)} + \frac{y^{n+1}}{1-y} - \frac{y^{n+1}}{1-yz}\,(1-(n+1)(1-z)) + O(s^{-1+27\epsilon})\\
&\qquad = -\frac{y(1-z)}{(1-y)(1-yz)}\,(1-y^n(n+1-ny)) + O(s^{-1+27\epsilon}).
\end{aligned}
$$

We can apply a similar argument to show that y^n may be replaced by $e^{-n(1-y)}$, and it is also easy to see that

$$
\frac{y(1-z)}{(1-y)(1-yz)} = \frac{1-z}{(1-y)^2} + O(s^{-1+27\epsilon}).
$$

Hence we finally obtain

$$
\begin{aligned}
S(i,j) = \iint_R (i-j)(1-z)\exp\Big(-s(1-y) - \frac{1-z}{(1-y)^2}\,\big(1-e^{-n(1-y)}(n+1-ny)\big)\Big)dy\,dz\\
\cdot\big(1+O(s^{-1+27\epsilon})\big) + O(s^{-4-\epsilon}).
\end{aligned}
$$

Now we complete the range of the integral again; the integral over the region $0 \le 1-y \le s^{-1-7\epsilon}$ (and $0 \le 1-z \le s^{-2+3\epsilon}$) only gives rise to an error term $O(s^{-4-\epsilon})$ by trivial estimates, as before. Since the integral over R is also easily estimated to be $O(s^{-4+7\epsilon})$, we thus have

$$
\begin{aligned}
S(i,j) = (i-j)\int_{y=1-s^{-1+\epsilon}}^{1}\int_{z=1-s^{-2+3\epsilon}}^{1}(1-z)\\
\cdot\exp\Big(-s(1-y) - \frac{1-z}{(1-y)^2}\,\big(1-e^{-n(1-y)}(n+1-ny)\big)\Big)dy\,dz + O(s^{-4-\epsilon}+s^{-5+34\epsilon}).
\end{aligned}
$$

The change of variables $y = 1-u$, $z = 1-v$ transforms this to

$$
S(i,j) = (i-j)\int_{u=0}^{s^{-1+\epsilon}}\int_{v=0}^{s^{-2+3\epsilon}} v\exp\Big(-su - \frac{v}{u^2}\,\big(1-e^{-nu}(1+nu)\big)\Big)\,du\,dv + O(s^{-4-\epsilon}).
$$

We further extend the range of integration to the entire quarter-plane $[0,\infty)\times[0,\infty)$:

- For $u \geq s^{-1+\epsilon}$ and $v \leq s^{-2+3\epsilon}$, we estimate the integrand by ve^{-su} and obtain an error term of order $O(\exp(-s^{\epsilon}))$.
- For $u \leq n^{-1}$ and $v \geq s^{-2+3\epsilon}$, we use the inequality

$$\big(1 - e^{-nu}(1+nu)\big) \geq \left(1 - \frac{2}{e}\right)(nu)^2 \geq Cs^2u^2$$

  for some $C > 0$ to obtain an estimate of the form $O(\exp(-Cs^{3\epsilon}))$.
- For $n^{-1} \leq u \leq s^{-1+\epsilon}$ and $v \geq s^{-2+3\epsilon}$, we use the fact that $(1 - e^{-nu}(1+nu)) \geq C$ for some $C > 0$ to estimate the integral by

$$\begin{aligned}\int_{u=n^{-1}}^{s^{-1+\epsilon}} \int_{v=s^{-2+3\epsilon}}^{\infty} & v \exp\left(-su - \frac{Cv}{u^2}\right) du\, dv \\ &= \int_{u=n^{-1}}^{s^{-1+\epsilon}} \frac{u^2(Cs^{3\epsilon} + s^2u^2)}{C^2s^2} \exp\left(-su - \frac{Cs^{-2+3\epsilon}}{u^2}\right) du \\ &\ll \int_{u=n^{-1}}^{s^{-1+\epsilon}} s^{-4+5\epsilon} \exp\left(-su - Cs^{\epsilon}\right) du \\ &\ll \exp(-Cs^{\epsilon}).\end{aligned}$$

- Finally, if $u \geq s^{-1+\epsilon}$ and $v \geq s^{-2+3\epsilon}$, we also use the inequality $(1 - e^{-nu}(1+nu)) \geq C$ to obtain an estimate of the form $O(\exp(-s^{\epsilon}))$.

Putting everything together, we find

$$\begin{aligned}S(i,j) &= (i-j) \int_{u=0}^{\infty} \int_{v=0}^{\infty} v \exp\left(-su - \frac{v}{u^2}\big(1 - e^{-nu}(1+nu)\big)\right) du\, dv + O(s^{-4-\epsilon}) \\ &= (i-j) \int_{u=0}^{\infty} \frac{u^4 e^{-su}}{\big(1 - e^{-nu}(1+nu)\big)^2}\, du + O(s^{-4-\epsilon}).\end{aligned}$$

Performing the final change of variables $u = w/s$ yields

$$S(i,j) = (i-j)s^{-5} \int_{w=0}^{\infty} \frac{w^4 e^{-w}}{\big(1 - e^{-nw/s}(1+nw/s)\big)^2}\, dw + O(s^{-4-\epsilon}).$$

The integral represents a function $I(s/n)$ that is easily seen to be convex and increasing; furthermore, $I(0) = \Gamma(5) = 24$. Since s/n is also bounded above (by 2), we obtain the final estimate

$$S(i,j) = (i-j)s^{-5}\left(24 + O(s/n)\right) + O(s^{-4-\epsilon}).$$

It remains to take the sum

$$\begin{aligned}\sum_{1 \leq j < i \leq n} ijS(i,j) &= \sum_{1 \leq j < i \leq n} \left(\frac{ij(i-j)}{(i+j)^5}\left(24 + O((i+j)/n)\right) + O((i+j)^{-2-\epsilon})\right) \qquad (12) \\ &= \frac{3 \log n}{4} + O(1)\end{aligned}$$

to complete the proof of (9) and thus the entire theorem. □

Although it is not needed for our asymptotic formula it is interesting to note that the sum on the right hand side of (12) simplifies with Sigma to

$$\sum_{1\le j<i\le n} \frac{ij(i-j)}{(i+j)^5} = \frac{H_{2n}}{32} - \frac{(4n^2+4n-1)(2n+1)^2}{32} H_{2n}^{(5)} + \frac{512n^4+1024n^3+512n^2-1}{1024} H_n^{(5)}$$
$$+ \frac{n(n+1)(2n+1)}{2}\big(H_{2n}^{(4)} - H_n^{(4)}\big) + \frac{64n^2+64n-1}{128} H_n^{(3)} - \frac{8n^2+8n+1}{16} H_{2n}^{(3)}$$
$$= \frac{1}{32}(\log n + \gamma) + \frac{31\zeta(5)}{1024} - \frac{9\zeta(3)}{128} + \frac{1}{384}(-5+12\log 2) + O(1/n);$$

here $\zeta(x)$ denotes the Riemann Zeta function.

Remark. Intuitively, the difference between the number of inversions (whose average differs from the average for fair permutations) and the number of ascents (whose average differs only in the lower terms) lies in the fact that the number of inversions is a "global" permutation statistic as opposed to the "local" ascents. However, it seems very likely that the distribution of the number of ascents is also Gaussian. The method used to prove Theorem 3 cannot be used, though (since the dependence structure is more complicated). Considering that lengthy calculations are needed just to determine the asymptotic behaviour of the mean, it might be very hard to actually prove a central limit theorem for the number of ascents. Even numerical values for the variance seem to be hard to obtain.

Aside from this seemingly intractable problem, many questions remain for further study: for instance,

- to study the number of fixed points or the probability that a random unfair permutation is fixed-point free,
- to determine the probability that a random unfair permutation is an involution,
- more generally, to study the distribution of cycle lengths,
- to determine the distribution of the distance to the identity permutation, i.e., the sum
$$\sum_{i=1}^{n} |\sigma(i) - i|$$
for a random unfair permutation σ,

to name but a few.

References

[1] A. D. Barbour and G. K. Eagleson. Multiple comparisons and sums of dissociated random variables. *Adv. in Appl. Probab.*, 17(1):147–162, 1985.

[2] J. Blümlein and S. Kurth. Harmonic sums and Mellin transforms up to two-loop order. *Phys. Rev.*, D60, 1999.

[3] M. Bóna. *Combinatorics of permutations*. Discrete Mathematics and its Applications (Boca Raton). Chapman & Hall/CRC, Boca Raton, FL, 2004.

[4] D. Chebikin. Variations on descents and inversions in permutations. *Electron. J. Combin.*, 15(1):Research Paper 132, 34, 2008.

[5] W. M. B. Dukes. Permutation statistics on involutions. *European J. Combin.*, 28(1):186–198, 2007.

[6] P. Flajolet and R. Sedgewick. *Analytic combinatorics*. Cambridge University Press, Cambridge, 2009.
[7] P. Grabner, A. Knopfmacher, and S. Wagner. A general asymptotic scheme for moments of partition statistics. Preprint, 2010.
[8] M. Karr. Summation in finite terms. *J. ACM*, 28:305–350, 1981.
[9] D. E. Knuth. *The Art of Computer Programming*, volume 1: Fundamental Algorithms. Addison-Wesley, 1968. Third edition, 1997.
[10] D. E. Knuth. *The Art of Computer Programming*, volume 3: Sorting and Searching. Addison-Wesley, 1973. Second edition, 1998.
[11] M. Kuba and A. Panholzer. On edge-weighted recursive trees and inversions in random permutations. *Discrete Math.*, 308(4):529–540, 2008.
[12] G. Louchard and H. Prodinger. The number of inversions in permutations: a saddle point approach. *J. Integer Seq.*, 6:Article 03.2.8, 19 pp. (electronic), 2003.
[13] N. Nielson. *Handbuch der Theorie der Gamma Funktion*. B.G. Teubner, Leipzig, 1906.
[14] M. Petkovšek, H. Wilf, and D. Zeilberger. $A = B$. A.K. Peters, Wellesley, MA, 1996.
[15] A. Regev and Y. Roichman. Permutation statistics on the alternating group. *Adv. in Appl. Math.*, 33(4):676–709, 2004.
[16] V. N. Sachkov. *Probabilistic methods in combinatorial analysis*, volume 56 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 1997.
[17] C. Schneider. Symbolic summation assists combinatorics. *Sém. Lothar. Combin.*, 56:1–36, 2007. Article B56b.
[18] R. P. Stanley. *Enumerative combinatorics. Vol. 1*, volume 49 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1997.
[19] J. A. M. Vermaseren. Harmonic sums, Mellin transforms and integrals. *Int. J. Mod. Phys.*, A14:2037–2976, 1999.

Helmut Prodinger, Department of Mathematical Sciences, Stellenbosch University, 7602 Stellenbosch, South Africa
E-mail address: hproding@sun.ac.za

Carsten Schneider, Research Institute for Symbolic Computation, Johannes Kepler University Linz, A–4040 Linz, Austria
E-mail address: Carsten.Schneider@risc.uni-linz.ac.at

Stephan Wagner, Department of Mathematical Sciences, Stellenbosch University, 7602 Stellenbosch, South Africa
E-mail address: swagner@sun.ac.za