THE ACCURACY OF THE NORMAL APPROXIMATION TO THE DISTRIBUTION OF THE TRACES OF POWERS OF RANDOM ORTHOGONAL MATRICES

BY

CHARLES STEIN

TECHNICAL REPORT NO. 470

MARCH 1995

PREPARED UNDER THE AUSPICES OF

NATIONAL SCIENCE FOUNDATION GRANT DMS92-04864

DEPARTMENT OF STATISTICS

STANFORD UNIVERSITY

STANFORD, CALIFORNIA
The Accuracy of the Normal Approximation to the
Distribution of the Traces of Powers of Random
Orthogonal Matrices

By
Charles Stein

Technical Report No. 470
March 1995

Prepared Under the Auspices
Of
National Science Foundation Grant DMS92-04864

Department of Statistics
Stanford University
Stanford, California
The Accuracy of the Normal Approximation to the Distribution of the Traces of Powers of Random Orthogonal Matrices

by

Charles Stein
Stanford University

Abstract

The error in the normal approximation to the distribution of the trace of a power of a random orthogonal matrix in $n$ dimensions decreases faster than any power of $n$. More precisely, let $G_n$ be a random $n$-dimensional real orthogonal matrix distributed according to the invariant probability measure in the group of such matrices. Then, for all natural numbers $k$ and $r$, there exist constants $C_{k,r}$ such that the distance, in total variation, between the conditional distribution of $\text{tr} \ G_n^k$ given its determinant and the normal distribution, with variance $k$ and mean 0 or 1 depending on whether $k$ is odd or even, is less than $C_{k,r}(n - 1)^{-r}$. 
The Accuracy of the Normal Approximation to the Distribution of the Traces of Powers of Random Orthogonal Matrices

by
Charles Stein
Stanford University

1. Introduction

For each natural number \( n \), let \( G_n \) be a random orthogonal matrix, distributed according to the invariant probability measure in the orthogonal group. I shall prove that, for all \( n \) and every pair of natural numbers \( k \) and \( r \), the error in total variation of the normal approximation, with mean \( [1 + (-1)^k]/2 \) and variance \( k \), to the distribution of \( \operatorname{tr} G_n^k \) is less than \( C_{k,r}n^{-r} \) for appropriately chosen constants \( C_{k,r} \). In particular, \( C_{1,2} \) can be taken to be 15. In fact these results hold conditionally given the determinant of \( G_n \).

The work uses the methods of my book, Stein (1986), but the results needed from that work are summarized in Section 2. A basic identity concerning the joint distribution of the \( \operatorname{tr} G_n^k \) is derived in Section 3. This identity implies that all sufficiently regular functions of the \( \operatorname{tr} G_n^k \) that would have expectation 0 if the \( \operatorname{tr} G_n^k \) were independently normally distributed with means and variances described above do have expectation of the order of \( 1/n \). More precisely these expectations are evaluated as \( 1/(n - 1) \) times the expectation of certain related random variables. In Section 4 these related random variables are shown to be linear combinations of random variables occurring on the left-hand side of such identities. This permits an inductive evaluation of the expectations on the left-hand side as \( (n - 1)^{-r} \) times the expectations of certain related random variables. In Section 5 this leads to the theorem announced above. Presumably a similar result could be proved for the joint distribution of the \( \operatorname{tr} G_n^k \) with the aid of the work of Götze (1991), but some technical difficulties remain. In Section 6 I look at some special cases of the basic identity of Section 3, partly to convince myself that these rather tedious calculations are correct.

I am indebted to Persi Diaconis for suggesting the problem and to Susan Holmes for finding a serious blunder in the first version of this work. The results complement the work of Diaconis and Shahshahani (1994), who showed that the joint moments of weight not exceeding \( 2n \) of the \( \operatorname{tr} G_n^k \) are exactly equal to those of the corresponding normal distribution. A better understanding of the relation between their results and mine would be useful. Diaconis (personal communication) has also mentioned that Johansson has recently proved that the error in the normal approximation to the distribution of the trace of a uniformly distributed random \( n \) by \( n \) unitary matrix goes down exponentially with \( n \).
2. **Background material on normal approximation**

This section contains background material on the method of studying normal approximation introduced in Stein (1972) and studied more thoroughly in Stein (1986). The present treatment is essentially self-contained. Later sections use the notation and results of this section.

Let \( Z_1 \) be a real standard normal random variable and \( f \) a function on \( \mathcal{R} \) to \( \mathcal{R} \) that is an indefinite integral of a function \( f' \) for which \( E|f'(Z_1)| < \infty \). The linear space of such functions will be denoted by \( \mathcal{F}_1 \), and the linear space of all functions \( h \) for which \( E|h(Z_1)| < \infty \) will be denoted by \( \mathcal{H}_1 \). Then, using integration by parts or Fubini's theorem, it is not difficult to show that

\[
(2.1) \quad Ef'(Z_1) = EZ_1 f(Z_1).
\]

This suggests the following approach to the problem of obtaining a bound for the error of the standard normal approximation to \( Eh(W_1) \), where \( W_1 \) is a given random variable and \( h \) belongs to \( \mathcal{H}_1 \). Suppose that, for all \( f \) in \( \mathcal{F}_1 \), we can derive an identity of the form

\[
(2.2) \quad E[f'(W) - Wf(W)] = E(R_1 f)(W),
\]

where \( R_1 \) is a linear mapping on \( \mathcal{F}_1 \) to \( \mathcal{H}_1 \). Typically \( R_1 f \) will in some sense be small if the normal approximation is valid and we have made a good choice of \( R_1 \). Substitute for \( f \) in (2.2) the unique solution in \( \mathcal{F}_1 \) of the differential equation

\[
(2.3) \quad f'(w) - wf(w) = h(w) - Eh(Z),
\]

obtaining

\[
(2.4) \quad Eh(W) - Eh(Z) = E[f'(W) - Wf(W)] = E(R_1 f)(W).
\]

Then it remains only to bound the right-hand side.

In order to develop the details of this argument, it will be useful to introduce some linear mappings. Define \( E_1 \) on \( \mathcal{H}_1 \) to \( \mathcal{R} \) by

\[
(2.5) \quad E_1 h = Eh(Z_1),
\]

define \( T_1 \) on \( \mathcal{F}_1 \) to \( \mathcal{H}_1 \) by

\[
(2.6) \quad (T_1 f)(w) = f'(w) - wf(w),
\]
$U_1$ on $\mathcal{H}_1$ to $\mathcal{F}_1$ by

\begin{equation}
(U_1 h)(w) = e^{w^2/2} \int \{x < w\} [h(x) - E_1 h] e^{-x^2/2} dx,
\end{equation}

and $V_1$ on $\mathcal{H}_1$ to $\mathcal{H}_1$ by

\begin{equation}
(V_1 h)(w) = h(w) - E_1 h - we^{w^2/2} \int \{x < w\} [h(x) - E_1 h] e^{-x^2/2} dx.
\end{equation}

Then $f = U_1 h$ is the unique solution in $\mathcal{F}_1$ of the first order linear differential equation $T_1 f = h$ and $V_1 h$ is the derivative of that solution. From the well-known fact that, for $w > 0$,

\begin{equation}
1 - \Phi(w) < \frac{e^{-w^2/2}}{w\sqrt{2\pi}},
\end{equation}

where $\Phi$ is the standard cumulative normal distribution function, it follows that

\begin{align}
\sup |U_1 h| &< \sqrt{\frac{\pi}{2}} \sup |f - E_1 h| \\
\sup |V_1 h| &< \sup |h - E_1 h|.
\end{align}

This will be used in Section 5, with some change of notation, to prove a normal approximation theorem for a single tr $G_n^1$.

We shall need some extensions of this notation. Let $Z$ be a random function on $\{1, 2, \ldots\}$ having the $Z(k)$ independently normally distributed with means $\alpha(k) = \frac{1+(-1)^k}{2}$ and variances $k$. Let $\mathcal{H}$ be the linear space of all real-valued functions $h$ on $\mathcal{R}^{\{1, 2, \ldots\}}$ that are determined by a finite number of coordinates of their argument and satisfy

\begin{equation}
E|h(Z)| < \infty.
\end{equation}

For each natural number $k$, let $\mathcal{F}_k$ be the linear space of all real-valued functions $f$ on $\mathcal{R}^{\{1, 2, \ldots\}}$ for which $D_k f$ belongs to $\mathcal{H}$, where $D_k$ denotes partial differentiation with respect to the $k$-th coordinate, and define a linear mapping $T_k$ on $\mathcal{F}_k$ to $\mathcal{H}$ by

\begin{equation}
(T_k f)(w) = k(D_k f)(w) - w_k^* f(w),
\end{equation}

where $w_k = w(k)$ and $w_k^* = w_k - \alpha(k)$ and similar notation will be used for $W$. It follows from (2.1) that

\begin{equation}
E(T_k f)(Z) = 0.
\end{equation}
For each $k$, define the linear mapping $U_k$ on $\mathcal{H}$ to $\mathcal{F}_k$ by

\begin{equation}
U_k f(w) = e^{\frac{w_k^2}{2k}} \int_{-\infty}^{w_k^*} e^{-\frac{w_k^2}{2k}} (h(w) - E_k h) dw_k.
\end{equation}

I hope the poor notation causes no serious difficulty. The $w$, $w_k$ and the $w_k^*$ in $e^{-\frac{w_k^2}{2k}}$ refer to the variable of integration. The other two occurrences of $w_k^*$ on the r.h.s. of (2.15) refer to the argument $w$ on the l.h.s. As in (2.1),

\begin{equation}
T_k U_k h = h - E_k h.
\end{equation}

I shall ordinarily omit the parentheses in fragments such as $(T_k f)$ and $(D_k f)$ and I may also write things such as $T_k w_j f(w)$, which are to be interpreted in the obvious way.

3. A basic identity

As indicated in the introduction, we are interested in the joint distribution of the traces of the powers of a random orthogonal matrix $G$ in $n$ dimensions, uniformly distributed over a component of the orthogonal group. The notation used here is that of Section 2.

**THEOREM 3.1.** Let $G$ be a random $n$ by $n$ matrix uniformly distributed (according to the invariant probability measure) in a component of the orthogonal group, let $W_j = \text{tr} G^j$ for all $j$ in $\mathbb{Z}^+$ and let $W$ be the random element of $\mathcal{R}^N$ having these coordinates. Let $f$ be a function on $\mathcal{R}^N$ to $\mathcal{R}$ that is determined by the first $k_0$ coordinates of its argument and has the property that for every $k \leq k_0$, $f$ is an indefinite integral, with respect to the $k$-th coordinate of its argument, of a function that will be denoted by $D_k f$. Then, for all natural numbers $k \leq k_0$,

\begin{equation}
ET_k f(W) = \frac{1}{n-1} \left( k(W_{2k} - 1)D_k + \sum_{i \leq k_0 \& i \neq k} (W_{i+k} - W_{|k-i|})D_i + \alpha(k)((W_{k/2})^2 - 1) + 2\sum_{a \leq k/2} (W_a W_{k-a} - W_{k-2a}) \right) f(W).
\end{equation}

**REMARKS.** (a) If $Z$ is a normally distributed random vector with independent coordinates and with $EZ_j = \alpha(j)$ and $\text{Var} Z_j = j$ for all $j$ then, by the results of Section 2, $ET_k f(W) = 0$ for all $k$, which strongly suggests that the error in the normal approximation to the distribution of the restriction of $W$ to $\{1, \ldots, k_0\}$ is $O(1/n)$, with the implied constant depending on $k_0$. In Section 5, we shall see that the error of the normal approximation to the distribution of any single $W_k$ goes down faster than any power of $n$. It is highly plausible, but not yet proved, that the same holds for the restriction of $W$ to $\{1, \ldots, k\}$. 

4
(b) In applications we shall often have to use this result for a number of different values of \( k \), with a different choice of \( f \) for each \( k \).

(c) In the proof below I shall write, for example, \( O(\varepsilon^3) \) for a quantity that is bounded in absolute value by \( \varepsilon^3 \) times an expression that may depend on various quantities. It is easy to verify that the validity of the limiting results obtained is not affected by this dependence.

**PROOF.** In the following argument I shall assume \( f \) to be twice continuously differentiable. The theorem follows as stated by a simple approximation argument using Lebesgue’s bounded convergence theorem. For \( \varepsilon > 0 \), let \( G_\varepsilon \) be the matrix obtained by multiplying \( G \) on the left by a random rotation through an angle \( \pm \sin^{-1} \varepsilon \) so that \((G, G_\varepsilon)\) is an exchangeable pair. More precisely, with \( U \) uniformly distributed in the two point set \([-1, 1]\) independent of \( G \), let

\[ G_\varepsilon = H(I + A_\varepsilon)H^t G, \]

where \( H \) is a random \( n \) by \( n \) orthogonal matrix uniformly distributed over the orthogonal group independent of \( G \) and \( U \), and \( A_\varepsilon \) is the \( n \) by \( n \) matrix having as its upper left-hand corner the 2 by 2 matrix

\[ B_\varepsilon = \begin{pmatrix} (1 - \varepsilon^2)^{-1/2} \varepsilon U & \varepsilon U \\ -\varepsilon U & \sqrt{(1 - \varepsilon^2)^{-1} - 1} \end{pmatrix} = \begin{pmatrix} -\varepsilon^2/2 & \varepsilon U \\ -\varepsilon U & -\varepsilon^2/2 \end{pmatrix} + O(\varepsilon^3) \]

\[ = -\varepsilon^2/2 I + \varepsilon c + O(\varepsilon^3), \]

and zeros elsewhere, where \( e_2 \) is the 2 by 2 identity matrix and

\[ c = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}. \]

Also let \( K \) be the random \( n \) by 2 matrix consisting of the first two columns of \( H \), so that \( HA_\varepsilon K^t = KB_\varepsilon K \), let \( W_{\varepsilon,j} = \text{tr} \ G_j \) for all \( j \), and let \( s \) be an upper bound for the second partial derivatives of \( f \). Then for all \( k \),

\[ 0 = E(W_{\varepsilon,k} - W_k)[f(W) + f(W_\varepsilon)] \]

\[ = E(W_{\varepsilon,k} - W_k)[2f(W) + f(W_\varepsilon) - f(W))] \]

\[ = E[2(E^G(W_{\varepsilon,k} - W_k))f(W) \]

\[ + (W_{\varepsilon,k} - W_k)\Sigma_m \{ m \leq k_0 \}(W_{\varepsilon,m} - W_m)D_m f(W) \]

\[ + O(s\Sigma_m \{ m \leq k_0 \}|W_{\varepsilon,m} - W_m|^3)]. \]

5
Multiplying by $n/(e^2 k)$ and letting $\epsilon$ approach 0,

$$
E \lim_{\epsilon \to 0} \left[ 2nE^G(W_{\epsilon,k} - W_k)/(e^2 k) \right] f(W) \\
+ \sum_m \lim_{\epsilon \to 0} \left[ nE^G(W_{\epsilon,k} - W_k)(W_{\epsilon,m} - W_m)/(e^2 k) \right] D_m f(W).
$$

The following well-known lemma will be used in evaluating the coefficients in equation (3.6).

**Lemma 1.** Let $k$ be a natural number and let $x$ and $y$ be $n$ by $n$ matrices. Then, as $\epsilon$ approaches 0,

$$
(x + \epsilon y)^k = x^k + \epsilon \sum_{i=0}^{k-1} x^i y x^{k-i-1} \\
+ \epsilon^2 \sum_{i,j=0}^k \{0 \leq i \leq k-2, 0 \leq j \leq k-2\} x^i y x^j y x^{k-2-i-j} \\
+ O(\epsilon^3).
$$

This can be proved by induction on $k$. Of course, there are $k$ terms in the first sum and $k(k-1)/2$ terms in the second sum, in agreement with the commutative case.

Taking the trace of (3.7) and collecting terms appropriately,

$$
\text{tr} \ (x + \epsilon y)^k \\
= \text{tr} \ x^k + k\epsilon \text{tr} \ x^{k-1} y \\
+ \epsilon^2 \sum_{j=0}^{k-2} (k-2) \{k-1-(k-2)/2\} \text{tr} \ y x^{(k-2)/2} \\
+ \epsilon^2 \sum_{j=0}^{k-2} (k-2) \{k-1-(k-2)/2\} \text{tr} \ y x^{j} y x^{k-2-j} \\
+ O(\epsilon^3)
$$

In order to evaluate the coefficients in (3.6), replace $x$ by $G$ and replace $y$ by the expression $K((-\epsilon^2/2)e_2 + Uc)K'G$ in (3.8), obtaining

$$
W_{\epsilon,k} - W_k = \text{tr} \ (G_{\epsilon}^k - G^k) \\
= \text{tr} \ \left[ (G + K((-\epsilon^2/2)e_2 + \epsilon Uc)K'G)^k - G^k \right] + O(\epsilon^3) \\
= k\epsilon U \text{tr} \ KcK'G^k - (ke^2/2) \text{tr} \ KK'G^k \\
+ (k\epsilon^2/2) [\alpha(k) \text{tr} \ (KcK'G^{k/2})^2 \\
+ 2\sum_a \{1 \leq a < k/2\} \text{tr} \ KcK'G^a KcK'G^{k-a}] + O(\epsilon^3).
$$
Consequently

\[
\lim_{\epsilon \to 0} \frac{n}{(e^2 k)} E^G(W_{\epsilon,k} - W_k) \\
= (n/2)E^G [2\Sigma_a \{1 \leq a < k/2\} \text{ tr } KcK'G^aKcK'G^{k-a} \\
+ \alpha(k) \text{ tr } (KcK'G^{k/2})^2] - (n/2) \text{ tr } (E^G K K')G^k,
\]

and

\[
\lim_{\epsilon \to 0} \frac{n}{(e^2 k)} E^G(W_{\epsilon,k} - W_k)(W_{\epsilon,i} - W_i) \\
= niE^G(\text{tr } KcK'G^k) \text{ tr } KcK'G^i.
\]

In order to evaluate the conditional expectations on the right-hand side of (3.10) and (3.11) we shall need two lemmas. The need for Lemma 2 (which is well known) will be clear from the statement of Lemma 3.

**LEMMA 2.** If $H$ is uniformly distributed over the real orthogonal group in $n$ dimensions, then

\[
\gamma_n = E(H_{1,1}^2, H_{2,2}^2 - H_{1,1}H_{1,2}H_{2,1}H_{2,2}) = \frac{1}{n(n-1)}.
\]

**PROOF.** Using elementary properties of a random orthogonal matrix,

\[
EH_{1,1}^2 H_{1,2}^2 = \frac{1}{n-1} EH_{1,1}^2(1 - H_{1,1}^2) \\
= \frac{1}{n-1} \left[ \frac{1}{n} - \frac{3}{n(n+2)} \right] \\
= \frac{1}{n(n+2)}
\]

\[
EH_{1,1}^2 H_{2,2}^2 = \frac{1}{n-1} EH_{1,1}^2(1 - H_{1,2}^2) \\
= \frac{n+1}{(n-1)n(n+2)}
\]

\[
EH_{1,1} H_{1,2} H_{2,1} H_{2,2} = \frac{1}{n(n-1)} E\Sigma_{i,j}\{i \neq j\} H_{i,1} H_{i,2} H_{j,1} H_{j,2} \\
= \frac{1}{n(n-1)} E[(\Sigma_i H_{i,1} H_{i,2})^2 - \Sigma_i H_{i,1}^2 H_{i,2}^2] \\
= -\frac{1}{n-1} EH_{1,1}^2 H_{1,2}^2 \\
= \frac{-1}{(n-1)n(n+2)}.
\]
Then (3.12) follows from (3.14) and (3.15).

**Lemma 3.** For all natural numbers \(a\) and \(b\),

\[
(3.16) \quad E^G \text{tr } KcK'G^aKcK'G^b = \frac{2}{n(n-1)} \left[ \text{tr } G^{a-b} - (\text{tr } G^a)(\text{tr } G^b) \right],
\]

and

\[
(3.17) \quad E^G(\text{tr } KcK'G^a)(\text{tr } KcK'G^b) = \frac{2}{n(n-1)} \left[ \text{tr } G^{a-b} - \text{tr } G^{a+b} \right].
\]

**Proof of (3.16):**

\[
(3.18) \quad E^G \text{tr } KcK'G^aKcK'G^b = \sum_{i,j} E^G(KcK'G^a)_{i,j}(KcK'G^b)_{j,i}
\]

\[
= \sum_{i,j,i',j'} E^G(KcK'_{i,i'}G^a_{i',j'}(KcK'_{j,j'})_{j',i})
\]

\[
= \sum_{i,j,i',j'} q_{i,i',j,j'}(G^a_{i',j'}G^b_{j',i})
\]

where, because \(K\) is independent of \(G\),

\[
(3.19) \quad q_{i,i',j,j'} = E^G(KcK'_{i,i'}G^a_{i',j'}(KcK'_{j,j'})_{j',i})
\]

\[
= E(K_{i,1}^2K_{i',2}^2 - K_{i',1}K_{i,2}^2) - K_{i,1}K_{i',2}K_{j,1}K_{j',2}
\]

\[
= E[K_{i,1}K_{i',1}K_{i',2}K_{j,2} + K_{i',1}K_{j',1}K_{i,2}K_{j,2}]
\]

But the multiplication of a fixed row or column of a uniformly distributed random orthogonal matrix by -1 does not destroy the uniform distribution. Consequently the mixed moments of elements of a random orthogonal matrix are 0 unless each index occurs an even number of times in the product of which we take the expectation. Also observe that if either \(i = i'\) or \(j = j'\), then \(q_{i,i',j,j'}\) vanishes. Thus non-zero expectations can occur only if \(i = j\) and \(i' = j'\), or \(i = j'\) and \(j = i'\), but not both. Carrying out the expectation on the right-hand side of (3.20), we obtain

\[
(3.20) \quad q_{i,i',j,j'} = E\left[ 2(K_{i1}^2K_{i2}^2\delta_{i,j}\delta_{i',j'} + K_{11}K_{12}K_{21}K_{22}\delta_{i,j}\delta_{i',j'})
\]

\[
- 2(K_{11}^2K_{12}^2\delta_{i,j}\delta_{i',j'} + K_{11}K_{12}K_{21}K_{22}\delta_{i,j}\delta_{i',j'}) \right]
\]

\[
= \frac{2}{n(n-1)} \left[ \delta_{i,j}\delta_{i',j'} - \delta_{i,j'}\delta_{i',j} \right].
\]

This is antisymmetric in \((i,i')\) and in \((j,j')\) as anticipated. Substituting in (3.18),
\[(3.21) \quad E^G \text{tr } KcK'G^aKcK'G^b \]
\[= \frac{2}{n(n-1)} \sum_{i,j,i',j'} [\delta_{i,j}\delta_{i',j'} - \delta_{i,j'}\delta_{i',j}] (G^a)_{i',j}(G^b)_{j',i} \]
\[= \frac{2}{n(n-1)} \left[ \sum_{i,i'} (G^a)_{i',i}(G^b)_{i',i} - \sum_{i,i'} (G^a)_{i',i'}(G^b)_{i,i} \right] \]
\[= \frac{2}{n(n-1)} \left[ \text{tr } G^a(G^b)' - \text{tr } G^a \text{tr } G^b \right] \]
\[= \frac{2}{n(n-1)} \left[ \text{tr } G^{a-b} - \text{tr } G^a \text{tr } G^b \right]. \]

**PROOF of (3.17):**

\[(3.22) \quad E^G(\text{tr } KcK'G^a)(\text{tr } KcK'G^b) \]
\[= \sum_{i,i',j,j'} (G^a)_{i',i}(G^b)_{j',j} E^G(KcK')_{i,i'}(KcK')_{j,j} \]
\[= \frac{2}{n(n-1)} \sum_{i,i'} [(G^a)_{i',i}(G^b)_{i',i} - (G^a)_{i',i}(G^b)_{i,i}] \]
\[= \frac{2}{n(n-1)} \left[ \text{tr } G^{a-b} - \text{tr } G^{a+b} \right]. \]

By (3.9) and Lemma 3 and the fact that \(K\) is independent of \(G\) and \(EKK' = \frac{2}{n} I,\)

\[(3.23) \quad \lim_{\varepsilon \to 0} \frac{n}{\varepsilon^2} E^G(W_{\varepsilon,k} - W_k) \]
\[= -\text{tr } G^k + \frac{1}{n-1} \left[ 2\Sigma_a \{1 \leq a < k/2 \} [\text{tr } G^{2a-k} - (\text{tr } G^a) \text{ tr } G^{k-a}] \right. \]
\[+ \alpha(k)[n - (\text{tr } G^{k/2})^2] \].

Also, again using (3.11) and (3.8),

\[(3.24) \quad \lim_{\varepsilon \to 0} \frac{n}{\varepsilon^2} E^G(W_{\varepsilon,k} - W_k)(W_{\varepsilon,m} - W_m) \]
\[= nm E^G(\text{tr } KcK'G^k) \text{ tr } KcK'G^m \]
\[= \frac{2m}{n-1} \sum_{i,j} \{ i \neq j \} [(G^k)_{i,j}(G^m)_{i,j} - (G^k)_{i,j}(G^m)_{j,i}] \]
\[= \frac{2m}{n-1} \left[ \text{tr } G^k(G^m)^t - \text{tr } G^{k+m} \right] \]
\[= \frac{2m}{n-1} \left[ \text{tr } G^{k-m} - \text{tr } G^{k+m} \right]. \]
Substituting (3.23) and (3.24) in (3.6) and dividing by 2,

\begin{equation}
(3.25) \quad 0 = E \left\{ \left[ -\text{tr} \, G^k + \frac{1}{n-1} [2 \Sigma_a \{1 \leq a < k/2\} \text{tr} \, G^{k-2a} - (\text{tr} \, G^a) \, \text{tr} \, G^{k-a}] \right. \\
+ \alpha(k)[n - (\text{tr} \, G^{k/2})^2] \right\} f(W) \\
+ \frac{1}{n-1} \Sigma_m \{1 \leq m \leq k_0\} m[\text{tr} \, G^{k-m} - \text{tr} \, G^{k+m}] D_m f(W) \right\}.
\end{equation}

which implies (3.1).

4. Reformulation of the basic identity

From the basic theorem of the last section it seems plausible that, if \( G_n \) is a random \( n \) by \( n \) orthogonal matrix distributed in a component of that group according to the invariant probability measure, then the joint distribution of the \( \text{tr} \, G^k \) for \( k \) in \( \{1, \ldots, k_0\} \) is approximately normal, with mean 0 for odd \( k \) and 1 for even \( k \) and variance \( k \), the error of approximation being \( O(n^{-1}) \). The main theorem of Section 5 shows that the error in the normal approximation to the distribution of a single \( \text{tr} \, G^k_n \) is \( O(n^{-r}) \) for arbitrary \( r \). The essential step in this proof is to rewrite the quantity under the expectation sign on the right-hand side of the basic identity of the last section as a linear combination of expressions having the same form as the left-hand side, thus permitting an inductive argument. Recall that the linear mapping \( T_k \) is defined by

\begin{equation}
(4.1) \quad T_k f(w) = kD_k f(w) - w_k^* f(w).
\end{equation}

**THEOREM 4.1.** For every \( f \) in \( C^1(R^{k_0}) \) and every \( k \leq k_0 \),

\begin{equation}
(4.2) \quad E^* T_k f(W) = \frac{1}{n-1} E^* \left[ \Sigma_m \{m \leq k_0\} T_m W_{k+m}^* - \Sigma_m \{m \leq k_0 + k \land m \neq k\} T_m W_{k+m}^* \right] f(W).
\end{equation}

**REMARKS.** Here the \( m \)-th term in the second summation is to be evaluated by applying the linear mapping \( T_m \) to the function that takes each \( w \) into \( w_{k-m}^* f(w) \), then substituting \( W \) for \( w \) and taking the conditional expectation of the resulting expression given the determinant of \( G \). The condition \( k \leq k_0 \) is needed only to make sure that \( T_k W_{2k}^* \) is included in the first summation in square brackets and can be achieved by replacing the original \( k_0 \) by the larger of \( k_0 \) and \( k \).

**PROOF.** It will be convenient to start from the right-hand side of (4.2) and show that it is equal to the r.h.s. of (3.1). I shall omit the \( \frac{1}{n-1}, E^* \), and \( f(W) \).
of $T_m$, separating the second sum at $k$ and rearranging terms,

\begin{equation}
\sum_{m \leq k_0} m W_{k+m} - \sum_{m \leq k_0 + k & m \neq k} T_m W_{k-[k-m]}^* \\
= \sum_{m \leq k_0} m W_{k+m} D_m - \sum_{m < k} m W_{k-m}^* D_m - \frac{k}{2} \alpha(k) \\
- \sum_{m < k} k m W_{k-m}^* D_m + \sum_{m \leq k_0} W_{m-k}^* W_{k+m}^* \\
- \sum_{m < k} W_{m}^* W_{k-m}^* - \sum_{m < k} k m W_{m-k}^* W_{m-k}^*.
\end{equation}

The fourth and sixth sums cancel. For the fifth term, observe that

\begin{equation}
\sum_{m < k} W_{m}^* W_{k-m}^* \\
= \sum_{m < k} [W_{m} W_{k-m} - \alpha(m) W_{k-m} - \alpha(k-m) W_{m} + \alpha(m) \alpha(k-m)] \\
= 2 \sum_{a < \frac{k}{2}} (W_{a} W_{k-a} - W_{k-2a}) + \alpha(k) (W_{\frac{k}{2}}^2 + \frac{k}{2} - 1).
\end{equation}

Because the sum in the third term on the r.h.s. of (4.3) can be restricted to $m \leq k_0$, the sum of the first three terms on the r.h.s. of (4.3) can be simplified to

\begin{equation}
\sum_{m \leq k_0 & m \neq k} m (W_{k+m} - W_{k-[k-m]}) D_m + k (W_{1k-1}) D_k - \frac{k}{2} \alpha(k).
\end{equation}

Combining this with (4.4),

\begin{equation}
\sum_{m \leq k_0} m W_{k+m}^* - \sum_{m \leq k_0 + k & m \neq k} T_m W_{k-[k-m]}^* \\
= \sum_{m \leq k_0 & m \neq k} m (W_{k+m} - W_{k-[k-m]}) D_m + k (W_{2k-1}) D_k \\
+ 2 \sum_{a < \frac{k}{2}} (W_{a} W_{k-a} - W_{k-2a}) + \alpha(k) (W_{\frac{k}{2}}^2 - 1).
\end{equation}

The theorem follows because the right-hand side of (3.1) is the expectation of the result of applying this to $f(W)$.

5. The accuracy of the normal approximation to the distribution of $\text{tr} \ G^k_n$

Theorem 4.1 will be applied inductively with the aid of Theorem 3.1 to prove that in Theorem 5.1, the error of the normal approximation to the distribution of $\text{tr} \ G^k_n$ decreases faster than any power of $n$. In Theorem 5.2, an explicit bound of order $n^{-2}$ is given for the special case $k = 1$.

THEOREM 5.1. For all natural numbers $k$ and $r$ there exist constants $C_{k,r}$ such that, for every measurable subset $A$ of the real line

\begin{equation}
\left| P\{\text{tr} \ G^k_n \in A\} - P\{Z_k \in A\} \right| \leq \frac{C_{k,r}}{(n-1)^2}.
\end{equation}
PROOF. It follows by induction from Theorem 4.1 that there exists a function \( \alpha \) on \( \mathbb{Z}^+ \) to \( \mathbb{Z}^+ \) and, for every \( i \leq \alpha(k,r) \), a polynomial \( \beta_{i,k,r}(w) \) of degree at most \( \alpha(k,r) \) in the first \( \alpha(k,r) \) coordinates of \( w \) such that, for every function \( f \) in \( F_k \)

\[
E^* T_k f(W_k) = \frac{1}{(n-1)^{r-1}} E^* \Sigma_i \{ i \leq \alpha(k,r) \} T_i [\beta_{i,k,r}(W)f(W_r)].
\]

By Theorem 3.1, this implies the existence of polynomials \( \gamma_{r,k,1} \) and \( \gamma_{r,k,2} \) of degree at most \( 2\alpha(k,r) \) in the first \( 2\alpha(k,r) \) coordinates of \( w \) such that

\[
E^* T_k f(W_k) = \frac{1}{(n-1)^{r}} E^* [\gamma_{r,k,1}(W)D_k f(W_k) + \gamma_{r,k,2}(W)f(W_k)].
\]

Substituting \( f = U_k h \) and using (2.14),

\[
E^* h(W_k) - Eh(Z_k) = E^* T_k (U_k h)(W_k)
= \frac{1}{(n-1)^{r}} E^* [\gamma_{r,k,1}(W)(U_k h)'(W_k) + \gamma_{r,k,2}(W)(U_k h)(W_k)].
\]

Consequently,

\[
|E^* h(W_k) - Eh(Z_k)|
\leq \frac{2}{(n-1)^{r}} [E |\gamma_{r,k,1}(W)| \sup |(U_k h)'| + \sup |\gamma_{r,k,2}(W)| \sup |U_k h|].
\]

Then (5.1) follows with

\[
C_{k,r} = 2E |\gamma_{r,k,1}(W)| + \sqrt{2\pi} E |\gamma_{r,k,2}(W)|.
\]

As a partial check on the validity of this argument, let us carry out the first two steps of the induction in the special case \( k = 1 \). Specializing (3.2) to the case \( k = 1 \) and to a function \( f \) determined by its first coordinate,

\[
E^* (T_1 f)(W_1) = \frac{1}{n-1} E^* (W_2 - 1) D_1 f(W_1).
\]

For an arbitrary bounded piecewise continuous function \( h \) on \( \mathcal{R} \) to \( \mathcal{R} \), substitute \( f = U_1 h \) in (5.7) to obtain

\[
E^* h(W_1) - Eh(Z_1) = E^* T_1 (U_1 h)(W_1)
= \frac{1}{n-1} E^* (W_2 - 1)(U_1 h)'(W_1).
\]
Taking \( h \) to be the indicator function of a measurable set \( A \) and using (5.9)

\[
|P^*\{\text{tr } G_n \in A\} - P\{Z_1 \in A\}| = \frac{1}{n-1} |E^* (W_2 - 1) (U_1 h)'(W_1)| \\
\leq \frac{1}{n-1} \sup |(U_1 h)'| E^* |W_2 - 1| \leq \frac{2}{n-1}.
\]

because

\[
E^* |W_2 - 1| \leq \sqrt{E^* (W_2 - 1)^2} \leq \sqrt{2E(W_2 - 1)^2} = 2.
\]

**THEOREM 5.2.** For every measurable subset \( A \) of the real line,

\[
|P^*\{\text{tr } G_n \in A\} - P\{Z_1 \in A\}| \leq \frac{15}{(n-1)^2}.
\]

**PROOF.** Specializing (4.2) to the case \( k = k_0 = 1 \),

\[
E^* T_1 f(W_1) = \frac{1}{n-1} E^* [T_1 W_2^* - T_2 W_1] f(W).
\]

By the special case \( k = 1, k_0 = 2 \) of (3.1), with \( f(W) \) replaced by \( W_2^* f(W), \)

\[
E^* T_1 W_2^* f(W) = \frac{1}{n-1} E^* [W_2^* D_1 + 2(W_3 - W_1)] f(W),
\]

and, by the special case \( k = 2, k_0 = 1 \) of (3.1), with \( f(W) \) replaced by \( W_1 f(W_1), \)

\[
E^* T_2 W_1 f(W_1) = \frac{1}{n-1} E^* (W_1^2 - 1) W_1 f(W_1).
\]

Substituting (5.13) and (5.14) into (5.12),

\[
E^* T_1 f(W_1) = \frac{1}{(n-1)^2} E^* [W_2^* f(W_1) + (2 W_3 - W_1 - W_1) f(W_1)].
\]

Again substituting \( U_1 h \) for \( f \), with \( h \) the indicator function of the set \( A \) and using (2.10), (2.11) and bounds on the moments,

\[
|P^*\{\text{tr } G_n \in A\} - P\{Z_1 \in A\}| = |E^* h(W_1) - E_1 h| = |E^* T_1 U_1 h(W_1)|
\]

\[
= \frac{1}{(n-1)^2} \left| E^* [W_2^* (U_1 h)'(W_1) + (2 W_3 - W_1^3 - W_1) (U_1 h)(W_1)] \right|
\]

\[
\leq \frac{1}{(n-1)^2} \left[ \sup |(U_1 h)'| 2E W_2^* + \sup |U_1 h| \sqrt{2E(2W_3 - W_1^3 - W_1)^2} \right]
\]

\[
= \frac{1}{(n-1)^2} \left[ 4 + \sqrt{2} \sqrt{2(34)} \right] < \frac{15}{(n-1)^2}.
\]
6. The two- and three-dimensional cases

It may be instructive to give a separate derivation of the two-dimensional case. Later in this section, I shall also try to specialize the general results of Sections 3 and 4 to the three-dimensional case. I start by studying the distribution of the trace of the matrix $G$ of a two-dimensional rotation through a uniformly distributed random angle $T$. Of course this problem is trivial because the trace of such a rotation is $2 \cos T$, but my aim is to see this in relation to the method of the rest of this paper. Clearly

\begin{equation}
G = \begin{pmatrix}
\cos T & \sin T \\
-\sin T & \cos T
\end{pmatrix}.
\end{equation}

Let $G_\epsilon$ be the matrix of the transformation obtained by applying first $G$ and then a random rotation through an angle $\pm \epsilon$ with equal probabilities for the two possible signs, independent of $G$, so that

\begin{equation}
\text{tr } G_\epsilon - \text{tr } G = 2 \cos (T \pm \epsilon) - 2 \cos T \\
= \pm 2 \epsilon \sin T - \epsilon^2 \cos T + O(\epsilon^3),
\end{equation}

and consequently

\begin{equation}
E^G(\text{tr } G_\epsilon - \text{tr } G) = -\epsilon^2 \cos T + O(\epsilon^3) \\
= -(\epsilon^2/2) \text{tr } G + O(\epsilon^3)
\end{equation}

and

\begin{equation}
E^G(\text{tr } G_\epsilon - \text{tr } G)^2 = 4\epsilon^2 \sin^2 T + O(\epsilon^3) \\
= \epsilon^2[4 - (\text{tr } G)^2] + O(\epsilon^3).
\end{equation}

Because of the exchangeability of the pair $(G, G_\epsilon)$ and the antisymmetry of the function $(g, g^*) \rightarrow (\text{tr } g^* - \text{tr } g)[f(\text{tr } g) + f(\text{tr } g^*)]$, for arbitrary twice continuously differentiable $f$, say, with second derivative bounded in absolute value by $r$,

\begin{equation}
0 = \lim_{\epsilon \to 0} E(\text{tr } G_\epsilon - \text{tr } G)\{2f(\text{tr } G) + [f(\text{tr } G_\epsilon) - f(\text{tr } G)]\}/(2\epsilon^2) \\
= \lim_{\epsilon \to 0} E(\text{tr } G_\epsilon - \text{tr } G)\{2f(\text{tr } G) + (\text{tr } G_\epsilon - \text{tr } G)f'(\text{tr } G) + O(\epsilon^2)\}/(2\epsilon^2) \\
= E[-(\text{tr } G)f(\text{tr } G) + (4 - (\text{tr } G)^2)f'(\text{tr } G)].
\end{equation}

The probability density function $p$ of $W = \text{tr } G$ can be derived from the identity (6.5), essentially by solving the dual differential equation. We can rewrite (6.5) in the form
\[(6.6) \quad 0 = E[-W f(W) + (4 - W^2)f'(W)]
= \int_{-2}^{2}[-w f(w) + (4 - w^2)f'(w)]p(w) \, dw
= \int_{-2}^{2}[-wp(w) - (2wp(w) + (4 - w^2)p'(w))]f(w) \, dw
= \int_{-2}^{2}[wp(w) - (4 - w^2)p'(w)]f(w) \, dw.
\]

Because this must hold for all reasonable functions \(f\),

\[(6.7) \quad wp(w) - (4 - w^2)p'(w) = 0.
\]

The solution of this differential equation is given by

\[(6.8) \quad p(w) = \frac{C}{\sqrt{4 - w^2}}.
\]

Let us compare this with the result of computing the density \(p_w\) of \(W = 2 \cos T\), directly from the density \(p_T\) of \(T\), where \(T\) may be taken to be uniformly distributed in \((0, \pi)\). Since

\[(6.9) \quad dw/dt = -2 \sin t = \pm \sqrt{4 - w^2},
\]

it follows that

\[(6.10) \quad p_w(w)dw = p_T(t)|dt| = |dt|/\pi,
\]

and consequently

\[(6.11) \quad p_w(w) = \frac{1}{\pi \sqrt{4 - w^2}},
\]

in agreement with \((6.8)\).

We shall see that the basic identity \((6.5)\) can be rewritten in a form that is also valid for \(G\), a random reflection, where \(\text{tr} \, G = 0\) with probability 1. This is the form obtained directly from the general result of Theorem 3.1. With \(G\) still a random rotation, using the notation introduced earlier,

\[(6.12) \quad \text{tr} \, G^2 = 2 \cos (2T) = 2(\cos^2 T - \sin^2 T) = 4 \cos^2 T - 2 = (\text{tr} \, G)^2 - 2.
\]
Thus (6.5) is equivalent to

\begin{equation}
E[-(\text{tr} \ G)f(\text{tr} \ G) + (2 - \text{tr} \ G^2)f'(\text{tr} \ G)] = 0.
\end{equation}

However, if \( G \) is a random reflection (indeed any reflection, but that is not the point to be emphasized here), then \( \text{tr} \ G = 0 \) and \( \text{tr} \ G^2 = \text{tr} \ I = 2 \), so that (6.13) remains true. Before turning to the case \( n = 3 \), let us look at the more general case of Theorem 3.1 in which \( f(W) \) is allowed to depend on both \( W_1 = \text{tr} \ G \) and \( W_2 = \text{tr} \ G^2 \). The general result must remain valid despite the perfect dependence between \( W_1 \) and \( W_2 \).

Now let us turn to the case \( n = 3 \). Here both components of the orthogonal group have non-trivial distributions of the trace. It is at first a bit surprising that the conditional distribution of the angle of rotation given that \( G \) is a rotation is not uniform, but this is obvious from the fact that the variance of the trace of \( G \) is 1. Let us proceed with the details, starting with the case where \( G \) is a rotation through an angle \( Q \), so that

\begin{equation}
G \sim \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos Q & \sin Q \\ 0 & -\sin Q & \cos Q \end{pmatrix}
\end{equation}

where \( \sim \) denotes equivalence under an inner automorphism, and

\begin{equation}
W_1 = \text{tr} \ G = 1 + 2\cos Q
\end{equation}

and

\begin{equation}
W_2 = \text{tr} \ G^2 = 1 + 2\cos(2Q) = 1 + 2(2\cos^2 Q - 1) = 4\cos^2 Q - 1
= W_1^2 - 2W_1.
\end{equation}

For \( n = 3 \) and \( k_0 = k = 1 \), (3.1) becomes

\begin{equation}
E^*[f'(W_1) - W_1 f(W_1)] = (1/2)E^*(W_2 - 1)f'(W_1)
= (1/2)E^*(W_1^2 - 2W_1 - 1)f'(W_1)
\end{equation}

and consequently, if \( p \) is the density of \( W_1 

\begin{equation}
0 = E\{-2W_1 f(W_1) + (3 + 2W_1 - W_1^2)f'(W_1)\}
= \int\{-1 < w < 3\}[-2w f(w) + (3 + 2w - w^2)f'(w)]p(w) \, dw
= \int\{-1 < w < 3\} f(w)[-2wp(w) - (-2w + 2)p(w) - (3 + 2w - w^2)p'(w)] \, dw
= \int\{-1 < w < 3\} f(w)[2p(w) + (w + 1)(w - 3)p'(w)] \, dw.
\end{equation}
Since this holds for arbitrary $f$, $p$ must satisfy the first order linear differential equation

\begin{equation}
-2p(w) + (w + 1)(w - 3)p'(w) = 0,
\end{equation}

from which it follows that (writing $p_+$ rather than $p$), for some constant $C$,

\begin{equation}
p_+(w) = C\sqrt{\frac{3 - w}{w + 1}},
\end{equation}

for $-1 < w < 3$, and similarly, for the conditional density $p_-$ of $\text{tr } G$ given $\det G = -1$,

\begin{equation}
p_-(w) = C\sqrt{\frac{3 + w}{1 - w}},
\end{equation}

for $-3 < w < 1$.

References


