An Experimental Mathematics Approach to the Area Statistic of Parking Functions

YUKUN YAO and DORON ZEILBERGER

Once upon a time, way back in the 1960s, there was a one-way street (with no passing allowed) with \( n \) parking spaces bordering the sidewalk. Entering the street were \( n \) cars, each driven by a loyal husband, and sitting next to him, at times dozing off, was his capricious (and somewhat bossy) wife. At a random point along the street, the wife would wake up and order her husband, “Park here, darling.” If that space was unoccupied, the hubby gladly obliged, but if the parking space was occupied, he would park at the first vacant parking space—if there was one before the end of the street. Alas, if all the subsequent parking spaces were occupied, he would have to drive around the block, return to the beginning of the one-way street, and then look for the first available spot. Due to construction, this trip around the block wasted half an hour, making the wife very cranky.

Question: What is the probability that no one has to go around the block?

Answer: \((n + 1)^n / n^n \approx e / (n + 1)\).

Both the question and its elegant answer are due to Alan Konheim and Benji Weiss [KW].

Parking Functions

Suppose wife \( i \) (\( 1 \leq i \leq n \)) wakes up at parking space \( p_i \). Then the preferences of the wives can be summarized as an array \( (p_1, \ldots, p_n) \), where \( 1 \leq p_i \leq n \). So altogether, there are \( n^n \) possible preference vectors, starting from \((1, \ldots, 1)\), where it is clearly possible for everyone to park on the first pass, and ending with \((n, \ldots, n)\) (all \( n \)'s), where every wife prefers the last parking space, and of course every husband after the first has to drive around the block. Given a preference vector \((p_1, \ldots, p_n)\), let \((p_{(1)}, \ldots, p_{(n)})\) be its sorted version, arranged in (weakly) increasing order. For example, if \((p_1, p_2, p_3, p_4) = (3, 1, 1, 4)\), then \((p_{(1)}, p_{(2)}, p_{(3)}, p_{(4)}) = (1, 1, 3, 4)\).

We invite our readers to convince themselves that a parking-space preference vector \((p_1, \ldots, p_n)\) makes it possible for every husband to park without inconveniencing his wife if and only if \( p_{(i)} \leq i \) for \( 1 \leq i \leq n \). This naturally leads to the following definition.

Definition 1 (Parking function). A vector of positive integers \((p_1, \ldots, p_n)\) with \( 1 \leq p_i \leq n \) is a parking function if its (nondecreasing) sorted version \((p_{(1)}, \ldots, p_{(n)})\) (i.e.,

---

¹This article is accompanied by a Maple package, ParkingStatistics.txt, available at http://sites.math.rutgers.edu/~zeilberg/tokhniot/ParkingStatistics.txt.
There are \((n+1)^{n-1}\) parking functions of length \(n\).

There are many proofs of this lovely theorem; possibly the slickest is due to the brilliant Henry Pollak (who apparently did not deem it worthy of publication. It is quoted, e.g., in [FR]). It is nicely described in [St1, pp. 4–5] (see also [St2]), and hence we will not repeat it here. Instead, as a warm-up to the “statistical” part, and to illustrate the power of experiments, we will give a much uglier proof, one that, however, is motivated.

Before going on to present our (very possibly not new) “humble” proof, we should mention that one natural way to prove the Konheim–Weiss theorem is by a bijection with labeled trees on \(n+1\) vertices, which Arthur Cayley famously proved is also enumerated by \((n+1)^{n-1}\). The first such bijection, as far as we know, was given by the great formal linguist Marco Schützenberger [Sc]. This was followed by an elegant bijection produced by the classical combinatorial giants Dominique Foata and John Riordan [FR], as well as others.

Since we know (at least) sixteen different proofs of Cayley’s formula (see, e.g., [Z3]) and at least four different bijections between parking functions and labeled trees, there are at least 64 different proofs (see also [St3, ex. 5.49]) of the parking enumeration theorem. To these one must add proofs like Pollak’s and a few others.

Curiously, our “new” proof has some resemblance to the very first one in [KW], since both use recurrences (one of the greatest tools in the experimental mathematician’s tool kit), but our proof is both motivated and experimental (yet fully rigorous).

### An Experimental Mathematics Motivated Proof of the Kohnheim–Weiss Parking Enumeration Theorem

On encountering a new combinatorial family, one’s first task is to write a computer program to enumerate as many terms as possible and hope to be able to conjecture a nice formula. One can also try to “cheat” and use the great Online Encyclopedia of Integer Sequences (OEIS) to see whether anyone has come up with this sequence before and to see whether this new combinatorial family is mentioned there.

A very brute force approach, which will not go very far (but would suffice to get the first five terms needed for OEIS) is to list the superset, in this case all the \(n^n\) vectors in \(\{1, \ldots, n\}^n\), and for each of them, sort it and see whether the condition \(p_{(i)} \leq i\) holds for all \(1 \leq i \leq n\). Then count the vectors that pass this test.

But a much better way is to use dynamic programming to express the desired sequence, let’s call it \(a(n)\), in terms of values \(a(i)\) for \(i < n\).

Let us analyze the anatomy of a typical parking function of length \(n\). A natural parameter is the number of 1’s that show up; let’s call it \(k\) \((0 \leq k \leq n\), i.e.,

\[
p_{(1)} = 1, \ldots, p_{(k)} = 1, \quad 2 \leq p_{(k+1)} \leq k+1, \quad \ldots, p_{(n)} \leq n.
\]

Removing the 1’s yields a shorter weakly increasing vector,

\[
2 \leq p_{(k+1)} \leq p_{(k+2)} \leq \cdots \leq p_{(n)}
\]

satisfying

\[
p_{(k+1)} \leq k + 1, \quad p_{(k+2)} \leq k + 2, \quad \ldots, \quad p_{(n)} \leq n.
\]

Define

\[
(q_1, \ldots, q_{n-k}) := (p_{(k+1)} - 1, \ldots, p_{(n)} - 1).
\]

The vector \((q_1, \ldots, q_{n-k})\) satisfies

\[
1 \leq q_1 \leq \cdots \leq q_{n-k}
\]

and

\[
q_1 \leq k, \quad q_2 \leq k + 1, \quad \ldots, \quad q_{n-k} \leq n - 1.
\]

We see that the set of parking functions with exactly \(k\) 1’s may be obtained by taking the above set of vectors of length \(n-k\), adding 1 to each component, scrambling it every which way, and inserting the 1’s every which way. Alas, the “scrambling” of the set of such \(q\)-vectors is not of the original form. We are forced to consider a more general object, namely scramblings of vectors of the form

\[
p_{(1)} \leq \cdots \leq p_{(n)}\]

with the condition

\[
p_{(1)} \leq a, \quad p_{(2)} \leq a + 1, \quad \ldots, \quad p_{(a)} \leq a + n - 1
\]

for a general positive integer \(a\), not just for \(a = 1\). So in order to get the dynamical programming recurrence rolling, we are forced to introduce a more general object, called an \(a\)-parking function. This leads to the following definition.

### Definition 2 (\(a\)-parking function).

A vector of positive integers \((p_1, \ldots, p_n)\) with \(1 \leq p_i \leq n + a - 1\) is an \(a\)-parking function if its (nondecreasing) sorted version \((p_{(1)}, \ldots, p_{(n)})\) (i.e., \(p_{(1)} \leq p_{(2)} \leq \cdots \leq p_{(n)}\)), and the latter is a permutation of the former) satisfies

\[
p_{(i)} \leq a + i - 1, \quad 1 \leq i \leq n.
\]

Note that every \(a\)-parking function is also a \(b\)-parking function for \(b > a\). For example, 4355 is a 3-parking function, a 4-parking function, etc., but it is neither a 2-parking function nor a 1-parking function (aka parking function).

The usual parking functions are the special case \(a = 1\). So if we were able to find an efficient recurrence for counting \(a\)-parking functions, we would be able to answer our original question.

So let’s redo the above “anatomy” for these more general creatures, and hope that the two parameters \(n\) and \(a\)
will suffice to establish a recursive scheme, and then we won't need to introduce yet more general creatures.

Let's analyze the anatomy of a typical \( a \)-parking function of length \( n \). Again, a natural parameter is the number of 1's that show up; let's call that number \( k \) (\( 0 \leq k \leq n \)), i.e.,
\[
p(1) = 1, \quad \ldots, \quad p(k) = 1, \quad 2 \leq p(k+1) \leq a+k, \quad \ldots, \quad p(n) \leq a+n-1.
\]
Removing the 1's yields a sorted vector,
\[
2 \leq p(k+1) \leq p(k+2) \leq \cdots \leq p(n),
\]
satisfying
\[
p(k+1) \leq k+a, \quad p(k+2) \leq k+a+1, \quad \ldots, \quad p(n) \leq n+a-1.
\]
Define
\[
(q_1, \ldots, q_{n-k}) := (p(k+1) - 1, \ldots, p(n) - 1).
\]
The vector \((q_1, \ldots, q_{n-k})\) satisfies
\[
q_1 \leq \cdots \leq q_{n-k}
\]
and
\[
q_1 \leq k+a-1, \quad q_2 \leq k+a, \quad \ldots, \quad q_{n-k} \leq n+a-1.
\]
We see that the set of \( a \)-parking functions with exactly \( k \) 1's may be taken by taking the above set of vectors of length \( n-k \), adding 1 to each component, scrambling it every which way, and inserting the \( k \) 1's every which way.

But now the set of scramblings of the vectors \((q_1, \ldots, q_{n-k})\) is an old friend. It is the set of \((a+k-1)\)-parking functions of length \( n-k \). To get all \( a \)-parking functions of length \( n \) with exactly \( k \) ones, we need to take each and every member of the set of \((a+k-1)\)-parking functions of length \( n-k \), add 1 to each component, and insert \( k \) ones every which way. There are \( \binom{n}{k} \) ways of doing so. Hence the number of \( a \)-parking functions of length \( n \) with exactly \( k \) ones is \( \binom{n}{k} \) times the number of \((a+k-1)\)-parking functions of length \( n-k \). Summing over all \( k \) between 0 and \( n \), we get the following recurrence.

**Theorem 2** (Fundamental recurrence for \( a \)-parking functions). Let \( p(n,a) \) be the number of \( a \)-parking functions of length \( n \). We have the recurrence
\[
p(n,a) = \sum_{k=0}^{n} \binom{n}{k} p(n-k, a+k-1),
\]
subject to the boundary conditions \( p(n,0) = 0 \) for \( n \geq 1 \), and \( p(0,a) = 1 \) for \( a \geq 0 \).

Note that in the sense of Wilf [W], this already answers the enumeration problem of computing \( p(n,a) \), and hence \( p(n,1) = p(n) \), since this gives us a polynomial-time algorithm to compute \( p(n) \) (and \( p(n,a) \)). Moving the term \( k = 0 \) from the right to the left and denoting \( p(n,a) \) by \( p_n(a) \), we have
\[
p_n(a) - p_n(a-1) = \sum_{k=0}^{n} \binom{n}{k} p_{n-k}(a+k-1).
\]
Hence we can express \( p_n(a) \) as follows, in terms of \( p_m(a) \) with \( m < n \):
\[
p_n(a) = \sum_{k=0}^{a} \sum_{m=0}^{n-k} \binom{n}{k} p_{m-k}(b+k-1).
\]
Here is the Maple code that implements it:

```maple
def p(n,a):
    if n=0 then
        RETURN (1)
    else
        factor (subs(b=a, sum(add(binomial(n,k)
            *subs(a=a+k-1,p(n-k,a)),k=1..n)),a=1..b))
    fi:
end:
```

If you copy and paste this into a Maple session followed by the line
\[
[seq(p(i,a),i=1..10)];
\]
you will immediately get
\[
[a, a(a+2), a(a+3)^2, a(a+4)^3, a(a+5)^4, a(a+6)^5, a(a+7)^6, a(a+8)^7, a(a+9)^8, a(a+10)^9],
\]
Note that these are rigorously proved exact expressions in terms of general \( a \) (i.e., symbolic \( a \)) for \( p_n(a) \), for \( 1 \leq n \leq 10 \), and we can easily get more. The following guess immediately comes to mind:
\[
p(n,a) = p_n(a) = a(a+n)^{n-1}.
\]
How to prove this rigorously? If you set
\[
q(n,a) := a(a+n)^{n-1},
\]
that the fact that \( p(n,a) = q(n,a) \) will follow by induction once you prove that \( q(n,a) \) also satisfies the same fundamental recurrence:
\[
q(n,a) = \sum_{k=0}^{n} \binom{n}{k} q(n-k,a+k-1).
\]
In other words, in order to prove that \( p(n,a) = a(a+n)^{n-1} \), we have to prove the identity
\[
a(a+n)^{n-1} = \sum_{k=0}^{n} \binom{n}{k} (a+k-1)(a+n-1)^{n-k-1},
\]
but this is an immediate consequence of the binomial theorem, hence trivial to both humans and machines.

We have just rigorously re-proved, via experimental mathematics, the following well-known theorem.
The number of $a$-parking functions of length $n$ is
\[ p(n, a) = a^n (a + n)^{n-1}. \]

In particular, by substituting $a = 1$, we have re-proved the original Konheim–Weiss theorem that $p(n, 1) = (n + 1)^{n-1}$.

**From Enumeration to Statistics in General**

Often in enumerative combinatorics, the class of interest has natural “statistics,” such as height, weight, and IQ for human beings, and one is interested in—rather than, for a finite set $A$,
\[ |A| := \sum_{a \in A} 1, \]
called the naive counting, resulting in a number (obviously a nonnegative integer)—the so-called weighted counting
\[ |A|_x := \sum_{a \in A} x^f(a), \]
where $f : A \rightarrow \mathbb{Z}$ is the statistic in question. To go from the weighted enumeration (a certain Laurent polynomial) to straight enumeration, one sets $x = 1$, i.e., $|A|_1 := |A|$.

Since this is mathematics, and not accounting, the usual scenario is not just one specific set $A$, but a sequence of sets $\{A_n\}_{n=0}^\infty$, and then the enumeration problem is to have an efficient description of the numerical sequence $a_n := |A_n|$, ready to be looked up (or submitted) to OEIS, and its corresponding sequence of polynomials $P_n(x) := |A_n|_x$.

It often happens that the statistic $f$, defined on $A_n$, has a scaled limiting distribution. In other words, if you draw a histogram of $f$ on $A_n$ and do the obvious scaling, then $f$ gets closer and closer to a certain continuous curve as $n$ goes to infinity.

The scaling is as follows. Let $E_n(f)$ and $\text{Var}_n(f)$ denote the expectation and variance of the statistic $f$ defined on $A_n$, and define the scaled random variable, for $a \in A_n$, by
\[ X_n(a) := \frac{f(a) - E_n(f)}{\sqrt{\text{Var}_n(f)}}. \]

If you draw the histograms of $X_n(a)$ for large $n$, they look practically the same, and they converge to some continuous limit.

A famous example is coin tossing. If $A_n$ is $\{-1, 1\}^n$, $v \in A_n$, and $f(v)$ is the sum of $v$, then the limiting distribution is the bell-shaped curve called the standard normal distribution, aka the Gaussian distribution.

As explained in [Z4], a purely finitistic approach to finding, and proving, a limiting scaled distribution is via the method of moments. Using symbolic computation, the computer can rigorously prove exact expressions for as many moments as desired, and often (as in the above case; see [Z4]) find a recurrence for the sequence of moments. This enables one to identify the limits of the scaled moments with the moments of the continuous limit; in the example of coin tossing (and many other cases),
\[ e^{-x^2/2} \sqrt{2\pi}, \]
whose moments are famously $1, 0, 1 \cdot 3, 0, 1 \cdot 3 \cdot 5, 0, 1 \cdot 3 \cdot 5 \cdot 7, 0, \cdots$. Whenever this is the case, the discrete family of random variables is said to be asymptotically normal. Whenever this is not the case, it is interesting and surprising.

**The Sum and Area Statistics for $a$-Parking Functions**

Let $P(n, a)$ be the set of $a$-parking functions of length $n$. A natural statistic is the sum
\[ \text{Sum}(p_1, \ldots, p_n) := p_1 + p_2 + \cdots + p_n = \sum_{i=1}^{n} p_i. \]

Another one, even more natural (see the beautiful article [Dj]), happens to be
\[ \text{Area}(p) := \frac{n(2a + n - 1)}{2} - \text{Sum}(p). \]

Let $P(n, a)(x)$ be the weighted analogue of $p(n, a)$, according to Sum, i.e.,
\[ P(n, a)(x) := \sum_{p \in P(n, a)} x^{\text{Sum}(p)}. \]

Analogously, let $Q(n, a)(x)$ be the weighted analogue of $p(n, a)$, according to Area, i.e.,
\[ Q(n, a)(x) := \sum_{p \in P(n, a)} x^{\text{Area}(p)}. \]

Clearly, one can easily go from one to the other:
\[ Q(n, a)(x) = x^{(2an + n - 1)n/2} P(n, a)(x^{-1}), \]
\[ P(n, a)(x) = x^{(2an + n - 1)n/2} Q(n, a)(x^{-1}). \]

How do we compute $P(n, a)(x)$ (or equivalently, $Q(n, a)(x)$)? It is readily seen that the analogue of (1) for the weighted counting is
\[ P(n, a)(x) = x^n \sum_{k=0}^{n} \binom{n}{k} P(n - k, a + k - 1)(x), \]
subject to the initial conditions $P(0, a)(x) = 1$ and $P(n, 0)(x) = 0$. So it is almost the same; the “only” change is sticking $x^n$ in front of the sum on the right-hand side. Equivalently,
\[ Q(n, a)(x) = \sum_{k=0}^{n} \binom{n}{k} x^{2k + 2a - 3/2} Q(n - k, a + k - 1)(x), \]
subject to the initial conditions $Q(0, a)(x) = 1$ and $Q(n, 0)(x) = 0$.

Once again, in the sense of Wilf, this is already an answer, but because of the extra variable $x$, one cannot go as far as we did before for the naive, merely numeric, counting.
It is very unlikely that there is a “closed-form” expression for $P(n, a(x))$ (and hence $Q(n, a(x))$), but for statistical purposes, it would be nice to get “closed-form” expressions for the expectation, the variance, and as many factorial moments as possible, from which the “raw” moments, and later the centralized moments, and finally the scaled moments, could be obtained. Then we could take the limits as $n$ goes to infinity, see whether they matched the moments of any of the known continuous distributions, and prove rigorously that at least for that many moments, the conjectured limiting distribution matches.

In our case, the limiting distribution is the intriguing Airy distribution, which Svante Janson prefers to call “area under Brownian excursion”. This result was stated and proved in [DJ], using deep and sophisticated continuous probability theory and continuous martingales. Here we will “almost” prove this result, in the sense of showing that the limits of the scaled moments of the area statistic on parking functions coincide with the scaled moments of the Airy distribution up to the 30th moment, and we can go much further.

But we can do much more than continuous probabilists. We (or rather our computers, running Maple) can find exact polynomial expressions in $n$ and the expectation $E_1(n)$. We can do so for any desired number of moments, say 30. Unlike the methods of continuous probability theorists, ours are entirely elementary, using only high-school algebra.

We can also do the same thing for the more general $a$-parking functions. Now the expressions are polynomials in $n$, $a$, and the expectation $E_1(n, a)$.

Finally, we believe that our approach, using the recurrence (3), can be used to give a full proof (for all moments) by doing it asymptotically and deriving a recurrence for the leading terms of the asymptotics for the factorial moments that would coincide with the well-known recurrence for the moments of the Airy distribution given, for example in Svante Janson’s article [J, eqs. 4, 5]. This is left as a challenge to our readers.

Finding the Expectation

The expectation of the sum statistic, let’s call it $E_{\text{sum}}(n, a)$, is given by

$$E_{\text{sum}}(n, a) = \frac{P(n, a)(1)}{P(n)(1)} = \frac{P(n, a)(1)}{a(n + 1)} \cdot$$

where the prime denotes, as usual, differentiation with respect to $x$.

Can we get a closed-form expression for $P(n, a)(1)$, and hence for $E_{\text{sum}}(n, a)$? Differentiating (2) with respect to $x$ using the product rule, we get

$$P(n, a)'(x) = x^n \sum_{k=0}^{n} \binom{n}{k} P(n-k, a+k-1)'(x)$$

$$+ nx^{n-1} \sum_{k=0}^{n} \binom{n}{k} P(n-k, a+k-1)(x).$$

Plugging in $x = 1$, we get that $P(n, a)'(1)$ satisfies the recurrence

$$P(n, a)'(1) - \sum_{k=0}^{n} \binom{n}{k} P(n-k, a+k-1)'(1)$$

$$= n \sum_{k=0}^{n} \binom{n}{k} P(n-k, a+k-1)(1) = n p(n, a).$$

(4)

Using this recurrence, we can, just as we did for $p(n, a)$ above, get expressions, as polynomials in $a$, for numeric $1 \leq n \leq 10$, say, and then conjecture that

$$P'(n, a)(1) = \frac{1}{2} an(a+n-1)(a+n)^{n-1}$$

$$- \frac{1}{2} \sum_{j=1}^{n} \binom{n}{j} j! a(a+n)^{n-j}.$$

To prove it, one plugs the left-hand side into (4), changes the order of summation, and simplifies. This is rather tedious, but since at the end of the day, these are equivalent to polynomial identities in $n$ and $a$, checking it for sufficiently many special values of $n$ and $a$ would be a rigorous proof.

It follows that

$$E_{\text{sum}}(n, a) = \frac{n(a+n+1)}{2} - \frac{1}{2} \sum_{j=1}^{n} \frac{n!}{(n-j)!(a+n)^{n-j}}.$$

This formula first appeared in [KY1]. Equivalently,

$$E_{\text{area}}(n, a) = \frac{n(a-2)}{2} + \sum_{j=1}^{n} \frac{n!}{(n-j)!(a+n)^{n-j}}.$$

In particular, for the primary object of interest, the case $a = 1$, we get

$$E_{\text{area}}(n, 1) = -\frac{n}{2} + \sum_{j=1}^{n} \frac{n!}{(n-j)!(n+1)^{n-j}}.$$

This rings a bell! It may written as

$$E_{\text{area}}(n, 1) = -\frac{n}{2} + \frac{1}{2} W_{n+1},$$

where $W_n$ is the iconic quantity

$$W_n = \frac{n!}{n^{n-1}} \sum_{k=0}^{n-1} \frac{n^k}{k!},$$

proved by Riordan and Sloane [RS] to be the expectation of another very important quantity, the sum of the heights on rooted labeled trees on $n$ vertices. In addition to its considerable mathematical interest, this quantity, $W_n$, has great historical significance. It was the first sequence, sequence A455, of the amazing On-Line Encyclopedia of Integer Sequences, now with almost 300, 000 sequences! See [EZ] for details and far-reaching extensions analogous to the present paper.2

---

2The reason it is not sequence A1 is that initially, the sequences were arranged in lexicographic order.
Another fact that will be of great use later in this paper is that as noted in [RS], Ramanujan and Watson proved that \( W_n \) (and hence \( W_{n+1} \)) is asymptotic to
\[
\frac{\sqrt{2\pi}}{2} n^{3/2}.
\]
It is very possible that the formula
\[
E_{\text{area}}(n, 1) = -\frac{n}{2} + \frac{1}{2} W_{n+1}
\]
may also be deduced from the Riordan–Sloane result via one of the numerous known bijections between parking functions and rooted labeled trees. More generally, the results below, for the special case \( a = 1 \), might be deduced from those of [EZ], but we believe that the present methodology is interesting for its own sake, and besides, in our current approach (which uses recurrences rather than the Lagrange inversion formula), it is much faster to compute higher moments, and hence going in the other direction would produce many more moments for the statistic on rooted labeled trees considered in [EZ], provided that there is indeed such a correspondence that sends the area statistic on parking functions (suitably tweaked) to the Riordan–Sloane statistic on rooted labeled trees.

The Limiting Distribution
Given a combinatorial family, one can easily get an idea of the limiting distribution by taking a large enough \( n \), say \( n = 100 \), generating a large enough number of random objects, say 50,000, and drawing a histogram; see Figure 2 in Diaconis and Hicks’s insightful article [DJ]. But one does not have to resort to simulation. While it is impractical to consider all \( 101^{100} \) parking functions of length 100, the generating function \( Q(100, 1)(x) \) contains the exact count for each conceivable area from 0 to \( \binom{109}{2} \). See Figure 1, which is a histogram of the area distribution of all parking functions of length 100.

But an even more informative way to investigate the limiting distribution is to draw the histogram of the probability generating function of the scaled distribution
\[
X_n(p) := \frac{\text{Area}(p) - E_n}{\sqrt{\text{Var}_n}},
\]
where \( E_n \) and \( \text{Var}_n \) are the expectation and variance respectively. See Figure 2, which is a histogram of the scaled area distribution of all parking functions of length 100.

As proved in [DJ] (using deep results in continuous probability due to David Aldous, Svante Janson, and Chassaing and Marckert), the limiting distribution is the Airy distribution. We will soon “almost” prove this, but we shall do much more by discovering exact expressions for the first 30 moments, not just their limiting asymptotics.

Truly Exact Expressions for the Factorial (and Hence Centralized) Moments
In [KY2] there is an “exact” expression for the general moment that is not very useful for our purposes. If one traces the proof, one can conceivably obtain explicit expressions for each specific moment, but the authors did not bother to implement it, and the asymptotics are not immediate.

We discovered the following important fact.

Proposition 1. Let \( E_1(n, a) := E_{\text{area}}(n, a) \) be the expectation of the area statistic on \( a \)-parking functions of length \( n \), given above, and let \( E_k(n, a) \) be the \( k \)th factorial moment
\[
E_k(n, a) := \frac{Q^{(k)}(n, a)(1)}{a(a + n)^{a-k+1}}.
\]

Then there exist polynomials \( A_k(n, a) \) and \( B_k(n, a) \) such that
\[
E_k(n, a) = A_k(n, a) + B_k(n, a)E_1(n, a).
\]

The beauty of experimental mathematics is that these polynomials can be found by cranking out enough data, using the sequence of probability generating functions \( Q(n, a)(x) \) obtained using the recurrence (3), calculating sufficiently many numerical data for the moments, and using undetermined coefficients. These can be proved a posteriori by taking these truly exact formulas and verifying the implied recurrences for the \( k \)th factorial moment (obtained from differentiating (3) \( k \) times, using Leibniz’s rule) in terms of the previous ones. But this is not necessary. Since at the end of the day, it all boils down to verifying

Figure 1. The area distribution of all parking functions of length 100.

Figure 2. The scaled area distribution of all parking functions of length 100.
Theorem 7. The fourth factorial moment of the area statistic on parking functions of length $n$ is
\[
\frac{221}{1008} n^6 + \frac{64377}{30240} n^5 + \frac{101897}{15120} n^4 + \frac{22217}{5040} n^3
\]
\[
- \frac{1375}{189} n^2 + \frac{187463}{30240} n
\]
\[
+ \left( - \frac{35}{16} n \right) + \frac{449}{27} n^2 - \frac{130243}{2520} n^2 - \frac{749}{105} n - \frac{503803}{15120} \right) E_1(n),
\]
and asymptotically it equals $\frac{221}{1008} n^6 + O(n^{11/2})$.

Theorem 8. The fifth factorial moment of the area statistic on parking functions of length $n$ is
\[
\frac{105845}{110592} n^7 + \frac{2170159}{290304} n^6 - \frac{9995651}{387072} n^5 + \frac{20773609}{725760} n^4
\]
\[
+ \frac{94846903}{11612160} n^3 + \frac{24676991}{483840} n^2 + \frac{392763901}{11612160} n
\]
\[
+ \left( \frac{565}{128} n^6 + \frac{1005}{16} n^5 + \frac{9832585}{128} n^4 + \frac{1111349}{165988} n^3
\]
\[
+ \frac{826358527}{1935360} n^2 + \frac{159943787}{362880} n + \frac{1024580441}{5800608} \right) E_1(n),
\]
and asymptotically it equals $\frac{565}{8972} n^{5/2} + O(n^7)$.

Theorem 9. The sixth factorial moment of the area statistic on parking functions of length $n$ is
\[
\frac{82825}{576576} n^8 + \frac{375340075}{110702592} n^7 + \frac{9401544029}{33210776} n^6 + \frac{14473244813}{127733760} n^5
\]
\[
+ \frac{14413396709}{1660538880} n^4 + \frac{88215445651}{33210776} n^3 + \frac{8878316473}{33210776} n^2
\]
\[
+ \frac{64359542029}{1660538880} n^2 + \frac{35893654049}{1660538880} + \frac{259283273}{11612160} n
\]
\[
+ \left( \frac{3955}{2048} n^7 + \frac{186549}{6144} n^6 - \frac{250283273}{11612160} n^5 - \frac{119912501}{129024} n^4
\]
\[
+ \frac{149860633081}{63866880} n^4 - \frac{60179426581}{1660538880} n^3 - \frac{86000570107}{276756480} n^2
\]
\[
+ \frac{92139038089}{830269440} \right) E_1(n),
\]
and asymptotically it equals $\frac{82825}{576576} n^8 + O(n^{11/2})$.

For the analogous theorems on the factorial moments 7–30, see the output file http://sites.math.rutgers.edu/~zeilberg/okniot/0ParkingStatistics7.txt.

Let \( \{e_k\}_{k=1}^\infty \) be the sequence of moments of the Airy distribution, defined by the recurrence given in Svante Janson’s interesting survey paper [1, eqs. 4, 5]. Our computers, using our Maple package, proved that
\[
E_k(n) = e_k n^{2k} + O\left( n^{2k - 1} \right),
\]
for $1 \leq k \leq 30$. It follows that the limiting distribution of the area statistic is (most probably) the Airy distribution, since the first 30 moments match. Of course, this was already known to continuous probability theorists, and we have proved it only for the first 30 moments. But:

- Our methods are purely elementary and finitistic.
- We can easily go much further, i.e., prove it for more moments.
- We believe that our approach, using recurrences, can be used to derive a recurrence for the leading asymptotics of the factorial moments $E_k(n)$, which would turn out to be the same as the above-mentioned recurrence [J, eqs. 4, 5]. We leave this as a challenge to the reader.

To see expressions in $a$, $n$, and $E_1(n,a)$ for the first ten moments of $a$-parking, see http://sites.math.rutgers.edu/~zeilberg/tokhniot/oParkingStatistics8.txt.

**ACKNOWLEDGMENTS**

Many thanks are due to Valentin Féray and Svante Janson for insightful information and useful references. Also thanks to Benji Weiss for comments on a previous version. Finally, we thank an anonymous referee, and the nonanonymous editor, Sergei Tabachnikov, for comments that improved the clarity of our article.

Yukun Yao  
Department of Mathematics  
Rutgers University (New Brunswick)  
Hill Center-Busch Campus  
110, Frelinghuysen Rd.  
Piscataway, NJ 08854-8019  
USA  
e-mail: yao@math.rutgers.edu

Doron Zeilberger  
Department of Mathematics  
Rutgers University (New Brunswick)  
Hill Center-Busch Campus  
110, Frelinghuysen Rd.  
Piscataway, NJ 08854-8019  
USA  
e-mail: DoronZeil@gmail.com

**REFERENCES**

[DJ] Persi Diaconis and Angela Hicks. Probabilizing parking functions. *Adv. in Appl. Math.* 89 (2017), 125-155.

[EZ] Shalosh B. Ekhad and Doron Zeilberger. Going back to Neil Sloane’s first love (OEIS Sequence A435): on the total heights in rooted labeled trees. *The Personal Journal of Shalosh B. Ekhad and Doron Zeilberger*, July 19, 2016. Available at http://sites.math.rutgers.edu/~zeilberg/mamarim/mamarimhtml/a435.html (accessed May 24, 2018).

[FR] Dominique Foata and John Riordan. Mapping of acyclic and parking functions. *Aequationes Mathematicae* 10 (1974), 490-515.

[J] Svante Janson. Brownian excursion area, Wright’s constants in graph enumeration, and other Brownian areas. *Probability Surveys* 4 (2007), 80-145.

[KW] Alan G. Konheim and Benjamin Weiss. An occupancy discipline and applications. *SIAM J. Applied Math.* 14 (1966), 1266-1274.

[KY1] J.P. Kung and C. Yan. Expected sums of general parking functions. *Annals of Combinatorics* 7 (2003), 481-493.

[KY2] J.P. Kung and C. Yan. Exact formulas for the moments of sums of classical parking functions. *Advances in Applied Mathematics* 31 (2003), 215-241.

[RS] John Riordan and Neil J.A. Sloane. The enumeration of rooted trees by total height. *J. Australian Math. Soc.* 10 (1969), 278-282.

[Sc] Marcel-Paul Schützenberger. On an enumeration problem. *J. Combinatorial Theory* 4 (1968), 219-221.

[St1] Richard Stanley. Parking functions. Available at www-math.mit.edu/~rstan/transparencies/parking.pdf (accessed May 24, 2018).

[St2] Richard Stanley. A survey of parking functions. Available at www-math.mit.edu/~rstan/transparencies/parking3.pdf (accessed May 24, 2018).

[St3] Richard Stanley. *Enumerative Combinatorics*, Volume 2. Cambridge University Press, 1999.

[W] Herbert S. Wilf. What is an answer? *American Mathematical Monthly* 89 (1982), 289-292.

[Z1] Doron Zeilberger. Symbolic moment calculus I: Foundations and permutation pattern statistics. *Annals of Combinatorics* 8 (2004), 369-378.

[Z2] Doron Zeilberger. Symbolic moment calculus II: Why is Ramsey theory sooooo eeeenormously hard? *Integers* 7(2) (2007), A34.

[Z3] Doron Zeilberger. The $n^{-2}$th proof for the number of labeled trees. *The Personal Journal of Shalosh B. Ekhad and Doron Zeilberger*, undated (c. 1998). Available at http://sites.math.rutgers.edu/~zeilberg/mamarim/mamarimhtml/labtree.html (accessed May 24, 2018).

[Z4] Doron Zeilberger. The automatic central limit theorems generator (and much more)! In *Advances in Combinatorial Mathematics: Proceedings of the Waterloo Workshop in Computer Algebra 2008 in honor of Georgy P. Egorychev*, edited by I. Kotsireas and E. Zima, Chapter 8, pp. 165-174. Springer, 2009.