**| Annals of Combinatorics**

# On a Theorem of Baxter and Zeilberger via a Result of Roselle

Joshua P. Swanson

**Abstract.** We provide a new proof of a result of Baxter and Zeilberger
showing that inv and maj on permutations are jointly independently
asymptotically normally distributed. The main feature of our argument
is that it uses a generating function due to Roselle, answering a question
raised by Romik and Zeilberger.

## 1. Introduction

For a permutation $w$ in the *symmetric group* $S_n$ written in *one-line notation*
$w = w_1 \ldots w_n$, the *inversion* and *major index* statistics are given by

$$\mathrm{inv}(w) := \#\{i < j : w_i > w_j\} \quad \text{and} \quad \mathrm{maj}(w) := \sum_{\substack{1 \le i \le n-1 \\ w_i > w_{i+1}}} i.$$

It is well known that inv and maj are equidistributed on $S_n$ [Mac, §1] with
common mean and standard deviation

$$\mu_n = \frac{n(n-1)}{4} \quad \text{and} \quad \sigma_n^2 = \frac{2n^3 + 3n^2 - 5n}{72}.$$

These results also follow easily from our arguments; see Remark 2.7. In [BZ10],
Baxter and Zeilberger proved that inv and maj are jointly independently
asymptotically normally distributed as $n \to \infty$. More precisely, define normalized random variables on $S_n$

$$X_n := \frac{\mathrm{inv} - \mu_n}{\sigma_n}, \quad Y_n := \frac{\mathrm{maj} - \mu_n}{\sigma_n}. \tag{1}$$

**Theorem 1.1** (Baxter–Zeilberger [BZ10]). *For each $u, v \in \mathbb{R}$, we have*

$$\lim_{n \to \infty} \mathbb{P}[X_n \le u, Y_n \le v] = \frac{1}{2\pi} \int_{-\infty}^{u} \int_{-\infty}^{v} e^{-x^2/2} e^{-y^2/2} \, \mathrm{d}y \, \mathrm{d}x.$$

See [BZ10] for further historical background. Baxter and Zeilberger's argument involves mixed moments and recurrences based on combinatorial manipulations with permutations. Romik suggested a generating function due to Roselle, quoted as Theorem 2.2 below, should provide another approach. Zeilberger subsequently offered a \$300 reward for such an argument, which has happily now been collected. Our overarching motivation, and Romik's original impetus for suggesting Roselle's formula, is to give a *local limit theorem*, i.e. a formula for the counts $\#\{w \in S_n : \mathrm{inv}(w) = u, \mathrm{maj}(w) = v\}$ with an explicit error term, which will be the subject of a future article. For further context, see [Zei] and [Thi16].

## 2. Consequences of Roselle's Formula

Here we recall Roselle's formula, originally stated in different but equivalent terms, and derive a generating function expression which quickly motivates Theorem 1.1.

**Definition 2.1.** Let $H_n$ be the bivariate inv, maj generating function on $S_n$, i.e.

$$H_n(p,q) := \sum_{w \in S_n} p^{\mathrm{inv}(w)} q^{\mathrm{maj}(w)}.$$

**Theorem 2.2** (Roselle [Ros74]). *We have*

$$\sum_{n \geq 0} \frac{H_n(p,q)z^n}{(p)_n(q)_n} = \prod_{a,b \geq 0} \frac{1}{1 - p^a q^b z}, \tag{2}$$

*where* $(p)_n := (1-p)(1-p^2)\cdots(1-p^n)$.

The following is the main result of this section. An *integer partition* $\mu \vdash n$ is a weakly decreasing list of positive integers $\mu_1 \geq \mu_2 \geq \cdots \geq \mu_k > 0$ summing to $n$. The *length* of $\mu$ is $\ell(\mu) = k$.

**Theorem 2.3.** *There are constants* $c_\mu \in \mathbb{Z}$ *indexed by integer partitions* $\mu \vdash n$ *such that*

$$\frac{H_n(p,q)}{n!} = \frac{[n]_p![n]_q!}{n!^2} F_n(p,q), \tag{3}$$

*where*

$$F_n(p,q) = \sum_{d=0}^{n} [(1-p)(1-q)]^d \sum_{\substack{\mu \vdash n \\ \ell(\mu) = n-d}} \frac{c_\mu}{\prod_i [\mu_i]_p [\mu_i]_q} \tag{4}$$

*and* $[n]_p! := [n]_p[n-1]_p \cdots [1]_p$, $[c]_p := 1 + p + \cdots + p^{c-1} = (1-p^c)/(1-p)$.

An explicit expression for $c_\mu$ is given below in (12). The rest of this section is devoted to proving Theorem 2.3. Straightforward manipulations with (2) immediately yield (3), where

$$F_n(p,q) := (1-p)^n(1-q)^n n! \cdot \{z^n\} \left( \prod_{a,b \geq 0} \frac{1}{1 - p^a q^b z} \right) \tag{5}$$

and $\{z^n\}$ here refers to extracting the coefficient of $z^n$. Thus it suffices to show (5) implies (4). By standard arguments, the $z^n$ coefficient of the product over $a, b$ in (5) is the bivariate generating function of size-$n$ multisets of pairs $(a, b) \in \mathbb{Z}_{\geq 0}^2$, where the weight of such a multiset is $p^{\sum_i a_i} q^{\sum_i b_i}$. We will use this same componentwise-sum weight throughout.

**Definition 2.4.** For $\lambda \vdash n$, let $M_\lambda$ be the bivariate generating function for multisets of pairs $(a, b) \in \mathbb{Z}_{\geq 0}^2$ of *type* $\lambda$, meaning some element has multiplicity $\lambda_1$, another element has multiplicity $\lambda_2$, etc.

We clearly have

$$\{z^n\} \left( \prod_{a,b \geq 0} \frac{1}{1 - p^a q^b z} \right) = \sum_{\lambda \vdash n} M_\lambda(p, q), \tag{6}$$

though the $M_\lambda$ are inconvenient to work with, so we perform a change of basis.

**Definition 2.5.** Let $P[n]$ denote the lattice of *set partitions* of $[n] := \{1, 2, \ldots, n\}$ with minimum $\widehat{0} = \{\{1\}, \{2\}, \ldots, \{n\}\}$ and maximum $\widehat{1} = \{\{1, 2, \ldots, n\}\}$. Here $\Lambda \leq \Pi$ means that $\Pi$ can be obtained from $\Lambda$ by merging blocks of $\Lambda$. The *type* of a set partition $\Lambda$ is the integer partition obtained by rearranging the list of the block sizes of $\Lambda$ in weakly decreasing order. For $\lambda \vdash n$, set

$$\Pi(\lambda) := \{\{1, 2, \ldots, \lambda_1\}, \{\lambda_1 + 1, \lambda_1 + 2, \ldots, \lambda_1 + \lambda_2\}, \ldots\},$$

which has type $\lambda$.

**Definition 2.6.** For $\Pi \in P[n]$, let $R_\Pi$ denote the bivariate generating function for lists $L \in (\mathbb{Z}_{\geq 0}^2)^n$ where for each block of $\Pi$ the entries in $L$ from that block are all equal. Similarly, let $S_\Pi$ denote the bivariate generating function of lists $L$ where in addition to entries from the same block being equal, entries from two different blocks are not equal.

We easily see that

$$R_\Lambda = \prod_{A \in \Lambda} \frac{1}{(1 - p^{\#A})(1 - q^{\#A})} \tag{7}$$

and that

$$R_\Lambda = \sum_{\Pi : \Lambda \leq \Pi} S_\Pi, \tag{8}$$

so that, by Möbius inversion on $P[n]$,

$$S_\Pi = \sum_{\Lambda : \Pi \leq \Lambda} \mu(\Pi, \Lambda) R_\Lambda, \tag{9}$$

where $\mu(\Pi, \Lambda)$ is the Möbius function of the lattice of set partitions. Under the "forgetful" map from lists to multisets, a multiset of type $\lambda \vdash n$ has fiber of size $\binom{n}{\lambda}$. It follows that

$$S_{\Pi(\lambda)} = \frac{n!}{\lambda!} M_\lambda, \tag{10}$$

where $\lambda! := \lambda_1! \lambda_2! \cdots$. Combining in order (5), (6), (10), (9), and (7) gives

$$F_n(p,q) = \sum_{d=0}^{n} [(1-p)(1-q)]^d \sum_{\lambda \vdash n} \lambda! \sum_{\substack{\Lambda : \Pi(\lambda) \leq \Lambda \\ \#\Lambda = n-d}} \frac{\mu(\Pi(\lambda), \Lambda)}{\prod_{A \in \Lambda} [\#A]_p [\#A]_q}. \qquad (11)$$

Now (4) follows from (11), where

$$c_\mu = \sum_{\lambda \vdash n} \lambda! \sum_{\substack{\Lambda : \Pi(\lambda) \leq \Lambda \\ \text{type}(\Lambda) = \mu}} \mu(\Pi(\lambda), \Lambda). \qquad (12)$$

This completes the proof of Theorem 2.3.

*Remark 2.7.* From (12), $c_{(1^n)} = 1$ since the sum only involves $\Lambda = \widehat{0}$, where $(1^n)$ refers to the partition with $n$ parts of size 1. Letting $p \to 1$ (or, symmetrically, $q \to 1$) in (4), the only surviving term is $d = 0$ and $\mu = (1^n)$. Consequently, $H_n(1, q) = [n]_q!$, recovering a classic result of MacMahon [Mac, §1]. Explicitly,

$$\sum_{w \in S_n} q^{\mathrm{inv}(w)} = [n]_q! = \sum_{w \in S_n} q^{\mathrm{maj}(w)}.$$

The mean and standard deviation may be extracted by recognizing $[n]_q!/n!$ as the probability generating function of the sum of independent discrete random variables.

*Remark 2.8.* Using (3), we see that the probability generating function (discussed below in Example 4.3) $H_n(p, q)/n!$ differs from $[n]_p![n]_q!/n!^2$ by precisely the correction factor $F_n(p, q)$. Using (5), this factor has the following combinatorial interpretation:

$$F_n = \frac{n! \cdot \text{g.f. of size - } n \text{ multisets from } \mathbb{Z}_{\geq 0}^2}{\text{g.f. of size - } n \text{ lists from } \mathbb{Z}_{\geq 0}^2}.$$

Intuitively, the numerator and denominator should be the same "up to first order." Theorem 3.1 will give one precise sense in which they are asymptotically equal.

## 3. Estimating the Correction Factor

This section is devoted to showing that the correction factor $F_n(p, q)$ from Theorem 2.3 is negligible in an appropriate sense, Theorem 3.1. Recall that $\sigma_n$ denotes the standard deviation of inv or maj on $S_n$.

**Theorem 3.1.** *Uniformly on compact subsets of $\mathbb{R}^2$, we have*

$$F_n(e^{is/\sigma_n}, e^{it/\sigma_n}) \to 1 \quad \text{as} \quad n \to \infty.$$

We begin with some simple estimates starting from (11) which motivate the rest of the inequalities in this section. We may assume $|s|, |t| \leq M$ for some fixed $M$. Setting $p = e^{is/\sigma_n}, q = e^{it/\sigma_n}$, we have $|1 - p| = |1 - \exp(is/\sigma_n)| \leq$

$|s|/\sigma_n$. For $n$ sufficiently large compared to $M$ and $1 \leq c \leq n$, we claim that $|[c]_p| \geq 1$. Assuming this for the moment, for $n$ sufficiently large, (11) gives

$$|F_n(e^{is/\sigma_n}, e^{it/\sigma_n}) - 1| \leq \sum_{d=1}^{n} \frac{|st|^d}{\sigma_n^{2d}} \sum_{\lambda \vdash n} \lambda! \sum_{\substack{\Lambda : \Pi(\lambda) \leq \Lambda \\ \#\Lambda = n-d}} |\mu(\Pi(\lambda), \Lambda)|. \qquad (13)$$

As for the claim, one finds $|[c]_p| = |\sinh(cs/2\sigma_n)/\sinh(s/2\sigma_n)|$. For sufficiently large $n$, $cs/2\sigma_n \ll 1$, and $\sinh(z)$ is increasing near 0, which gives the claim. We now simplify the inner sum on the right-hand side of (13).

**Lemma 3.2.** *Suppose $\lambda \vdash n$ with $\ell(\lambda) = n - k$, and fix $d$. Then*

$$\sum_{\substack{\Lambda : \Pi(\lambda) \leq \Lambda \\ \#\Lambda = n-d}} \mu(\Pi(\lambda), \Lambda) = (-1)^{d-k} \sum_{\substack{\Lambda \in P[n-k] \\ \#\Lambda = n-d}} \prod_{A \in \Lambda} (\#A - 1)! \qquad (14)$$

*and the terms on the left all have the same sign $(-1)^{d-k}$. The sums are empty unless $n \geq d \geq k \geq 0$.*

*Proof.* The upper order ideal $\{\Lambda \in P[n] : \Pi(\lambda) \leq \Lambda\}$ is isomorphic to $P[n-k]$ by collapsing the $n-k$ blocks of $\Pi(\lambda)$ to singletons. This isomorphism preserves the number of blocks. Furthermore, recall that in $P[n]$ the Möbius function satisfies

$$\mu(\widehat{0}, \widehat{1}) = (-1)^{n-1}(n-1)!,$$

from which it follows easily that

$$\mu(\widehat{0}, \Lambda) = \prod_{A \in \Lambda} (-1)^{\#A-1}(\#A - 1)!. \qquad (15)$$

The result follows immediately upon combining these observations.                    $\square$

**Lemma 3.3.** *Let $\lambda \vdash n$ with $\ell(\lambda) = n - k$ and $n \geq d \geq k \geq 0$. Then*

$$\sum_{\substack{\Lambda : \Pi(\lambda) \leq \Lambda \\ \#\Lambda = n-d}} |\mu(\Pi(\lambda), \Lambda)| \leq (n-k)^{2(d-k)}. \qquad (16)$$

*Proof.* Using (14), we can interpret the sum as the number of permutations of $[n-k]$ with $n-d$ cycles, which is a Stirling number of the first kind. There are well-known asymptotics for these numbers, though the stated elementary bound suffices for our purposes. We induct on $d$. At $d = k$, the result is trivial. Given a permutation of $[n-k]$ with $n-d$ cycles, choose $i, j \in [n-k]$ from different cycles. Suppose the cycles are of the form $(i' \cdots i)$ and $(j \cdots j')$. Splice the two cycles together to obtain

$$(i' \cdots i \ j \cdots j').$$

This procedure constructs every permutation of $[n-k]$ with $n-(d+1)$ cycles and requires no more than $(n-k)^2$ choices. The result follows.                    $\square$

**Lemma 3.4.** *For $n \geq d \geq k \geq 0$, we have*

$$\sum_{\substack{\lambda \vdash n \\ \ell(\lambda) = n-k}} \lambda! \sum_{\substack{\Lambda : \Pi(\lambda) \leq \Lambda \\ \#\Lambda = n-d}} |\mu(\Pi(\lambda), \Lambda)| \leq (n-k)^{2d-k}(k+1)!. \tag{17}$$

*Proof.* For $\lambda \vdash n$ with $\ell(\lambda) = n - k$, $\lambda!$ can be thought of as the product of terms obtained from filling the $i$th row of the *Young diagram* of $\lambda$ with $1, 2, \ldots, \lambda_i$. Alternatively, we may fill the cells of $\lambda$ as follows: put $n - k$ one's in the first column, and fill the remaining cells with the numbers $2, 3, \ldots, k+1$ starting at the largest row and proceeding left to right. It's easy to see the labels of the first filling are bounded above by the labels of the second filling, so that $\lambda! \leq (k+1)!$. Furthermore, each $\lambda \vdash n$ with $\ell(\lambda) = n - k$ can be constructed by first placing $n - k$ cells in the first column and then deciding on which of the $n - k$ rows to place each of the remaining $k$ cells, so there are no more than $(n-k)^k$ such $\lambda$. The result follows from combining these bounds with (16). □

**Lemma 3.5.** *For $n$ sufficiently large, for all $0 \leq d \leq n$ we have*

$$\sum_{\lambda \vdash n} \lambda! \sum_{\substack{\Lambda : \Pi(\lambda) \leq \Lambda \\ \#\Lambda = n-d}} |\mu(\Pi(\lambda), \Lambda)| \leq 3n^{2d}.$$

*Proof.* For $n \geq 2$ large enough, for all $n \geq k \geq 2$ we see that $(k+1)! < n^{k-1}$. Using (17) gives

$$\sum_{\lambda \vdash n} \lambda! \sum_{\substack{\Lambda : \Pi(\lambda) \leq \Lambda \\ \#\Lambda = n-d}} |\mu(\Pi(\lambda), \Lambda)| \leq \sum_{k=0}^{d} (n-k)^{2d-k}(k+1)!$$

$$\leq n^{2d} + 2(n-1)^{2d-1} + \sum_{k=2}^{d} (n-k)^{2d-k} n^{k-1}$$

$$\leq n^{2d} + 2n^{2d-1} + \sum_{k=2}^{d} n^{2d-1}$$

$$= n^{2d} + 2n^{2d-1} + (d-1)n^{2d-1} \leq 3n^{2d}.$$

□

We may now complete the proof of Theorem 3.1. Combining Lemma 3.5 and (13) gives

$$|F_n(e^{is/\sigma_n}, e^{it/\sigma_n}) - 1| \leq 3 \sum_{d=1}^{n} \frac{(Mn)^{2d}}{\sigma_n^{2d}}.$$

Since $\sigma_n^2 \sim n^3/36$ and $M$ is constant, $(Mn)^{2d}/\sigma_n^{2d} \sim (36^2 M^2/n)^d$. Since $M$ is constant, using a geometric series it follows that

$$\lim_{n \to \infty} \sum_{d=1}^{n} \frac{(Mn)^{2d}}{\sigma_n^{2d}} = 0,$$

completing the proof of Theorem 3.1.

*Remark 3.6.* Indeed, the argument shows that $|F_n(e^{is/\sigma_n}, e^{it/\sigma_n}) - 1| = O(1/n)$. The above estimates are particularly far from sharp for large $d$, though for small $d$ they are quite accurate. Working directly with (11), one finds the $d = 1$ contribution to be

$$(1-p)(1-q)\frac{2 - \binom{n}{2}}{[2]_p[2]_q}.$$

Letting $p = e^{is/\sigma_n}, q = e^{it/\sigma_n}$, straightforward estimates shows that this is $\Omega(1/n)$. Consequently, the preceding arguments are strong enough to identify the leading term, and in particular

$$|F_n(e^{is/\sigma_n}, e^{it/\sigma_n}) - 1| = \Theta(1/n).$$

## 4. Deducing Baxter and Zeilberger's Result

We next summarize enough of the standard theory of characteristic functions to prove Theorem 1.1 using (3) and Theorem 3.1.

**Definition 4.1.** The *characteristic function* of an $\mathbb{R}^k$-valued random variable $X = (X_1, \ldots, X_k)$ is the function $\phi_X \colon \mathbb{R}^k \to \mathbb{C}$ given by

$$\phi_X(s_1, \ldots, s_k) := \mathbb{E}[\exp(i(s_1 X_1 + \cdots + s_k X_k))].$$

*Example 4.2.* It is well known that the characteristic function of the standard normal random variable with density $\frac{1}{\sqrt{2\pi}}e^{-x^2/2}$ is $e^{-s^2/2}$. Similarly, the characteristic function of a bivariate jointly independent standard normal random variable with density $\frac{1}{2\pi}e^{-x^2/2-y^2/2}$ is $e^{-s^2/2-t^2/2}$.

*Example 4.3.* If $W$ is a finite set and $\mathrm{stat} = (\mathrm{stat}_1, \ldots, \mathrm{stat}_k) \colon W \to \mathbb{Z}_{\geq 0}^k$ is some statistic, the *probability generating function* of stat on $W$ is

$$P(x_1, \ldots, x_k) := \frac{1}{\#W} \sum_{w \in W} x_1^{\mathrm{stat}_1(w)} \cdots x_k^{\mathrm{stat}_k(w)}.$$

The characteristic function of the corresponding random variable $X$ where the $w$ are chosen uniformly from $W$ is

$$\phi_X(s_1, \ldots, s_k) = P(e^{is_1}, \ldots, e^{is_k}).$$

From Example 4.3, Remark 2.7, and an easy calculation, it follows that the characteristic functions of the random variables $X_n$ and $Y_n$ from (1) are

$$\phi_{X_n}(s) = e^{-i\mu_n s/\sigma_n}\frac{[n]_{e^{is/\sigma_n}}!}{n!} \quad \text{and} \quad \phi_{Y_n}(t) = e^{-i\mu_n t/\sigma_n}\frac{[n]_{e^{it/\sigma_n}}!}{n!}. \quad (18)$$

An analogous calculation for the random variable $(X_n, Y_n)$ together with (18) and (3) gives

$$\phi_{(X_n,Y_n)}(s,t) = e^{-i(\mu_n s/\sigma_n + \mu_n t/\sigma_n)}\frac{H_n(e^{is/\sigma_n}, e^{it/\sigma_n})}{n!}$$
$$= \phi_{X_n}(s)\phi_{Y_n}(t)F_n(e^{is/\sigma_n}, e^{it/\sigma_n}). \quad (19)$$

**Theorem 4.4** (Multivariate Lévy Continuity [Bil95, p. 383]). *Suppose that $X^{(1)}$, $X^{(2)}$, ... is a sequence of $\mathbb{R}^k$-valued random variables and $X$ is an $\mathbb{R}^k$-valued random variable. Then $X^{(1)}, X^{(2)}, \ldots$ converges in distribution to $X$ if and only if $\phi_{X^{(n)}}$ converges pointwise to $\phi_X$.*

If the distribution function of $X$ is continuous everywhere, convergence in distribution means that for all $u_1, \ldots, u_k$ we have

$$\lim_{n\to\infty} \mathbb{P}[X_i^{(n)} \le u_i, 1 \le i \le k] = \mathbb{P}[X_i \le u_i, 1 \le i \le k].$$

Many techniques are available for proving that inv and maj on $S_n$ are asymptotically normal. The result is typically attributed to Feller.

**Theorem 4.5** ([Fel68, p. 257]). *The sequences of random variables $X_n$ and $Y_n$ from (1) each converge in distribution to the standard normal random variable.*

We may now complete the proof of Theorem 1.1. From Theorem 4.5 and Example 4.2, we have for all $s, t \in \mathbb{R}$

$$\lim_{n\to\infty} \phi_{X_n}(s) = e^{-s^2/2} \quad \text{and} \quad \lim_{n\to\infty} \phi_{Y_n}(t) = e^{-t^2/2}. \tag{20}$$

Combing in order (20), (19), and Theorem 3.1 gives

$$\lim_{n\to\infty} \phi_{(X_n, Y_n)}(s, t) = e^{-s^2/2 - t^2/2}.$$

Theorem 1.1 now follows from Example 4.2 and Theorem 4.4.

# Acknowledgements

## Declarations

**Conflict of interest** The author states that there is no conflict of interest.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

# References

[Bil95]   Patrick Billingsley. *Probability and measure.* Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York, third edition, 1995. A Wiley-Interscience Publication.

[BZ10]   Andrew Baxter and Doron Zeilberger. The Number of Inversions and the Major Index of Permutations are Asymptotically Joint-Independently Normal (Second Edition!), 2010. arXiv:1004.1160.

[Fel68]   William Feller. *An introduction to probability theory and its applications. Vol. I.* Third edition. John Wiley & Sons, Inc., New York-London-Sydney, 1968.

[Mac]    P. A. MacMahon. Two Applications of General Theorems in Combinatory Analysis: (1) To the Theory of Inversions of Permutations; (2) To the Ascertainment of the Numbers of Terms in the Development of a Determinant which has Amongst its Elements an Arbitrary Number of Zeros. *Proc. London Math. Soc.*, S2-15(1):314.

[Ros74]   D. P. Roselle. Coefficients associated with the expansion of certain products. *Proc. Amer. Math. Soc.*, 45:144–150, 1974.

[Thi16]   Marko Thiel. The inversion number and the major index are asymptotically jointly normally distributed on words. *Combin. Probab. Comput.*, 25(3):470–483, 2016.

[Zei]    Doron Zeilberger. The number of inversions and the major index of permutations are asymptotically joint-independently normal. Personal Web Page. http://sites.math.rutgers.edu/~zeilberg/mamarim/mamarimhtml/invmaj.html (version: 2018-01-25).

Joshua P. Swanson
University of Southern California
Los Angeles, CA
USA
e-mail: swansonj@usc.edu