Statistics and Probability Letters 137 (2018) 14–18

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

Asymptotic normality of the trace for a class of distributions on orthogonal matrices[☆]

Amir Sepehri

Department of Statistics, Sequoia Hall, Stanford, CA 94305, USA

A R T I C L E   I N F O

Article history:
Received 6 November 2017
Received in revised form 19 December 2017
Accepted 5 January 2018
Available online 16 January 2018

MSC:
60B20
60F05

Keywords:
Gram–Schmidt procedure
Trace test
Random rotations
Asymptotic normality

A B S T R A C T

This paper proves that the trace of the random orthogonal matrices generated by the QR procedure on a matrix with independent identically distributed entries is asymptotically standard normal. This generalizes a well known result of Diaconis and Mallows (1990). A more general case is considered and open problems are suggested.

1. Introduction

A well known result of Diaconis and Mallows (unpublished), states that the trace of a uniformly random orthogonal matrix is asymptotically normal. Their motivation was to study the mixing properties of a Markov chain Monte Carlo sampler on the orthogonal group, suggested in an applied problem in telephone encryption (Sloane, 1982). As explained in Diaconis (2003), the procedure requires a stream of random rotation matrices. The following is a simple algorithm to generate uniform random rotations: fill out an empty $n \times n$ matrix with independent draws from the standard Gaussian distribution, then apply the Gram–Schmidt procedure to construct an orthogonal matrix. This orthogonal matrix is uniformly distributed. This can be easily proved using the fact that the standard multivariate Gaussian distribution is invariant under rotations. This algorithm and its numerical aspects are discussed in detail in Edelman and Rao (2005). The number of operations needed is of order n^3. They required 256×256 random orthogonal matrices, which require order 16×10^6 operations. This was quite slow given the computational resources at the time. In fact, the fastest known, and most commonly used, algorithm is the subgroup algorithm (Diaconis and Shahshahani, 1987) which also requires $O(n^3)$ operations. It was suggested in Sloane (1982) to use the product of relatively few random reflections instead. The asymptotic distribution of the trace of the uniform orthogonal matrices was derived in Diaconis and Mallows (unpublished) in order to lower bound the number of random reflections needed to achieve approximate uniformity for the algorithm proposed in Sloane (1982). This result, along with amazing generalizations, are presented in Diaconis and Shahshahani (1994). Generalizations in different direction, as well as rates of convergence, have been studied in the literature, e.g. see Diaconis and Evans (2001), Johansson (1997), Meckes (2008) and Stein (1995). There is a sizable literature around this problem, and this paper by no means attempts a through review of the literature.

[☆] This work was done during author's doctoral studies at Stanford University.
E-mail address: asepehri@stanford.edu.

https://doi.org/10.1016/j.spl.2018.01.003
0167-7152/

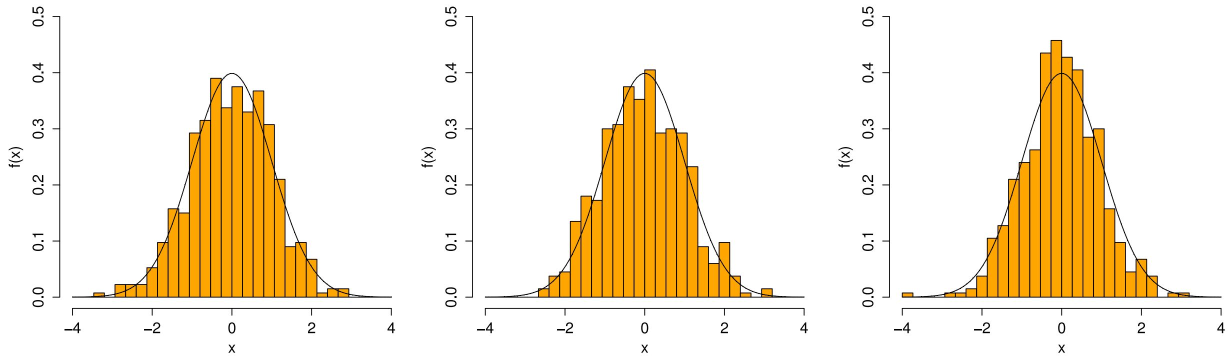

Fig. 1. Histogram of trace of the orthogonalized random $n \times n$ sign matrices for $n = 200$ (left) and $n = 1000$ (middle); and for i.i.d. normal observations (right). Each histogram is based on 500 samples. The standard normal density is also plotted for comparison.

Consider the orthogonal matrix generated by the Gram–Schmidt procedure. As stated above, it is Haar distributed given that the initial matrix has independent standard Gaussian entries. What happens if entries are independent from another distribution? For instance, fill out an empty matrix with independent draws from the uniform distribution on $[-1, 1]$ to get X. Apply Gram–Schmidt procedure to X to construct Q. It is easy to see that this does not generate a Haar distributed rotation, but can be easily mistaken by non-experts. How can we detect the difference between the induced distribution on the orthogonal group with the Haar measure? This problem is considered in Sepehri (under preparation) and different existing and new statistical tests of uniformity have been applied. One of the most basic tests to use is the test based on asymptotic normality of the trace. It is important to know whether this test can detect the Haar distribution from the alternatives described above. This paper provides a negative answer to that question, by proving that the trace is asymptotically standard Gaussian for a broad class of underlying entry distributions.

2. Main result

Theorem 1. *Let F be a symmetric distribution with $0 < \int x^4 F(dx) < \infty$, and X an $n \times n$ matrix with i.i.d. entries from F. Let $X = QR$ be the QR decomposition of X, with Q orthogonal and R upper-triangular with positive diagonal entries. Then, as n goes to infinity,*

$$\mathrm{Tr}(Q) \xrightarrow{w} \mathcal{N}(0, 1).$$

Example 2. An interesting example is letting F be the uniform distribution on $\{-1, 1\}$. This induces a distribution on the orthogonal group, very different from the Haar measure. However, the theorem implies that the Gram–Schmidt procedure applied to a random sign matrix generates an orthogonal component with asymptotically Gaussian trace. This is investigated numerically and results are represented in Fig. 1.

For a Haar distributed orthogonal matrix, M, the distribution of $\mathrm{Tr}(M)$ is close to the standard normal to a great approximation. In fact, Stein (1995) proved:

$$\left| \mathbb{P}\left(\mathrm{Tr}(M) \in B\right) - \int_B \frac{e^{-x^2/2}}{\sqrt{2\pi}} \right| \leq \frac{C_r}{(n-1)^r},$$

for any integer r and an appropriate constant C_r. In particular, $C_2 = 15$ works for $r = 2$. However, as suggested by the classical CLT rates, as well as Fig. 1, this may not be the case for the class of distributions concerned here. An interesting question is to prove an upper bound for the rate of convergence.

Remark 3 (*General Linear Functionals*)**.** It is well known (see D'Aristotile et al. (2003)) that if $M \in \mathcal{O}_n$ is Haar distributed and A is a matrix with $\mathrm{Tr}(A^T A) = n$, then $\mathrm{Tr}(A^T M)$ is asymptotically standard normal. For the class of distributions considered in the present paper, this is generally not the case. For example, let $A_{11} = \sqrt{n}$ and $Aij = 0$ otherwise. Then,

$$\mathrm{Tr}(A^T Q) = \sqrt{n} Q_{11} = \frac{X_{11}}{\sqrt{\frac{\sum_i X_{1i}^2}{n}}},$$

which converges to $\frac{X_{11}}{\sqrt{\mathbb{E}X_{11}^2}}$, because of SLLN and Slutsky's lemma. This is not Gaussian unless F is the Gaussian distribution. On the other hand, using a similar argument as in the proof of Theorem 1, it can be proved that $\mathrm{Tr}(A^T Q)$ is asymptotically standard normal for any matrix A with n entries equal to one such that each column has one nonzero coordinate (similarly for rows). Note that this class is more general than the set of all permutation matrices as it only requires each column to have exactly one non-zero coordinate without imposing a constraint on the rows. For instance, a matrix with one row equal