A simple case of the Mahonian statistic: A saddle point approach

Guy Louchard

March 14, 2010

Outline

1. Introduction
2. The Gaussian limit, $j = \mu + x\sigma$
3. The Large deviation, $j = \mu + xn^{7/4}$
4. Justification of the integration procedures

Introduction

In [1], Canfield, Janson and Zeilberger analyze the Mahonian distribution on multiset permutations: classic permutations on m objects can be viewed as words in the alphabet $\{1, \ldots, m\}$. If we allow *repetitions*, we can consider all words with a_1 occurences of 1, a_2 occurences of $2, \ldots$, a_m occurences of m. Let J_m denote the number of inversions. Assuming that *all words are equally likely*, the probability generating function of J_m is given, setting $N = a_1 + \cdots + a_m$, by

$$\phi_{a_1,\ldots,a_m}(z) = \frac{\prod_{i=1}^m a_i! \prod_{i=1}^N (1 - z^i)}{N! \prod_{j=1}^m \prod_{i=1}^{a_j} (1 - z^i)}.$$

The mean μ and variance σ^2 are given by

$$\mu(J_m) = e_2(a_1, \ldots, a_m)/2, \quad \sigma^2(J_m) \quad = \frac{(e_1 + 1)e_2 - e_3}{12},$$

where $e_k(a_1, \ldots, a_m)$ is the degree k elementary symmetric function.

Let $a^* = \max_j a_j$ and $N^* = N - a^*$. In [1], the authors prove that, if $N^* \to \infty$ then the sequence of normalized random variables

$$\frac{J_m - \mu(J_m)}{\sigma(J_m)}$$

tends to the *standard normal distribution*. They also conjecture a local limit theorem and prove it under additional hypotheses.
In this talk, we analyze simple examples of the Mahonian statistic, for instance, we consider the case

$$m = 2, a_1 = an, a_2 = bn, n \to \infty.$$

We analyze the central region $j = \mu + x\sigma$ and one large deviation $j = \mu + xn^{7/4}$. The exponent $7/4$ that we have chosen is of course not sacred, any fixed number below 2 and above $3/2$ could also have been considered.

We have here

$$\phi(z) = \frac{(an)!(bn)!\prod_{i=1}^{(a+b)n}(1-z^i)}{((a+b)n)!\prod_{i=1}^{an}(1-z^i)\prod_{i=1}^{bn}(1-z^i)},$$

$$\mu = \frac{abn^2}{2},$$

$$\sigma^2 = \frac{ab(a+b+1/n)n^3}{12}.$$

By Cauchy's theorem,

$$Z_2(j) := \mathbb{P}(J_2 = j) = \frac{1}{2\pi\mathbf{i}} \int_\Omega \frac{\phi(z)}{z^{j+1}} dz$$
$$= \frac{1}{2\pi\mathbf{i}} \int_\Omega e^{S(z)} dz, \qquad (1)$$

where Ω is inside the analyticity domain of the integrand and encircles the origin and

$$S(z) = S_1(z) + S_2(z), \qquad (2)$$
$$S_1(z) = \sum_{i=1}^{(a+b)n} \ln(1 - z^i) - \sum_{i=1}^{an} \ln(1 - z^i) - \sum_{i=1}^{bn} \ln(1 - z^i)$$
$$- \ln(((a + b)n)!) + \ln(an!) + \ln(bn!),$$
$$S_2(z) = -(j + 1) \ln(z).$$

Set

$$S^{(i)} := \frac{d^i S}{dz^i}.$$

Continuing the approach we used in [3] for classical inversions in permutations, we will use the *Saddle point method* (for a good introduction to this method, see Flajolet and Sedgewick [2, ch. *VIII*]. We obtain here *local limit theorems* with some corrections of order $1/n$.

The talk is organized as follows: Section 2 deals with the Gaussian limit. In Section 3, we analyze the case $j = \mu + xn^{7/4}$ (large deviation). Section 4 provides the justification of the integration procedures.

The Gaussian limit, $j = \mu + x\sigma$

The saddle point

To use the saddle point method, we must find the solution of

$$S^{(1)}(\tilde{z}) = 0 \quad (3)$$

with *smallest module*. Set $\tilde{z} := z^* - \varepsilon$, where $z^* = \lim_{n\to\infty} \tilde{z}$ .Here, it is easy to check that $z^* = 1$. This leads, to first order, to

$$[abn^2/2 - j - 1] + [-1 - n^3 ab(a+b)/12 + 5n^2 ab/12 - j]\varepsilon = 0. \quad (4)$$

Set $j = \mu + x\sigma$ in (4). This shows that, asymptotically, ε is given by a *Puiseux series* of powers of $n^{-1/2}$, starting with $-\frac{2x3^{1/2}}{[ab(a+b)]^{1/2} n^{3/2}}$.

To obtain the next terms, we compute the next terms in the expansion of (3). Even powers ε^{2k} lead to a $\mathcal{O}(n^{2k+1}) \cdot \varepsilon^{2k}$ term and odd powers ε^{2k+1} lead to a $\mathcal{O}(n^{2k+3}) \cdot \varepsilon^{2k+1}$ term. Now we set $j = \mu + x\sigma$, expand into powers of $n^{-1/2}$ and equate each coefficient with 0. This leads successively to a *full expansion of* ε. Note that to obtain a given precision of ε, it is enough to compute a given finite number of terms in the generalization of (4). We obtain (in all our asymptotics, we will provide only a few terms, but Maple knows more)

$$\varepsilon = -\frac{2x3^{1/2}}{[ab(a+b)]^{1/2}n^{3/2}} - \frac{(2x^2a^2 - 5ab + 2abx^2 + 2b^2x^2)3^{1/2}x}{5[ab(a+b)]^{3/2}n^{5/2}} - \frac{6(x^2+2)}{ab(a+b)n^3} + \dots$$

We have, with $\tilde{z} := z^* - \varepsilon = 1 - \varepsilon$,

$$Z_2(j) = \frac{1}{2\pi \mathbf{i}} \int_{\Omega} \exp\left[S(\tilde{z}) + S^{(2)}(\tilde{z})(z-\tilde{z})^2/2! + \sum_{l=3}^{\infty} S^{(l)}(\tilde{z})(z-\tilde{z})^l/l! \right] dz.$$

(note carefully that the linear term vanishes). Set $z = \tilde{z} + \mathbf{i}\tau$. This gives

$$Z_2(j) = \frac{1}{2\pi} \exp[S(\tilde{z})] \int_{-\infty}^{\infty} \exp\left[S^{(2)}(\tilde{z})(\mathbf{i}\tau)^2/2! + \sum_{l=3}^{\infty} S^{(l)}(\tilde{z})(\mathbf{i}\tau)^l/l! \right] d\tau. \tag{5}$$

Let us first analyze $S(\tilde{z})$. To compute $S_1(\tilde{z})$, we first compute the asymptotics of the i term, this leads to a $\ln(i)$ contribution, which will be cancelled by the factorials. Next we sum on i. We obtain

$$S_1(\tilde{z}) = \frac{3^{1/2}(ab)^{1/2}x}{(a+b)^{1/2}}n^{1/2} + \frac{x^2}{2} + \frac{(2x^2a^2 - 5ab + 2abx^2 + 2b^2x^2)3^{1/2}x}{10[ab(a+b)]^{3/2}n^{1/2}} + \frac{a^2x^4 + 30ab + abx^4 + b^2x^4}{5ab(a+b)n} + \mathcal{O}(1/n^{3/2}),$$

$$S_2(\tilde{z}) = -\frac{3^{1/2}(ab)^{1/2}x}{(a+b)^{1/2}}n^{1/2} - x^2 - \frac{(2x^2a^2 - 5ab + 2abx^2 + 2b^2x^2)3^{1/2}x}{10[ab(a+b)]^{3/2}n^{1/2}} - \frac{a^2x^4 + 30ab + abx^4 + b^2x^4}{5ab(a+b)n} + \mathcal{O}(1/n^{3/2}),$$

and so

$$S(\tilde{z}) = -x^2/2 + \mathcal{O}(n^{-3/2}).$$

Also, again computing the asymptotics of the i term and summing on i,

$$S^{(2)}(\tilde{z}) = \frac{ab(a+b)n^3}{12} + \left[\frac{ab}{12} - \frac{1}{20}(a^2+b^2+ab)\right] x^2 n^2 + \mathcal{O}(n^{3/2}),$$

$$S^{(3)}(\tilde{z}) = -\frac{(a^3+2ba^2+2b^2a+b^3)(ab)^{1/2}3^{1/2}x}{60(a+b)^{1/2}} n^{7/2} + \mathcal{O}(n^3),$$

$$S^{(4)}(\tilde{z}) = -\frac{(a^3+2ba^2+2b^2a+b^3)abn^5}{120} + \mathcal{O}(n^4),$$

$$S^{(l)}(\tilde{z}) = \mathcal{O}(n^{l+1}), \quad l \geq 5.$$

Note that, with $z = \tilde{z}e^{\mathbf{i}\theta}$, this leads to

$$S^{(2)}(\tilde{z})\frac{(z-\tilde{z})^2}{2} \sim -n^3\frac{ab(a+b)}{24}\theta^2. \qquad (6)$$

Integration

We can now compute (5), for instance by using the classical trick of setting

$$S^{(2)}(\tilde{z})(\mathbf{i}\tau)^2/2! + \sum_{l=3}^{\infty} S^{(l)}(\tilde{z})(\mathbf{i}\tau)^l/l! = -u^2/2,$$

and computing (by inversion) τ as a *truncated series in u*. This leads to

$$\tau = \frac{1}{n^{3/2}}[\alpha_1 u + \alpha_2 u^2 + \alpha_3 u^3 + \alpha_4 u^4 + \dots],$$

$$\alpha_1 = \frac{3^{1/2} 10ab(a+b)}{5[ab(a+b)]^{3/2}} + \frac{3^{1/2}(3abx^2 + 3x^2a^2 - 5ab + 3b^2x^2)}{5[ab(a+b)]^{3/2} n} + \dots,$$

$$\alpha_2 = \frac{2\mathbf{i}/5(ab + a^2 + b^2)3^{1/2}x}{[ab(a+b)]^{3/2} n} + \dots,$$

$$\alpha_3 = -\frac{3^{1/2}(ab + a^2 + b^2)(a+b)^{1/2}}{10(a+b)^2(ab)^{3/2} n} + \dots,$$

$$\alpha_4 = \mathbf{i}\mathcal{O}(1/n^2).$$

Setting $d\tau = \frac{d\tau}{du} du$, expanding w.r.t. n , integrating on $u = [-\infty \ldots \infty]$, (Note that α_2 is not useful here), finally (5) leads to

Theorem 2.1

$$Z_2(j) \sim e^{-x^2/2}.$$
$$\exp\left[\left[-\frac{3a^2+13ab+3b^2}{20ab(a+b)} + \frac{3(a^2+b^2+ab)x^2}{10ab(a+b)}\right]\Big/ n + \mathcal{O}(n^{-3/2})\right]$$
$$\Big/(\pi ab(a+b)n^3/6)^{1/2} \,. \tag{7}$$

Note that $S^{(3)}(\tilde{z})$ does not contribute to the $1/n$ correction.

To check the effect of the correction, we first give in Figure 1, for $n = 150, a = b = 1/2$, (the same values are used in this section) the comparison between $J_n(j)$ and the asymptotics (2.1), without the $1/n$ term.

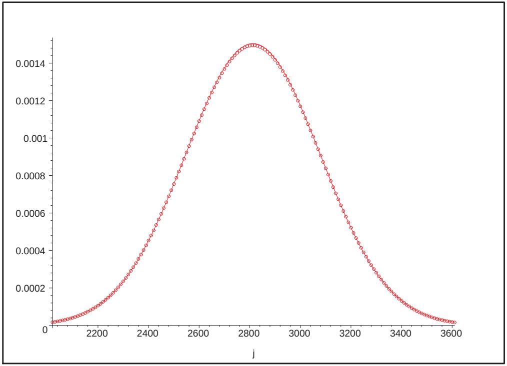

Figure 1: Comparison between $J_n(j)$ and the asymptotics (2.1), without the $1/n$ term

Figure 2 gives the same comparison, with the $1/n$ correction.

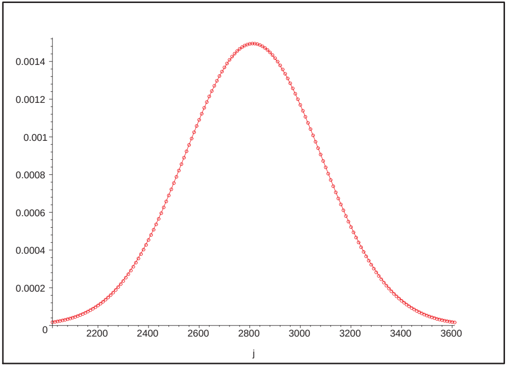

Figure 2: Comparison between $J_n(j)$ and the asymptotics (2.1), with the $1/n$ correction

Figure 3 shows the quotient of $Z_2(j)$ and the asymptotics (2.1), with the constant term $1/n$. The “hat” behaviour, already noticed in the classical permutation inversion analysis, is also apparent here
.

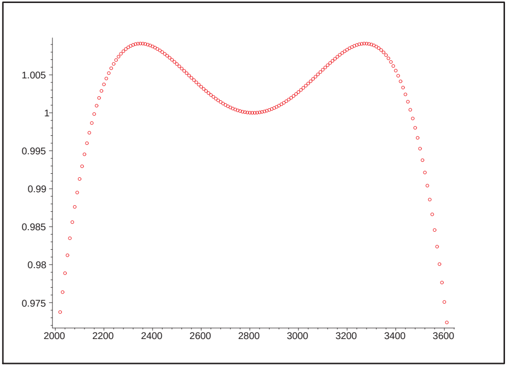

Figure 3: The quotient of $Z_2(j)$ and the asymptotics (2.1), with the constant $1/n$ term

Finally, Figure 4 shows the quotient of $Z_2(j)$ and the asymptotics (2.1), with the full $1/n$ correction (constant and x^2 term).

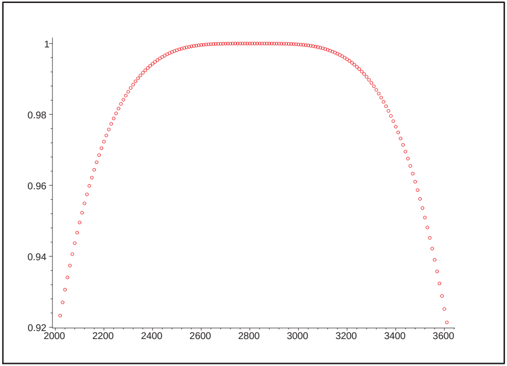

Figure 4: The quotient of $Z_2(j)$ and the asymptotics (2.1), with the full $1/n$ correction

The Large deviation, $j = \mu + xn^{7/4}$

The saddle point

Now we consider the case $j = \mu + xn^{7/4}$. Again, we have here $z^* = 1$. We observe the same behaviour as in Section 2 for the coefficients of ε in the generalization of (4).
Proceeding as before, we see that asymptotically, ε is now given by a *Puiseux series* of powers of $n^{-1/4}$, starting with $-\frac{12x}{ab(a+b)n^{5/4}}$.
This leads to (we provide only the first two terms, the other ones are rather complicated, but we use them up to the n^{-3} term)

$$\varepsilon = -\frac{12x}{ab(a+b)n^{5/4}} - \frac{144(a^2+ab+b^2)x^3}{5[ab(a+b)]^3 n^{7/4}} + \dots .$$

This gives ($C_i(x,a,b)$ are complicated functions, not given here)

$$S(\tilde{z}) = -\frac{6x^2}{ab(a+b)} n^{1/2} - \frac{36(a^2+ab+b^2)x^4}{5[ab(a+b)]^3} + C_1(x,a,b)/n^{1/2} + C_2(x,a,b)/n + \dots$$

Also,

$$S^{(2)}(\tilde{z}) = \frac{ab(a+b)}{12}n^3 - \frac{3(a^2+ab+b^2)x^2}{5ab(a+b)}n^{5/2} + C_3(x,a,b)n^2 - 2xn^{7/4} + \mathcal{O}(n^{3/2}),$$

$$S^{(3)}(\tilde{z}) = -\frac{(a^2+ab+b^2)x}{10}n^{15/4} - \frac{ab(a+b)}{4}n^3 + \mathcal{O}(n^{11/4}),$$

$$S^{(4)}(\tilde{z}) = -\frac{(a^3+2a^2b+2ab^2+b^3)ab}{120}n^5 + \mathcal{O}(n^{19/4}),$$

$$S^{(l)}(\tilde{z}) = \mathcal{O}(n^{l+1}), \quad l \geq 5,$$

Integration

Now

$$\tau = \frac{1}{n^{3/2}}[\alpha_1 u + \alpha_2 u^2 + \alpha_3 u^3 + \ldots],$$
$$\alpha_1 = C_4(x, a, b) + C_5(x, a, b)/n^{1/2} + C_6(x, a, b)/n + \ldots,$$
$$\alpha_2 = \mathbf{i} C_7(x, a, b)/n^{3/4} + \ldots,$$
$$\alpha_3 = C_8(x, a, b)/n^{3/2}.$$

and finally we obtain

Theorem 3.1

$$Z_2(j) \sim \exp\left[-\frac{6x^2}{ab(a+b)}n^{1/2} - \frac{36(a^2+ab+b^2)x^4}{5[ab(a+b)]^3} + C_9(x,a,b)/n^{1/2} + C_{10}(x,a,b)/n + \dots\right] \Big/ (\pi ab(a+b)n^3/6)^{1/2} . \quad (8)$$

Note that $S^{(3)}(\tilde{z})$ does not contribute to the correction. Of course, the dominant term is null for $x = 0$.

To check the effect of the correction, we first give in Figure 5, for $n = 50, a = b = 1/2$ and $x \in [0..0.2]$, the comparison between $J_n(j)$ and the asymptotics (8).

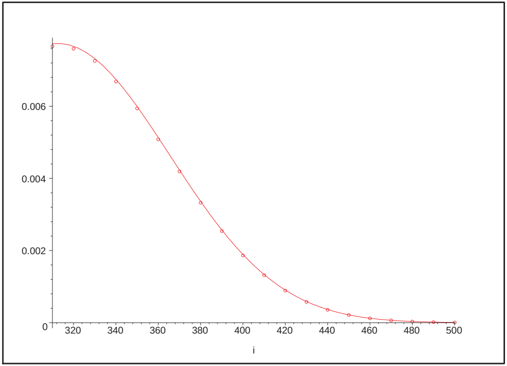

Figure 5: The comparison between $Z_2(j)$ and the asymptotics (8).

Figure 6 shows the quotient of $Z_2(j)$ and the asymptotics (8).

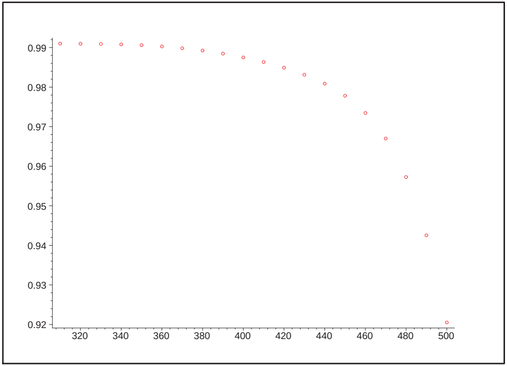

Figure 6: The quotient of $Z_2(j)$ and the asymptotics (2.1)

Justification of the integration procedures

Splitting value

Let us first analyze

$$F(r) := \sum_1^r \ln(1 - z^k)$$

for $z = e^{\mathbf{i}\theta}$. We have

$$F(j) = \sum_1^j \ln\left(1 - e^{\mathbf{i}k\theta}\right) = \sum_1^j \ln\left(\frac{1 - e^{\mathbf{i}k\theta}}{-\mathbf{i}k\theta}\right) + \sum_1^j \ln(k) + j\ln(-\mathbf{i}\theta)$$

$$= \sum_1^j \ln\left(\frac{e^{-\mathbf{i}k\theta/2} - e^{\mathbf{i}k\theta/2}}{-\mathbf{i}k\theta}\right) + \sum_1^j \frac{\mathbf{i}k\theta}{2} + \ln(j!) + j\ln(-\mathbf{i}\theta)$$

$$= \sum_1^j \ln\left(\frac{2\sin(k\theta/2)}{k\theta}\right) + \frac{j(j+1)}{2}\frac{\mathbf{i}\theta}{2} + \ln(j!) + j\ln(-\mathbf{i}\theta).$$

So, from (2), with $j = \mu + x\sigma$ or $j = \mu + xn^{7/4}$,

$$S(e^{\mathbf{i}\theta}) = \sum_{1}^{(a+b)n} \ln\left(\frac{2\sin(k\theta/2)}{k\theta}\right) - \sum_{1}^{an} \ln\left(\frac{2\sin(k\theta/2)}{k\theta}\right) - \sum_{1}^{bn} \ln\left(\frac{2\sin(k\theta/2)}{k\theta}\right) + \mathcal{O}(\mathbf{i}\theta n^{\alpha}),$$

where $\alpha = 3/2$ in the Gaussian case and $\alpha = 7/4$ in the large deviation case.

Note that, for small θ, we have

$$\frac{2\sin(k\theta/2)}{k\theta} \sim 1 - \frac{k^2\theta^2}{24},$$

$$\ln\left(\frac{2\sin(k\theta/2)}{k\theta}\right) \sim -\frac{k^2\theta^2}{24},$$

$$\sum_1^j \ln\left(\frac{2\sin(k\theta/2)}{k\theta}\right) \sim -\frac{j^3}{3}\frac{\theta^2}{24},$$

so

$$S(e^{\mathbf{i}\theta}) \sim -ab(a+b)\frac{\theta^2}{24}n^3$$

which conforms to (6).

Proceeding now as in [2, ch. *VIII*], we introduce a *splitting value θ_0* such that $n^3\theta_0^2 \to \infty, S^{(3)}\theta_0^3 \to 0, n \to \infty$, where $S^{(3)} \sim n^{7/2}$ in the Gaussian case and $S^{(3)} \sim n^{15/4}$ in the large deviation case. For instance, we choose $\theta_0 = n^{-4/3}$.
Let us now turn to (1) which leads to

$$\frac{1}{2\pi}\int_{\theta_0}^{2\pi-\theta_0} e^{S(e^{\mathbf{i}\theta})} e^{\mathbf{i}\theta} d\theta.$$

Singularity analysis

We will use some singularity analysis. We have

$$e^{S(e^{\mathrm{i}\theta})} \sim \frac{\prod_{k=bn+1}^{bn+an} \frac{2\sin(k\theta/2)}{k\theta}}{\prod_{j=1}^{an} \frac{2\sin(j\theta/2)}{j\theta}} e^{\mathcal{O}(\mathrm{i}\theta n^{\alpha})}. \tag{9}$$

For every fixed j^*, $\sin(j^*\theta/2)$ is null at $\theta = \ell\theta^*$, $\theta^* = \frac{2\pi}{j^*}$, $\ell = 1, \ldots, j^*$. But then, $\frac{k\theta^*}{2} = \frac{k\pi}{j^*}$ must be some integer multiple of π, $r^*\pi$, say, in order to *compensate the pole* in (9) at $\theta = \theta^*$. i.e. there must exist some k^* such that $k^* = r^*j^*$. But, as $bn + 1 \leq k \leq bn + an$, and $1 \leq j^* \leq an$, there exists always *at least one value k^**. Also, for integer ℓ (in the sequel, $\ell, \ell_1, \ldots$ are always integers),

$$\frac{\ell k^* \theta^*}{2} = \frac{\ell k^* \pi}{j^*} = \ell r^* \pi,$$

so multiples of θ^* are compensated by the same k^* as for θ^*.

Now,

$$e^{S(e^{\mathbf{i}\theta})} \sim \frac{(an)!(bn)!}{((a+b)n)!}\psi(n,\theta),$$

with

$$\psi(n,\theta) := \frac{\prod_{k=bn+1}^{bn+an} \sin(k\theta/2)}{\prod_{j=1}^{an} \sin(j\theta/2)} e^{\mathcal{O}(\mathbf{i}\theta n^{\alpha})},$$

and it remains to prove that

$$\frac{(an)!(bn)!}{((a+b)n)!} \int_{\theta_0}^{2\pi-\theta_0} \psi(n,\theta) d\theta$$

tends to 0.

Firstly, if we choose a pole $\theta = \ell_1 2\pi/j_1^*$, we can sometimes choose another value j_2^* leading to a pole of $\sin(\theta j/2)$. Indeed, it is enough to have $\theta = \ell_1 2\pi/j_1^* = \ell_2 2\pi/j_2^*$, or $\ell_1 j_2^* = \ell_2 j_1^*$. Now choose for instance, $j^* = 2, an = 10, bn = 20, \theta^* = \pi$. The possible values for j are $j \in \{2, 4, 6, 8, 10\}$ and we must choose $k \in \{22, 24, 26, 28, 30\}$. More generally, with $\theta = 2\pi/j^*, \beta = \lfloor an/j^* \rfloor$, we have β possible values of j leading to poles and there are at least β *possible compensating values for k*.
Let us consider

$$\frac{\sin(\frac{k_\ell \theta}{2})}{\sin(\frac{j_\ell \theta}{2})}, \quad \theta = \frac{2\pi}{j^*}, \quad \ell = 1, \ldots, \beta.$$

We have

$$j_\ell = \ell j^*, \quad k_\ell = \left\lceil \frac{bn+1}{j^*} \right\rceil j^* + (\ell - 1) j^*.$$

Also

$$|\sin(j_\ell(\theta + \varepsilon)/2)| \sim \ell j^* \varepsilon/2, \quad |\sin(k_\ell(\theta + \varepsilon)/2)| \sim k_\ell \varepsilon/2.$$

Bounds at the poles

So the contribution of the poles to $\psi(n,\theta)$ is bounded by

$$\mathcal{O}\left(\frac{(bn+an)^{an/j*}}{j^{*\,an/j^*}(an/j^*)!}\right).$$

This is maximum for $j^* = 2, \quad \theta = \pi$. So, at the poles, $|\exp(S(e^{\mathbf{i}\theta}))|$ is bounded by

$$\mathcal{O}\left(\frac{(an)!(bn)!}{((a+b)n)!}\frac{(bn+an)^{an/2}}{2^{an/2}(an/2)!}\right) \sim \mathcal{O}\left(\frac{e^{an/2}}{(1+b/a)^{an/2}(1+a/b)^{bn}}\right),$$

and, if we choose (as we may) $a \leq b$, this tends *exponentially* to 0 as $n \to \infty$ in the form $\exp(-Dn), D > 0$.

PROBLEM: BOUND $|\psi(n,\theta)|$ FOR θ DIFFERENT FROM THE POLES OF $\psi(n,\theta)$, SO THAT WE CAN OBTAIN CENTRAL APPROXIMATION AND TAIL COMPLETION

E. R. Canfield, S. Janson, and D. Zeilberger.
The mahonian probability distribution on words is asymptotically normal.
Technical report, 2009.

P. Flajolet and R. Sedgewick.
Analytic combinatorics.
Cambridge University press, 2009.

G. Louchard and H. Prodinger.
The number of inversions in permutations: A saddle point approach.
Journal of Integer Sequences, 6:03.2.8, 2003.