Journal of Integer Sequences, Vol. 4 (2001),
Article 01.2.4

Permutations with Inversions

Barbara H. Margolius
Cleveland State University
Cleveland, Ohio 44115

Email address: b.margolius@csuohio.edu

Abstract

The number of inversions in a random permutation is a way to measure the extent to which the permutation is "out of order". Let $I_n(k)$ denote the number of permutations of length n with k inversions. This paper gives asymptotic formulae for the sequences $\{I_{n+k}(n), n = 1, 2, \ldots\}$ for fixed k.

1. Introduction Let $a_1, a_2, \ldots, a_n$ be a permutation of the set $\{1, 2, \ldots, n\}$. If $i < j$ and $a_i > a_j$, the pair (a_i, a_j) is called an "inversion" of the permutation; for example, the permutation 3142 has three inversions: (3,1), (3,2), and (4,2). Each inversion is a pair of elements that is "out of sort", so the only permutation with no inversions is the sorted permutation.

2. Generating Function Let $I_n(k)$ represent the number of permutations of length n with k inversions.

Theorem 1 (Muir,1898). *[10] The numbers $I_n(k)$ have as generating function*

$$\Phi_n(x) = \sum_{k=0}^{\binom{n}{2}} I_n(k)x^k$$

$$= \prod_{j=1}^{n} \sum_{k=0}^{j-1} x^k$$
$$= \prod_{j=1}^{n} \frac{1-x^j}{1-x}.$$

Clearly the number of permutations with no inversions, $I_n(0)$, is 1 for all n, and in particular $I_1(0) = 1 = \Phi_1(x)$. So the formula given in the theorem is correct for $n = 1$. Consider a permutation of $n-1$ elements. We insert the nth element in the jth position, $j = 1, 2, \ldots, n$, choosing the insertion point randomly. Since the nth element is larger than the $n-1$ elements in the set $\{1, 2, \ldots, n-1\}$, by inserting the element in the jth position, $n-j$ additional inversions are added. The generating function for the number of additional inversions is $1 + x + x^2 + \cdots + x^{n-1}$ since each number of additional inversions is equally likely. The additional inversions are independent from the inversions present in the permutation of length $n-1$, so the total number of inversions has as its generating function the product of the generating function for $n-1$ inversions and the generating function for the additional inversions:

$$\Phi_n(x) = (1 + x + \cdots + x^{n-1})\Phi_{n-1}(x).$$

The required result then follows by induction.

Below is a table of values of the number of inversions (see sequence A008302 in [13], also [2], [3], [8], [11]):

Table 1 $I_n(k) = I_n(\binom{n}{2} - k)$

	k, number of inversions													
$n\backslash k$	0	1	2	3	4	5	6	7	8	9	10	11	12	13
1	**1**													
2	1	**1**												
3	1	2	**2**	1										
4	1	3	5	**6**	5	3	1							
5	1	4	9	15	**20**	22	20	15	9	4	1			
6	1	5	14	29	49	**71**	90	101	101	90	71	49	29	14
7	1	6	20	49	98	169	**259**	359	455	531	573	573	531	455
8	1	7	27	76	174	343	602	**961**	1415	1940	2493	3017	3450	3736
9	1	8	35	111	285	628	1230	2191	**3606**	5545	8031	11021	14395	17957
10	1	9	44	155	440	1068	2298	4489	8095	**13640**	21670	32683	47043	64889

Table 1 (continued) $I_n(k) = I_n(\binom{n}{2} - k)$										
	k, number of inversions									
$n\backslash k$	14	15	16	17	18	19	20	21	22	23
6	5	1								
7	359	259	169	98	49	20	6	1		
8	3836	3736	3450	3017	2493	1940	1415	961	602	343
9	21450	24584	27073	28675	29228	28675	27073	24584	21450	17957
10	86054	110010	135853	162337	187959	211089	230131	243694	250749	250749

3. Asymptotic Normality The unimodal behavior of the inversion numbers suggests that the number of inversions in a random permutation may be asymptotically normal. We explore this possibility by looking at the generating function for the probability distribution of the number of inversions. To get this generating function, we divide $\Phi_n(x)$ by $n!$ since each of the $n!$ permutations is equally likely.

$$\phi_n(x) = \Phi_n(x)/n!.$$

Following Vladimir Sachkov, we have the moment generating function [12]

$$\begin{aligned}
M_n(x) &= \phi_n(e^x) \\
&= \prod_{j=1}^{n} \frac{1-e^{jx}}{j(1-e^x)} \\
&= \exp\left\{\frac{1}{2}\sum_{j=0}^{n-1} jx\right\} \prod_{j=1}^{n} \frac{e^{-jx/2}-e^{jx/2}}{j(e^{-x/2}-e^{x/2})} \\
&= \exp\left\{\frac{1}{2}\sum_{j=0}^{n-1} jx\right\} \prod_{j=1}^{n} \frac{e^{jx/2}-e^{-jx/2}}{j(e^{x/2}-e^{-x/2})} \\
&= \exp\left\{\frac{n(n-1)x}{4}\right\} \prod_{j=1}^{n} \frac{\sinh(xj/2)}{j\sinh(x/2)}
\end{aligned}$$

An explicit formula for the generating function of the Bernoulli numbers is

$$\frac{x}{e^x-1} = \sum_{k=0}^{\infty} B_k \frac{x^k}{k!}.$$

Hence

$$\begin{aligned}
\frac{x}{e^x-1} + \frac{x}{1-e^{-x}} &= \sum_{k=0}^{\infty} B_k \frac{x^k}{k!} + \sum_{k=0}^{\infty} B_k \frac{(-x)^k}{k!} \\
\frac{xe^{-x/2}}{e^{x/2}-e^{-x/2}} + \frac{xe^{x/2}}{e^{x/2}-e^{-x/2}} &= 2\sum_{k=0}^{\infty} B_{2k} \frac{x^{2k}}{(2k)!}
\end{aligned}$$

$$\begin{aligned}
\frac{e^{-x/2}+e^{x/2}}{e^{x/2}-e^{-x/2}} &= 2\sum_{k=0}^{\infty} B_{2k}\frac{x^{2k-1}}{(2k)!} \\
\frac{e^{-x/2}+e^{x/2}}{2(e^{x/2}-e^{-x/2})} &= \frac{1}{x}+\sum_{k=1}^{\infty} B_{2k}\frac{x^{2k-1}}{(2k)!} \\
\frac{e^{-x/2}+e^{x/2}}{2(e^{x/2}-e^{-x/2})}-\frac{1}{x} &= \sum_{k=1}^{\infty} B_{2k}\frac{x^{2k-1}}{(2k)!} \\
\ln\left(\frac{\sinh(x/2)}{x/2}\right) &= \sum_{k=1}^{\infty} B_{2k}\frac{x^{2k}}{2k(2k)!},
\end{aligned}$$

where the final step follows from integrating both sides and noting that

$$\lim_{x\to 0}\frac{\sinh(x/2)}{x/2}=1,$$

so the constant of integration is zero.

Using this generating function, we find that the log of the moment generating function is

$$\begin{aligned}
\ln M_n(x) &= \frac{n(n-1)x}{4}+\sum_{j=1}^{n}\left(\ln\left(\frac{\sinh(xj/2)}{xj/2}\right)-\ln\left(\frac{\sinh(x/2)}{x/2}\right)\right) \\
&= \frac{n(n-1)x}{4}+\sum_{k=1}^{\infty} B_{2k}\frac{x^{2k}}{2k(2k)!}\sum_{j=1}^{n}(j^{2k}-1).
\end{aligned}$$

Now consider $\ln M_n(t/\sigma)$, where σ is the standard deviation of the number of inversions in a random equiprobable permutation with n elements,

$$\sigma = \sqrt{\frac{2n^3+3n^2-5n}{72}},$$

$$\ln M_n(t/\sigma) = \frac{n(n-1)t}{4\sigma}+\sum_{k=1}^{\infty} B_{2k}\frac{t^{2k}}{2k(2k)!\sigma^{2k}}\sum_{j=1}^{n}(j^{2k}-1).$$

The sum

$$\sigma^{-2k}\sum_{j=1}^{n}(j^{2k}-1),$$

for $k>1$ is bounded above by the following integral:

$$\sum_{j=1}^{n}(j^{2k}-1) < \int_1^{n+1}(t^{2k}-1)dt = \frac{(n+1)^{2k+1}-1}{2k+1}-n,$$

so

$$\sigma^{-2k}\sum_{j=1}^{n}(j^{2k}-1) = O(n^{1-k}).$$

Hence

$$\sum_{k=2}^{\infty} B_{2k} \frac{t^{2k}}{2k(2k)!\sigma^{2k}} \sum_{j=1}^{n} (j^{2k} - 1) \to 0, \quad \text{as} \quad n \to \infty \ ,$$

uniformly for t from any bounded set. Therefore

$$\begin{aligned} \lim_{n\to\infty} M_n(t/\sigma) \exp\left\{-\frac{n(n-1)t}{4\sigma}\right\} &= \lim_{n\to\infty} \exp\left\{\sum_{k=1}^{\infty} B_{2k} \frac{t^{2k}}{2k(2k)!\sigma^{2k}} \sum_{j=1}^{n} (j^{2k} - 1)\right\} \\ &= \lim_{n\to\infty} \exp\left\{B_2 \frac{t^2}{2(2)!\sigma^2} \sum_{j=1}^{n} (j^2 - 1)\right\} \\ &= e^{t^2/2}. \end{aligned}$$

This leads to the following theorem:

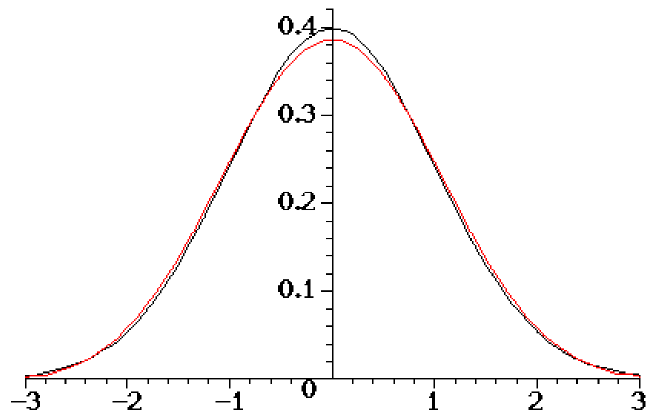
Theorem 2 (Sachkov). *[12] If ξ_n is a random variable representing the number of inversions in a random equiprobable permutation of n elements, then the random variable*

$$\eta_n \quad = \quad (\xi_n - E\xi_n)(\text{Var}\xi_n)^{-1}$$

has as $n \to \infty$ an asymptotically normal distribution with parameters $(0, 1)$.

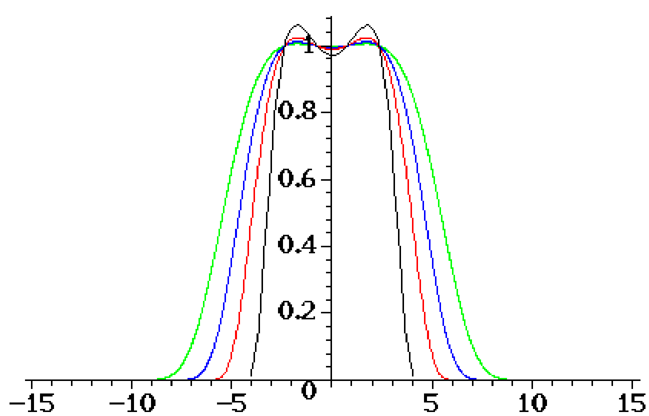
The graph below shows the density for a standard normal random variable in black. The red curve gives a continuous approximation for the discrete probability mass function for the number of inversions of a random permutation with n elements. The graph shown is for $n = 10$. As n increases, the red curve moves closer to the standard normal density so that it appears that the normal density may serve as a useful tool for approximating the inversion numbers.

Figure 1. Comparison of the inversion probability mass function to the standard normal density



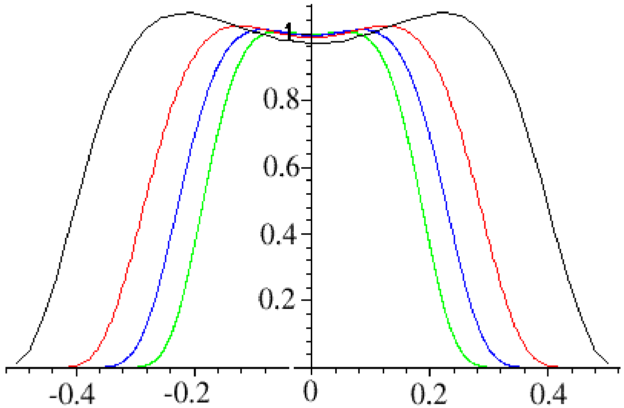
The figure below shows the ratio of the inversion numbers to the estimate provided by the normal density. The better the approximation, the closer the curve will be to 1. The graph is scaled so that the $x-$axis is the number of standard deviations from the mean.

Figure 2. The ratio of the inversion probability mass function to the standard normal density scaled by the number of standard deviations from the mean



The curves have roughly the shape of a cowboy hat. The top of the hat at about $y = 1$ seems to be getting broader as n increases (black is $n = 10$, red is $n = 25$, blue is $n = 50$, and green is $n = 100$), suggesting that the approximation improves with increasing n. Compare the figure above to the one below:

Figure 3. The ratio of the inversion probability mass function to the standard normal density scaled by the nonzero inversion numbers



The curves are rescaled in this figure so that 0 inversions is mapped to -0.5, and $\binom{n}{2}$ inversions is mapped to 0.5 on the $x-$axis. In this way, we can see whether the estimates for the nonzero inversion numbers improve as a percentage of the total nonzero inversion numbers as n increases. Note that the colored curves are in the opposite order of the preceding figure. The figure suggests that the estimates actually get worse as n increases. The width of the top of the cowboy hat is getting narrower as n increases. What this shows is that the relative error of the normal density approximation increases as n increases as we move further into the tails of the distribution. We can examine the asymptotic behavior of $I_n(k)$ for $k \leq n$ more closely.

4. An explicit formula for the inversion numbers Donald Knuth has made the observation that we may write an explicit formula for the kth coefficient of the generating function when $k \leq n$ ([8], p. 16). In that case,

Theorem 3 (Knuth, Netto). *[8],[11] The inversion numbers $I_n(k)$ satisfy the formula*

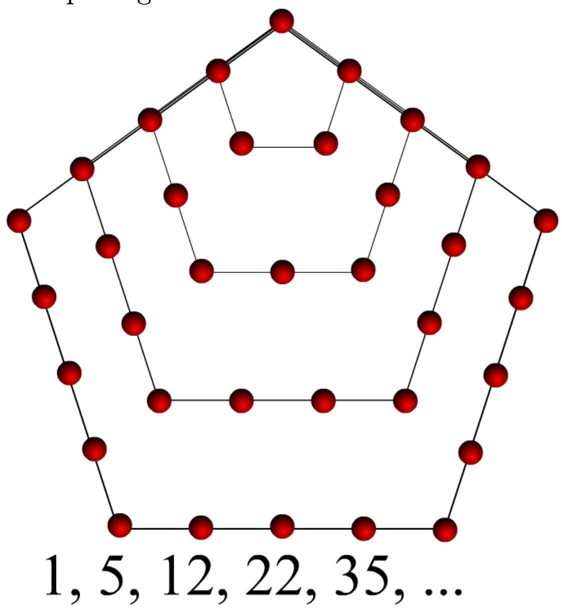
$$I_n(k) = \binom{n+k-1}{k} + \sum_{j=1}^{\infty}(-1)^j \binom{n+k-u_j-j-1}{k-u_j-j} + \sum_{j=1}^{\infty}(-1)^j \binom{n+k-u_j-1}{k-u_j}, \quad (1)$$

for $k \leq n$.

The binomial coefficients are defined to be zero when the lower index is negative, so there are only finitely many nonzero terms: $\lfloor -1/6 + \sqrt{1/36 + 2k/3} \rfloor$ in the first sum, and $\lfloor 1/6 + \sqrt{1/36 + 2k/3} \rfloor$ in the second. The u_j are the pentagonal numbers (sequence A001318 in [13]),

$$u_j = \frac{j(3j-1)}{2}.$$

Figure 4. The pentagonal numbers



Donald Knuth's formula follows from the generating function and Euler's pentagonal number theorem.

Theorem 4 (Euler). *[1][7][8]*

$$\prod_{j=1}^{\infty}(1-x^j) = 1 + \sum_{k=1}^{\infty}(-1)^k(x^{k(3k-1)/2} + x^{k(3k+1)/2}).$$

Recall the generating function

$$\begin{aligned} \Phi_n(x) &= \prod_{j=1}^{n} \frac{1-x^j}{1-x} \\ &= \left(\prod_{j=1}^{n}(1-x^j)\right)(1-x)^{-n} \\ &= \left(\prod_{j=1}^{n}(1-x^j)\right)\sum_{m=0}^{\infty}\binom{m+n-1}{m}x^m, \quad \text{for } |x|<1. \end{aligned}$$

The coefficients of $\prod_{j=1}^{n}(1-x^j)$ will match those in the power series expansion of the infinite product given by Euler's pentagonal number theorem up to the coefficient on x^n. We consider the product

$$\left(\prod_{j=1}^{\infty}(1-x^j)\right)\sum_{m=0}^{\infty}\binom{m+n-1}{m}x^m =$$
$$\left(1+\sum_{i=1}^{\infty}(-1)^i(x^{i(3i-1)/2}+x^{i(3i+1)/2})\right)\sum_{m=0}^{\infty}\binom{m+n-1}{m}x^m.$$

The coefficient on x^k is given by (1), for $k \le n$.

5. An asymptotic formula for the inversion numbers We are interested in the sequences $\{I_{n+k}(n),\ n=1,2,\ldots\}$. For $k \ge 0$, the nth term of the sequence is given by

$$\begin{aligned} I_{n+k}(n) &= \binom{2n+k-1}{n} + \sum_{j=1}^{\lfloor -1/6+\sqrt{1/36+2n/3}\rfloor}(-1)^j\binom{2n+k-u_j-j-1}{n-u_j-j} \\ &\quad + \sum_{j=1}^{\lfloor 1/6+\sqrt{1/36+2n/3}\rfloor}(-1)^j\binom{2n+k-u_j-1}{n-u_j} \end{aligned} \tag{2}$$

With $a=u_j+j$ or $a=u_j$, all terms are of the form

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

We can approximate this quantity using Stirling's approximation ([4], p.54 or [6], p.452):

$$n! = \sqrt{2\pi}n^{n+1/2}e^{-n}(1+(12n)^{-1}+O(n^{-2})).$$

So we have

$$\begin{aligned} &\binom{2n+k-a-1}{n-a} \\ &= \left(\frac{2n+k-a-1}{n-a}\right)^{n-a}\left(\frac{2n+k-a-1}{n+k-1}\right)^{n+k-1}\left(\frac{2n+k-a-1}{2\pi(n+k-1)(n-a)}\right)^{1/2}\times \\ &\quad\times\left(1-(8n)^{-1}+O(n^{-2})\right) \\ &= \frac{2^{2n+k-1-a}}{\sqrt{\pi n}}\left(1+\frac{(a+k-1)^2}{4(n-a)(n+k-1)}\right)^n\left(1-\frac{k+a-1}{2(n+k-1)}\right)^{k-1}\times \\ &\quad\times\left(1-\frac{n+k-1}{2n+k-a-1}\right)^a\left(\frac{1}{1-a/n}\left(\frac{k+a-1}{2(n+k-1)}\right)\right)^{1/2}\left(1-(8n)^{-1}+O(n^{-2})\right) \\ &= \frac{2^{2n+k-1-a}}{\sqrt{\pi n}}\left(1+\frac{n(a+k-1)^2}{4(n-a)(n+k-1)}\right)\left(1-\frac{(k-1)(k+a-1)}{2(n+k-1)}\right)\times \end{aligned}$$

$$\times\left(1-\frac{a(n+k-1)}{2n+k-a-1)}\right)\left(1+\frac{a-k+1}{4n}\right)\left(1-(8n)^{-1}+O(n^{-2})\right)$$
$$=\frac{2^{2n+k-1-a}}{\sqrt{\pi n}}\left(1-\frac{1}{8n}+\frac{1}{4n}(k+3a-(k+a)^2)+O(n^{-2})\right).$$

Using this asymptotic formula we can compute an asymptotic formula for the sum $I_{n+k}(n)$ given in equation (2):

$$I_{n+k}(n) = \frac{2^{2n+k-1}}{\sqrt{\pi n}}Q\left(1-\frac{C_1}{n}+\frac{C_2k-k^2}{4n}+O(n^{-2})\right)$$

where

$$\begin{aligned} Q &= \prod_{j=1}^{\infty}\left(1-\frac{1}{2^j}\right) \\ &= \sum_{i=1}^{\infty}(-1)^i\left(2^{-i(3i-1)/2}+2^{-i(3i+1)/2}\right) \\ &\approx 0.2887880951 \end{aligned}$$

is a digital search tree constant [5], and C_1 and C_2 are given by the convergent sums

$$\begin{aligned} C_1 &= \frac{1}{8}-\frac{1}{4Q}\sum_{i=1}^{\infty}(-1)^i\Big(2^{-i(3i-1)/2}(3(i(3i-1)/2)-(i(3i-1)/2)^2) \\ &\quad +2^{-i(3i+1)/2}(3(i(3i+1)/2)-(i(3i+1)/2)^2)\Big) \\ &\approx 1.855938894, \end{aligned}$$

and

$$\begin{aligned} C_2 &= 1+\frac{1}{Q}\sum_{i=1}^{\infty}(-1)^i(2^{-i(3i-1)/2}(i(3i-1))+2^{-i(3i+1)/2}(i(3i+1))) \\ &\approx 6.488067775, \end{aligned}$$

respectively. We summarize a less precise result in the following theorem:

Theorem 5.

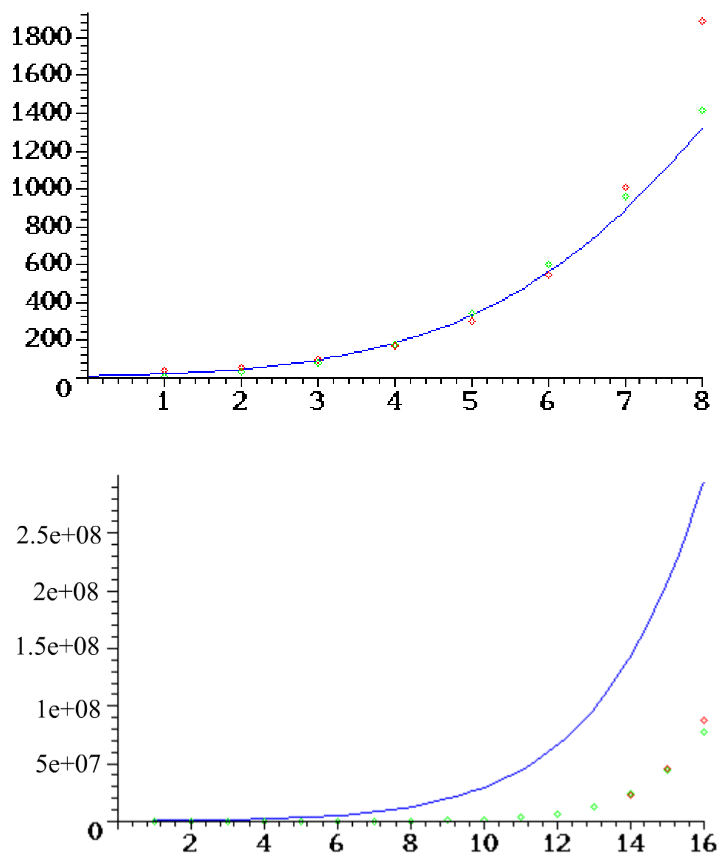
$$I_{n+k}(n) = \frac{2^{2n+k-1}}{\sqrt{\pi n}}Q\left(1+O(n^{-1})\right), \quad k \geq 0,$$

where $Q=\prod_{j=1}^{\infty}\left(1-\frac{1}{2^j}\right)$.

This formula provides asymptotic estimates for the sequences A000707, A001892, A001893, A001894, A005283, A005284 and A005285 of [13].

The figure below shows the behavior of the tail of the number of permutations with k inversions for $k \leq n$. The blue curve is $n!$ times normal density with mean $n(n-1)/4$ and variance $\frac{2n^3+3n^2-5n}{72}$, that is, the blue curve is the estimate of $I_n(k)$ based on the normal density. The red dots are the values of the asymptotic estimate; and the green dots are the exact values of $I_n(k)$. Where the red and green dots are not both visible, one dot covers the other. The figure shows the tail for $n = 8$ and $n = 16$.

Figure 4. Comparison of normal density estimate to asymptotic formula and actual inversion numbers



From our asymptotic formula for $I_n(n)$ we can see that

$$\lim_{n\to\infty} \frac{I_n(n)}{I_{n-1}(n-1)} = 4,$$

but the normal density approximation for the ratio $\frac{I_n(n)}{I_{n-1}(n-1)}$ gives the estimate $ne^{-9/8}$ as n tends to infinity. Hence the normal density approximation grows much faster than the inversion numbers in the tails do.

References

[1] G. E. Andrews, **The Theory of Partitions**, Cambridge University Press, 1998.

[2] L. Comtet, **Advanced Combinatorics**, Reidel, 1974, p. 240.

[3] F. N. David, M. G. Kendall and D. E. Barton, **Symmetric Function and Allied Tables**, Cambridge, 1966, p. 241.

[4] W. Feller, **An Introduction to Probability Theory and Its Applications**, second edition, John Wiley and Sons, New York, NY, 1971.

[5] S. Finch, **Digital Search Tree Constants**, published electronically at http://pauillac.inria.fr/algo/bsolve/constant/dig/dig.html.

[6] R. L. Graham, D. E. Knuth and O. Patashnik, **Concrete Mathematics**, 2d Ed., Addison-Wesley Publishing Company, Inc., Reading, MA, 1994.

[7] G. H. Hardy and E. M. Wright, **An Introduction to the Theory of Numbers**, Oxford, Clarendon Press, 1954.

[8] D. E. Knuth, **The Art of Computer Programming**. Addison-Wesley, Reading, MA, Vol. 3, p. 15.

[9] R. H. Moritz and R. C. Williams, "A coin-tossing problem and some related combinatorics", Math. Mag., 61 (1988), 24-29.

[10] Muir, "On a simple term of a determinant," *Proc. Royal S. Edinborough*, 21 (1898-9), 441-477.

[11] E. Netto, **Lehrbuch der Combinatorik.** 2nd ed., Teubner, Leipzig, 1927, p. 96.

[12] V. N. Sachkov, **Probabilistic Methods in Combinatorial Analysis**, Cambridge University Press, New York, NY, 1997.

[13] N. J. A. Sloane, **The On-Line Encyclopedia of Integer Sequences**, published electronically at http://www.research.att.com/ njas/sequences/, 2001.

(Concerned with sequence A000707, A001318, A001892, A001893, A001894, A005283, A005284, A005285, A008302.)

Received May 30, 2001; revised version received July 9, 2001. Published in Journal of Integer Sequences, Nov. 8, 2001.

Return to Journal of Integer Sequences home page.
