CHAPTER 7

Random Orthogonal Matrices

Orthogonal matrices, both fixed and random, play an important role in much of statistics, especially in multivariate analysis. Connections between the orthogonal group O_n and the multivariate normal distribution are explored in James (1954) and in Wijsman (1957), as well as in many texts on multivariate analysis. In this chapter, invariance arguments are used to derive the density of a subblock of a uniformly distributed element of O_n. This result is used to describe an upper bound on the rate at which one has convergence (as $n \to \infty$) to the multivariate normal distribution.

7.1. Generating a random orthogonal. Throughout this chapter, O_n denotes the group of $n \times n$ real orthogonal matrices. By the uniform distribution on the compact group O_n, we mean the unique left- (and right-) invariant probability measure. It is possible to represent this distribution using differential forms, but the approach taken here is to represent things in terms of random matrices.

Consider a random matrix X: $n \times q$ with $q \leq n$ and assume

$$\mathscr{L}(X) = N(0, I_n \otimes I_q).$$

Thus, the elements of X are iid $N(0,1)$ random variables. It is well known that X has rank q with probability 1 [for example, see Eaton (1983), Chapter 7]. Thus the random matrix

$$\Gamma_1 = X(X'X)^{-1/2}$$

is well defined. Since $\Gamma_1'\Gamma_1 = I_q$, Γ_1 is a random element of $F_{q,n}$ introduced in Example 2.3. Since O_n acts transitively on $F_{q,n}$, there exists a unique invariant probability measure on $F_{q,n}$, say ν. The following result shows that Γ_1 has distribution ν.

PROPOSITION 7.1. *The random matrix Γ_1 has the uniform distribution on $F_{q,n}$.*

PROOF. From the uniqueness of ν, it suffices to show that

$$\mathscr{L}(\Gamma_1) = \mathscr{L}(g\Gamma_1), \qquad g \in O_n.$$

But, it is clear that

$$\mathscr{L}(X) = \mathscr{L}(gX), \qquad g \in O_n.$$

Thus,

$$\begin{aligned}\mathscr{L}(g\Gamma_1) &= \mathscr{L}\big(gX(X'X)^{-1/2}\big)\\ &= \mathscr{L}\big(gX((gX)'gX)^{-1/2}\big)\\ &= \mathscr{L}\big(X(X'X)^{-1/2}\big) = \mathscr{L}(\Gamma_1). \qquad \square\end{aligned}$$

When $q = 1$, Γ_1 has the uniform distribution on the unit sphere in R^n and when $q = n$, Γ_1 has the uniform distribution on $O_n = F_{n,n}$. The two properties of X which lead to Proposition 7.1 are:

(i) X has rank q a.s.
(ii) $\mathscr{L}(X) = \mathscr{L}(gX)$, $g \in O_n$.

Any random matrix X satisfying (i) and (ii) yields a Γ_1 which is uniform on $F_{q,n}$.

Now, partition X as

$$X = \begin{pmatrix} Y \\ Z \end{pmatrix}$$

with Y: $p \times q$. Then

$$\Gamma_1 = \begin{pmatrix} Y \\ Z \end{pmatrix}(Y'Y + Z'Z)^{-1/2}$$

so that

$$\Delta = Y(Y'Y + Z'Z)^{-1/2} \tag{7.1}$$

is the $p \times q$ upper block of Γ_1. When $p + q \le n$, Khatri (1970) derived the density of Δ using the invariant differential on O_n. Here we derive the density of Δ using an invariance argument.

The following should be noted. If Γ is uniform on O_n and Γ is partitioned as

$$\Gamma = (\Gamma_1 \Gamma_2)$$

with Γ_1: $n \times q$ and Γ_2: $n \times (n - q)$, then Γ_1 has the uniform distribution on $F_{q,n}$. This follows from the observation that for $g \in O_n$,

$$\mathscr{L}(\Gamma) = \mathscr{L}(g\Gamma) = \mathscr{L}(g(\Gamma_1\Gamma_2)) = \mathscr{L}((g\Gamma_1 g\Gamma_2))$$

so that marginally

$$\mathscr{L}(\Gamma_1) = \mathscr{L}(g\Gamma_1).$$

This invariance characterizes $\mathscr{L}(\Gamma_1)$. Therefore, the matrix Δ: $p \times q$ can be thought of as the $p \times q$ upper left block of the random orthogonal matrix Γ.

7.2. The density of a block. Before turning to the density of Δ, we review a few basic facts about the multivariate beta distribution. Consider two independent Wishart matrices S_i, $i = 1, 2$, with

$$\mathscr{L}(S_i) = W(I_q, q, n_i), \qquad i = 1, 2.$$

That is, S_i: $q \times q$ has a Wishart distribution with n_i degrees of freedom and scale matrix I_q. When $n_1 + n_2 \geq q$, then $S_1 + S_2$ is positive definite with probability 1 and the random matrix

$$B = (S_1 + S_2)^{-1/2} S_1 (S_1 + S_2)^{-1/2}$$

is well defined. The matrix B has, by definition, a multivariate beta distribution which is written

$$\mathscr{L}(B) = \mathbf{B}(n_1, n_2; I_q).$$

This notation is from Dawid (1981). Since

$$I_q - B = (S_1 + S_2)^{-1/2} S_2 (S_1 + S_2)^{-1/2},$$

it follows that B can have a density with respect to Lebesgue measure [on the set of symmetric B's with all eigenvalues in (0,1)] iff $n_i \geq q$, $i = 1, 2$. In this case, the density of B is

$$\frac{\omega(n_1, q)\omega(n_2, q)}{\omega(n_1 + n_2, q)} |B|^{(n_1 - q - 1)/2} |I_q - B|^{(n_2 - q - 1)/2}, \tag{7.2}$$

where $\omega(\cdot, \cdot)$ is the Wishart constant [Eaton (1983), page 175]. This and related results can be found in Olkin and Rubin (1964), Mitra (1970) and Khatri (1970).

In what follows, we treat the case $q \leq p$ in the discussion of

$$\Delta = Y(Y'Y + Z'Z)^{-1/2},$$

where

$$\mathscr{L}(X) = \mathscr{L}\left(\begin{pmatrix} Y \\ Z \end{pmatrix}\right) = N(0, I_n \otimes I_q).$$

The case $p \leq q$ is treated by taking transposes.

PROPOSITION 7.2. *When $q \leq p$ and $p + q \leq n$, the random matrix $\Delta'\Delta$ has a $\mathbf{B}(p, n - p; I_q)$ distribution. Thus $\Delta'\Delta$ has a density (with respect to Lebesgue measure) given by*

$$f_0(x) = C_0 |x|^{(p - q - 1)/2} |I_q - x|^{(n - p - q - 1)/2} I_0(x), \tag{7.3}$$

where I_0 is the indicator function of the $q \times q$ symmetric matrices all of whose eigenvalues are in (0,1) *and the constant C_0 is*

$$C_0 = \omega(p, q)\omega(n - p, q)/\omega(n, q).$$

PROOF. Using the normal representation for Δ, we have

$$\begin{aligned}\mathscr{L}(\Delta'\Delta) &= \mathscr{L}\left((Y'Y + Z'Z)^{-1/2} Y'Y (Y'Y + Z'Z)^{-1/2}\right) \\ &= \mathscr{L}\left((S_1 + S_2)^{-1/2} S_1 (S_1 + S_2)^{-1/2}\right).\end{aligned}$$

Since S_1 is $W(I_q, q, p)$ and S_2 is $W(I_q, q, n-p)$, the first assertion follows. The expression for the density of $\Delta'\Delta$ follows immediately from (7.2). □

PROPOSITION 7.3. *For $q \le p$ and $p + q \le n$, the density of Δ is given by*

$$f_1(y) = C_1 |I_q - y'y|^{(n-p-q-1)/2} I_0(y'y), \tag{7.4}$$

where I_0 is given in Proposition 5.2 and the constant C_1 is

$$C_1 = (\sqrt{2\pi})^{-pq} \omega(n-p, q)/\omega(n, q). \tag{7.5}$$

PROOF. Let ψ: $p \times q$ have f_1 as a density. Since $f_1(gy) = f_1(y)$ for $g \in O_p$, it is clear that

$$\mathscr{L}(\psi) = \mathscr{L}(g\psi), \qquad g \in O_p.$$

Because $\psi'\psi$ is a maximal invariant under the group action

$$\psi \to g\psi, \qquad g \in O_p,$$

the results of Example 5.2 show that $\psi'\psi$ has the density f_0 given in (7.3). Therefore,

$$\mathscr{L}(\psi'\psi) = \mathscr{L}(\Delta'\Delta).$$

However, we know that

$$\mathscr{L}(g\Delta) = \mathscr{L}(\Delta), \qquad g \in O_p.$$

The results of Proposition 7.4 below imply that $\mathscr{L}(\psi) = \mathscr{L}(\Delta)$ so that Δ has density (7.3). □

Essentially, the argument used above to conclude that $\mathscr{L}(\psi) = \mathscr{L}(\Delta)$ consists of two parts:

(i) Both ψ and Δ have O_p-invariant distributions.
(ii) The distribution of the maximal invariant under the group action is the same for both ψ and Δ.

The result below shows that the argument is, in fact, a general argument and not specific to the case at hand. To describe the general situation, suppose that the compact group G acts measurably on measurable space $\mathbf{Y}$ and suppose that P_1 and P_2 are two G-invariant probability measures on $\mathbf{Y}$. Let τ: $\mathbf{Y} \to \mathbf{X}$ be a maximal invariant function.

PROPOSITION 7.4. *Let $Y_i \in \mathbf{Y}$ have distribution P_i, $i = 1, 2$. If $\mathscr{L}(\tau(Y_1)) = \mathscr{L}(\tau(Y_2))$, then $P_1 = P_2$.*

PROOF. Let Q be the common distribution of $\mathscr{L}(\tau(Y_i))$, $i = 1, 2$ and let μ be the invariant probability measure on G. Given a bounded measurable function f defined on $\mathbf{Y}$, the function

$$y \to \int_G f(gy)\mu(dg)$$

is invariant and is assumed to be measurable. Thus, this function can be written as $f_0(\tau(y))$ with f_0 defined on $\mathbf{X}$.

Now, using the invariance of P_i and the definition of Q, we have

$$\int f(y)P_i(dy) = \int f(gy)P_i(dy)$$
$$= \int\int f(gy)\mu(dg)P_i(dy) = \int f_0(\tau(y))P_i(dy) = \int f_0(x)Q(dx)$$

for $i = 1, 2$. Thus, for any bounded measurable function f,

$$\int f(y)P_1(dy) = \int f(y)P_2(dy)$$

so $P_1 = P_2$. □

Further properties of the matrix Δ and its distribution can be found in Eaton (1985).

7.3. Some asymptotics. Again consider Δ defined by (5.1) with p and q fixed, but n tending to infinity. The rows of X: $n \times q$, say $X_1', \dots, X_n'$, are iid $N(0, I_q)$. Thus, by the strong law of large numbers,

$$n^{-1}\sum_1^n X_i X_i' \to \mathbf{E}X_1X_1' = I_q.$$

Therefore,

$$\sqrt{n}\,\Delta = \sqrt{n}\,Y(Y'Y + Z'Z)^{-1/2}$$
$$= \sqrt{n}\,Y\left(\sum_1^n X_i X_i'\right)^{-1/2} = Y\left(n^{-1}\sum_1^n X_i X_i'\right)^{-1/2}$$

converges almost surely to Y which is $N(0, I_p \otimes I_q)$. Thus we have:

PROPOSITION 7.5. *Let Γ be uniform on O_n and let Δ be any $p \times q$ subblock of the matrix Γ. Then $\sqrt{n}\,\Delta$ converges in distribution to a $N(0, I_p \otimes I_q)$ distribution as $n \to \infty$.*

We now turn to the question of the rate of convergence of the distribution of $\sqrt{n}\,\Delta$ to the normal. The result described here is from Diaconis, Eaton and Lauritzen (1987). Recall that for two probability measures P_1 and P_2, the variation distance between P_1 and P_2 is defined by

$$\|P_1 - P_2\| = 2\sup_B |P_1(B) - P_2(B)|,$$

where the sup ranges over the relevant σ-algebra. When P_1 and P_2 are both absolutely continuous with respect to a σ-finite measure λ, say $p_i = dP_i/d\lambda$, then

$$\|P_1 - P_2\| = \int |p_1 - p_2|\,d\lambda = 2\int\left(\frac{p_1}{p_2} - 1\right)^+ p_2\,d\lambda,$$

where $f^+ = f$ when f is positive and $f^+ = 0$ otherwise. Therefore,

$$\|P_1 - P_2\| \le 2\sup_x \left(\frac{p_1(x)}{p_2(x)} - 1\right)^+. \tag{7.6}$$

Here is one technical fact concerning variation distance which is used henceforth without comment. For P_1 and P_2 defined on a measurable space $(\mathbf{X}_1, \mathbf{B}_1)$, consider a measurable map f from $(\mathbf{X}_1, \mathbf{B}_1)$ to $(\mathbf{X}_2, \mathbf{B}_2)$. Then $fP_i = Q_i$ is defined on $(\mathbf{X}_2, \mathbf{B}_2)$ by

$$Q_i(B) = (fP_i)(B) = P_i(f^{-1}(B)), \qquad i = 1, 2.$$

Because f is measurable, $f^{-1}(\mathbf{B}_2) \subseteq \mathbf{B}_1$ so that

$$\|Q_1 - Q_2\| \le \|P_1 - P_2\|.$$

In other words, variation distance cannot be increased by measurable transformations.

It is inequality (7.6) which is used to bound the variation distance between the distribution of $\sqrt{n}\,\Delta$ and the $N(0, I_p \otimes I_q)$ distribution. We now proceed with some of the technical details. First note that the variation distance between $\mathscr{L}(\sqrt{n}\,\Delta)$ and $N(0, I_p \otimes I_q)$ is the same as the variation distance between $\mathscr{L}(\Delta)$ and $N(0, n^{-1}I_p \otimes I_q)$ because variation distance is invariant under one-to-one bimeasurable transformations. In the calculation below [from Diaconis and Freedman (1987)], we treat the case of $q = 1$ and $p = 2r$ as an even integer. Under this assumption, the density of Δ is [from (7.4)]

$$p_1(x) = \frac{(\sqrt{2\pi})^{-p} 2^{p/2}\Gamma(n/2)}{\Gamma((n-p)/2)}(1 - x'x)^{(n-p-2)/2} I_0(x'x)$$

for $x \in R^p$. Dividing $p_1(x)$ by the density of a $N(0, n^{-1}I_p)$ distribution, say $p_2(x)$, we have

$$\frac{p_1(x)}{p_2(x)} = \left(\frac{2}{n}\right)^{p/2} \frac{\Gamma(n/2)}{\Gamma((n-p)/2)}(1 - x'x)^{(n-p-2)/2} \exp\left[\frac{nx'x}{2}\right] I_0(x'x).$$

When $p \le n - 3$, this ratio is maximized for $x'x = (p+2)/n$ so the ratio is bounded above by

$$B = \left(\frac{2}{n}\right)^{p/2} \frac{\Gamma(n/2)}{\Gamma((n-p)/2)}\left(1 - \frac{p+2}{n}\right)^{(n-p-2)/2} \exp\left[\frac{p+2}{2}\right].$$

Since $p = 2r$ is even, it follows easily that

$$\begin{aligned}\log\left(\left(\frac{2}{n}\right)^{p/2} \frac{\Gamma(n/2)}{\Gamma((n-p)/2)}\right) &= \sum_1^r \log\left(1 - \frac{2j}{n}\right) \\ &= \sum_1^{r+1} \log\left(1 - \frac{2j}{n}\right) - \log\left(1 - \frac{p+2}{n}\right).\end{aligned}$$

But

$$\sum_{1}^{r+1} \log\left(1 - \frac{2j}{n}\right) \le \int_0^{r+1} \log\left(1 - \frac{2x}{n}\right) dx$$

$$= -\frac{n-p-2}{2} \log\left(1 - \frac{p+2}{n}\right) - \frac{p+2}{2},$$

so

$$\log B \le -\log\left(1 - \frac{p+2}{n}\right).$$

Therefore,

$$B - 1 \le \frac{p+2}{n-p-2}.$$

Summarizing, we have:

PROPOSITION 7.6. [Diaconis and Freedman (1987)]. *If $q = 1$ and $p \le n - 4$, then the variation distance between $\mathscr{L}(\sqrt{n}\,\Delta)$ and a $N(0, I_p)$ distribution is bounded above by* $2(p+3)/(n-p-3)$.

PROOF. From inequality (7.6) when p is even, the variation distance is bounded above by $2(B-1)$ which is bounded above by $2(p+2)/(n-p-2) \le 2(p+3)/(n-p-3)$. The easy argument extending the bound to odd p is given in Diaconis and Freedman (1987). □

When $q > 1$, similar bounds have been established by Diaconis, Eaton and Lauritzen (1987), but the details are substantially more gory. Here is one version of the bound:

PROPOSITION 7.7. *For $p + q \le n - 3$ and $t = \min\{p, q\}$, the variation distance between $\mathscr{L}(\sqrt{n}\,\Delta)$ and a $N(0, I_p \otimes I_q)$ distribution is bounded above by*

$$\delta_n = 2\left[\exp\left[-c \log\left(1 - \frac{p+q+2}{n}\right)\right] - 1\right],$$

where

$$c = \frac{3t^2 + 5t}{8}.$$

PROOF. See Diaconis, Eaton and Lauritzen (1987). □

It is possible to bound δ_n above by an expression of the form $a(p + q + 2)/n$ (a is a constant) when $(p + q + 2)/n$ is bounded away from 1. More explicitly, assume $(p + q + 2)/n \leq \gamma < 1$ and set

$$\phi(x) = 2[\exp[-c\log(1 - x)] - 1], \qquad 0 \leq x \leq \gamma,$$

where the constant c is given in Proposition 7.7. Because ϕ is increasing and convex on $[0, \gamma]$,

$$\phi(x) \leq \frac{\phi(\gamma)x}{\gamma}, \qquad 0 \leq x \leq \gamma.$$

Setting $a = a(\gamma) = \phi(\gamma)/\gamma$ yields the inequality

$$\delta_n \leq a(\gamma)\frac{p + q + 2}{n}, \qquad \frac{p + q + 2}{n} \leq \gamma.$$

This bound is qualitatively the same as that given in Proposition 7.6.