CHAPTER 8

Finite de Finetti Style Theorems

The purpose of this chapter is to introduce the ideas surrounding the so called finite de Finetti style theorems. Four examples, one of which comes from the classical de Finetti theorem and three related to the normal distribution, are discussed here. These examples are introduced by first describing the "infinite version" of a result and then moving to the "finite version." In all of these examples, the infinite version came first, followed by a finite version. However, recent work on finite versions has suggested new infinite versions; some of these are discussed in the next chapter.

8.1. The de Finetti theorem. We begin with a review of the classical de Finetti theorem for an exchangeable infinite sequence of 0-1 valued random variables. Let $\mathbf{X} = \{0, 1\}$ and for each integer n, $1 \le n < +\infty$, let $\mathbf{X}^{(n)}$ be the n-fold product of $\mathbf{X}$ with itself. Given a probability P on the infinite product $\mathbf{X}^{\infty}$, $P^{(n)}$ denotes the projection of P onto $\mathbf{X}^{(n)}$. If $X = (X_1, X_2, \dots)$ is a sequence of random variables with values in $\mathbf{X}^{\infty}$, then $X^{(n)}$ denotes the first n coordinates of X. Thus, if the probability law of X in $\mathbf{X}^{\infty}$ is P, written $\mathscr{L}(X) = P$, then

$$\mathscr{L}(X^{(n)}) = P^{(n)}.$$

Recall that P, a probability on $\mathbf{X}^{\infty}$, is called *exchangeable* if for each n, $P^{(n)}$ on $\mathbf{X}^{(n)}$ is exchangeable, that is, if $P^{(n)}$ is invariant under the action of the permutation group on $\mathbf{X}^{(n)}$. Equivalently, if $X \in \mathbf{X}^{\infty}$, then X is *exchangeable* if for each n, the random vector $X^{(n)}$ has a distribution which is invariant under permutations. As an example, let $Z = (Z_1, Z_2, \dots)$ be a sequence of iid Bernoulli random variables with probability α of success and let P_α denote the distribution of Z on $\mathbf{X}^{\infty}$. Obviously P_α is exchangeable as is any mixture, over α, of P_α. That is, let μ be a probability measure defined on the Borel sets of $[0, 1]$ and define P_μ

on $\mathbf{X}^\infty$ by

$$P_\mu(B) = \int_0^1 P_\alpha(B)\mu(d\alpha) \tag{8.1}$$

for B in the σ-algebra of $\mathbf{X}^\infty$. Thus, for each n,

$$P_\mu^{(n)}(B) = \int_0^1 P_\alpha^{(n)}(B)\mu(d\alpha) \tag{8.2}$$

for relevant sets B. These two equations are often written as

$$P_\mu = \int_0^1 P_\alpha \mu(d\alpha) \tag{8.3}$$

and

$$P_\mu^{(n)} = \int P_\alpha^{(n)}\mu(d\alpha), \tag{8.4}$$

a notation which is adopted here. Thus, P_μ given in (8.3) is exchangeable. The important observation of de Finetti (1931) is:

THEOREM 8.1. *Suppose P on $\mathbf{X}^\infty$ is exchangeable. Then there is a unique probability measure μ on* $[0,1]$ *such that*

$$P = \int_0^1 P_\alpha \mu(d\alpha). \tag{8.5}$$

One consequence of (8.5) is that for each positive integer k,

$$P^{(k)} = \int_0^1 P_\alpha^{(k)}\mu(d\alpha). \tag{8.6}$$

In other words, all of the marginal distributions of P have the representation (8.6). Now, *fix* a finite integer n and assume $P^{(n)}$ on $\mathbf{X}^{(n)}$ is exchangeable. Thus all of the lower dimensional marginals, say $P^{(k)}$ with $1 \le k < n$, are exchangeable. It seems natural to ask if the $P^{(k)}$ have the representation (8.6). The answer is no; an example is given below. However, what is true is that the $P^{(k)}$ "almost" have such a representation when n is a lot bigger than k. The problem is to make this precise. We now turn to a careful discussion of this problem which was solved by Diaconis and Freedman (1980).

The sample space $\mathbf{X}^{(n)}$ consists of n dimensional vectors, which we write as column vectors

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix},$$

where each x_i is 0 or 1. The group of $n \times n$ permutation matrices $\mathscr{P}_n$ acts on the left of $\mathbf{X}^{(n)}$. Consider a probability measure $P^{(n)}$ on $\mathbf{X}^{(n)}$ which is exchangeable, that is, $P^{(n)}$ satisfies

$$gP^{(n)} = P^{(n)}, \qquad g \in \mathscr{P}_n, \tag{8.7}$$

or equivalently,

$$\mathscr{L}(X^{(n)}) = \mathscr{L}(gX^{(n)}), \qquad g \in \mathscr{P}_n,$$

where $P^{(n)} = \mathscr{L}(X^{(n)})$. The results of Example 4.2 give us a representation for $P^{(n)}$. A cross section in this example is

$$\mathbf{Y} = \{y_0, y_1, \ldots, y_n\},$$

where y_i has its first i elements equal to 1 and the remaining elements are 0. Thus $X^{(n)}$ has a representation as $X^{(n)} = UY$ where U is uniform on $\mathscr{P}_n$, Y is independent of U and Y has an arbitrary distribution on $\mathbf{Y}$. Let H_i denote the distribution of Uy_i on $\mathbf{X}^{(n)}$. Obviously H_i is the uniform distribution on the orbit

$$\{gy_i | g \in \mathscr{P}_n\}$$

and H_i puts mass $\binom{n}{i}^{-1}$ on each point in this orbit. Let

$$p_i = \text{Prob}\{Y = y_i\}.$$

From the representation $X^{(n)} = UY$, it is clear that

$$P^{(n)} = \sum_{i=0}^{n} p_i H_i. \tag{8.8}$$

Conversely, any probability measure of the form

$$\sum_{i=0}^{n} p_i H_i, \qquad 0 \le p_i, \Sigma p_i = 1,$$

is exchangeable. Further, the representation is unique because the H_i are mutually singular. Summarizing we have:

Theorem 8.2. *In order that $P^{(n)}$ on $\mathbf{X}^{(n)}$ be exchangeable it is necessary and sufficient that*

$$P^{(n)} = \sum_{i=0}^{n} p_i H_i \tag{8.9}$$

for some $p_i \ge 0$, $\Sigma p_i = 1$. The representation is unique.

It is clear that the set of exchangeable probabilities on $\mathbf{X}^{(n)}$ is a convex set. Theorem 8.2 shows that the extreme points of this convex set are $H_0, H_1, \ldots, H_n$. Now, focus on the exchangeable probability H_1 and let $\mathscr{L}(X^{(n)}) = H_1$. Consider the possibility of representing H_1 in the form (8.6), that is, suppose

$$H_1 = \int_0^1 P_\alpha^{(n)} \mu(d\alpha) \tag{8.10}$$

for some μ where $P_\alpha^{(n)}$ is the probability measure for iid Bernoullis with success probability α. The claim is that (8.10) cannot hold for any μ. On the contrary, if

(8.10) holds, observe that

$$1/n = \mathbf{E}X_1^{(n)} = \int_0^1 \alpha\mu(d\alpha)$$

and

$$0 = \mathbf{E}X_1^{(n)}X_2^{(n)} = \int_0^1 \alpha^2\mu(d\alpha).$$

The second equation implies that $\mu(\{0\}) = 1$ and this contradicts the first equation. This shows Theorem 8.1 is false for every finite n.

Again assume $P^{(n)} = \mathscr{L}(X^{(n)})$ is an exchangeable probability on $\mathbf{X}^{(n)}$. As usual, $X^{(k)}$ is the vector of the first k coordinates of $X^{(n)}$ where $P^{(k)} = \mathscr{L}(X^{(k)})$. Obviously $P^{(k)}$ is an exchangeable probability on $\mathbf{X}^{(k)}$ and $P^{(k)}$ is the "projection" of $P^{(n)}$ down to $\mathbf{X}^{(k)}$. More precisely, let π be the $k \times n$ matrix defined by

$$\pi = (I_k \quad 0): k \times n,$$

where I_k is the $k \times k$ identity matrix. Obviously

$$\pi X^{(n)} = X^{(k)}$$

so

$$\pi P^{(n)} = P^{(k)},$$

where

$$(\pi P^{(n)})(B) = P^{(n)}(\pi^{-1}(B))$$

for subsets B of $\mathbf{X}^{(k)}$. A main result in Diaconis and Freedman (1980) shows that

$$\Delta_{k,n} = \inf_{\mu}\left\| P^{(k)} - \int_0^1 P_\alpha^{(k)}\mu(d\alpha)\right\| \le 4k/n, \tag{8.11}$$

where $\|\cdot\|$ denotes variation distance (as discussed in Chapter 7) and the inf is over all the Borel measures on $[0,1]$. The interpretation of (8.11) is that when $P^{(k)}$ is the projection of an exchangeable probability on $\mathbf{X}^{(n)}$, then $P^{(k)}$ is within $4k/n$ of some mixture of iid Bernoullis. The basic step in the proof of (8.11) is the following:

THEOREM 8.3. *The variation distance between πH_i and $P_\alpha^{(k)}$ with $\alpha = i/n$ is bounded above by $4k/n$.*

PROOF. With $\mathscr{L}(X^{(k)}) = \pi H_i$, $X^{(k)}$ is the outcome of k draws made without replacement from an urn with i 1's and $n - i$ 0's. But $P_\alpha^{(k)}$ represents the probability measure of k draws made with replacement from the same urn. Bounding the variation distance between πH_i and $P_\alpha^{(k)}$, which involves some calculus, is carried out in Lemma 6 of Diaconis and Freedman (1980). □

THEOREM 8.4. *Given an exchangeable $P^{(n)}$ on $\mathbf{X}^{(n)}$ and $P^{(k)} = \pi P^{(n)}$, Equation* (8.11) *holds.*

PROOF. First use Theorem 8.2 to write

$$P^{(n)} = \sum_{i=0}^{n} p_i H_i$$

so that

$$P^{(k)} = \pi P^{(n)} = \sum_{i=0}^{n} p_i \pi H_i.$$

Let μ_0 be the probability on $[0,1]$ which puts mass p_i at the point $\alpha_i = i/n$, $i = 0, \ldots, n$. Then

$$\Delta_{k,n} = \inf_{\mu} \left\| P^{(k)} - \int_0^1 P_\alpha^{(k)} \mu(d\alpha) \right\| \le \left\| P^{(k)} - \int_0^1 P_\alpha^{(k)} \mu_0(d\alpha) \right\|$$

$$= \left\| \sum_0^n p_i \pi H_i - \sum_0^n p_i P_{\alpha_i}^{(k)} \right\| \le \sum_0^n p_i \left\| \pi H_i - P_{\alpha_i}^{(k)} \right\| \le 4k/n,$$

where the last inequality follows from Theorem 8.3. Then next to the last inequality is a consequence of the fact that variation distance is a norm and hence is a convex function. □

The argument given above shows that to bound $\Delta_{k,n}$ in (8.11), it is sufficient (and necessary) to bound $\Delta_{k,n}$ when $P^{(k)}$ is one of the projected extreme points πH_i, $i = 0, \ldots, n$. This type of argument is used in all of the examples in this and the next chapter. Theorem 8.4 is often called a finite style de Finetti theorem because n and k are both fixed and finite. This result can be used to provide an easy proof of the infinite de Finetti theorem. For example, see Theorem 14 in Diaconis and Freedman (1980) where the sort of argument used above provides an easy proof of the Hewitt–Savage (1955) generalization of the de Finetti theorem. An interesting related paper is Dubins and Freedman (1979).

Finally, a few remarks about extendability. Theorem 8.4 concerns those $P^{(k)}$ on $\mathbf{X}^{(k)}$ which are *n-extendable* in the sense that there exists an exchangeable $P^{(n)}$ on $\mathbf{X}^{(n)}$ such that

$$P^{(k)} = \pi P^{(n)}.$$

Thus, an n-extendability assumption on $P^{(k)}$ is equivalent to saying that $P^{(k)}$ is the projection of some exchangeable $P^{(n)}$ on $\mathbf{X}^{(n)}$. This latter condition is a bit more convenient and will appear throughout this and the next chapter. However, the reader should keep the equivalence in mind since n-extendability sometimes is a bit easier to think about.

The results of this section show that if $P^{(k)}$ is n-extendable for all large n, then $P^{(k)}$ has the representation (8.6). However, if $P^{(k)}$ is n-extendable for some fixed n, then (8.6) need not hold, but when n is much bigger than k, then (8.6) almost holds in the sense of Theorem 8.4.

8.2. Orthogonally invariant random vectors. The material in this section is related to Example 4.3. For $1 \le n \le \infty$, let R^n denote n dimensional coordinate space. Given $X = (X_1, X_2, \ldots)$ in R^∞, X has an *orthogonally invariant* distribution if for each finite n,

$$X^{(n)} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix}$$

has an O_n invariant distribution. If $X \in R^\infty$ and $P = \mathscr{L}(X)$, we say P is *orthogonally invariant* if X is orthogonally invariant. Of course, this means that for each finite n, the projected measures

$$P^{(n)} = \mathscr{L}(X^{(n)})$$

are O_n invariant.

For example, if $Z = (Z_1, Z_2, \dots)$ has iid coordinates which are $N(0, \sigma^2)$, let P_σ denote the probability on R^∞ of Z. Then $P_\sigma^{(n)}$ is the joint distribution of

$$Z^{(n)} = \begin{pmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_n \end{pmatrix}.$$

In other words,

$$\mathscr{L}(Z^{(n)}) = N(0, \sigma^2 I_n)$$

for $0 \le \sigma < +\infty$, so P_σ is orthogonally invariant. Given any probability μ on $[0, \infty)$, it is clear that

$$P_\mu = \int_0^\infty P_\sigma \mu(d\sigma) \tag{8.12}$$

is orthogonally invariant. Probabilities of the form (8.12) are called scaled mixtures of normals. When (8.12) holds, then for each n,

$$P_\mu^{(n)} = \int_0^\infty P_\sigma^{(n)} \mu(d\sigma). \tag{8.13}$$

In the present setting, here is the "infinite theorem."

THEOREM 8.5. *P on R^∞ is orthogonally invariant iff P has the representation* (8.12). *Further the representation in* (8.12) *is unique.*

This result is commonly attributed to Schoenberg, but see Section 6 in Diaconis and Freedman (1987). A proof of this theorem, based on the "finite version" given below, can be found in Theorem 3 in Diaconis and Freedman (1987). The representation has been rediscovered in a number of different contexts, for example, see Hill (1969), Andrews and Mallows (1974) and Eaton (1981). The uniqueness part of the theorem follows easily from the uniqueness of Laplace transforms because (8.12) implies

$$P_\mu^{(1)} = \int_0^\infty P_\sigma^{(1)} \mu(d\sigma).$$

Thus $P_\mu^{(1)}$ has characteristic function

$$t \to \int_0^\infty \exp\left[-\tfrac{1}{2}\sigma^2 t^2\right]\mu(d\sigma).$$

Therefore, if μ_1 and μ_2 both represent P_1, they have the same Laplace transforms and hence are equal.

We now turn to a finite version of Theorem 8.5. Fix a positive integer n and let $P^{(n)}$ be an O_n-invariant probability on R^n. Given $r \geq 0$, let H_r denote the uniform distribution on

$$\{x \mid x \in R^n,\ \|x\| = r\},$$

the sphere of radius r in R^n. Naturally H_0 is the probability degenerate at $0 \in R^n$. Clearly each H_r is O_n-invariant. The arguments given in Chapter 4 establish:

THEOREM 8.6. *A probability $P^{(n)}$ on $R^{(n)}$ is O_n-invariant iff for some Borel measure μ on $[0, \infty)$,*

$$P^{(n)} = \int_0^\infty H_r \mu(dr). \tag{8.14}$$

It is clear that the O_n-invariant probability H_1 cannot be represented in the form (8.12). Thus, Theorem 8.5 is false for any finite integer n. To establish an analog of Theorem 8.4 in the present context, fix an integer $k < n$ and let $P^{(k)}$ be the probability measure of the first k coordinates of $X^{(n)}$ where $P^{(n)} = \mathscr{L}(X^{(n)})$. Further, let

$$\pi = (I_k \quad 0) \colon k \times n$$

be a $k \times n$ real matrix so

$$X^{(k)} = \pi X^{(n)}$$

and

$$P^{(k)} = \pi P^{(n)}.$$

The main result below, due to Diaconis and Freedman (1987), shows that $P^{(k)}$ is close to a scale mixture of normals in the following sense:

THEOREM 8.7. *Assume $P^{(n)}$ is O_n-invariant and $k \leq n - 4$. Then, with $P^{(k)} = \pi P^{(n)}$,*

$$\Delta_{k,n} = \inf_\mu \left\| P^{(k)} - \int_0^\infty P_\sigma^{(k)} \mu(d\sigma) \right\| \leq \frac{2(k+3)}{n-k-3}, \tag{8.15}$$

where $\|\cdot\|$ denotes variation distance and the inf *is over all Borel measures on $[0, \infty)$.*

The proof of this theorem follows much the same lines as the proof of Theorem 8.4. Equation (8.15) is first established for πH_r and then (8.14) is used for the general case.

THEOREM 8.8. *Inequality* (8.15) *holds for* $P^{(k)} = \pi H_r$ *for each* $r \geq 0$.

PROOF. For $r = 0$, the result is obvious. For $r > 0$, H_r is the probability measure of the random vector

$$X^{(n)} = rU^{(n)},$$

where $U^{(n)}$ is uniform on the sphere of radius 1 in R^n. Thus, πH_r is the distribution of

$$\pi X^{(n)} = rU^{(k)}.$$

Taking $p = k$ in Proposition 7.6 shows that

$$-\left\| \mathscr{L}(\sqrt{n}\, U^{(k)}) - N(0, I_k) \right\| \leq 2(k+3)/(n-k-3).$$

Because variation distance is invariant under one-to-one bimeasurable transformations, this implies that

$$\left\| \mathscr{L}(rU^{(k)}) - N\left(0, n^{-1}r^2 I_k\right) \right\| \leq 2(k+3)/(n-k-3). \tag{8.16}$$

Hence (8.15) holds for πH_r because $\pi H_r = \mathscr{L}(rU^{(k)})$. □

PROOF OF THEOREM 8.7. Because $P^{(n)}$ is O_n-invariant, (8.14) implies that

$$P^{(k)} = \pi P^{(n)} = \int_0^\infty \pi H_r \mu_0(dr)$$

for some μ_0. Thus, using (8.16),

$$\begin{aligned}
\inf_\mu &\left\| P^{(k)} - \int_0^\infty N\left(0, r^2 I_k\right)\mu(dr) \right\| \\
&\leq \left\| P^{(k)} - \int_0^\infty N\left(0, n^{-1}r^2 I_k\right)\mu_0(dr) \right\| \\
&= \left\| \int_0^\infty \pi H_r \mu_0(dr) - \int_0^\infty N\left(0, n^{-1}r^2 I_k\right)\mu_0(dr) \right\| \\
&\leq \int_0^\infty \left\| \pi H_r - N\left(0, n^{-1}r^2 I_k\right) \right\| \mu_0(dr) \leq 2(k+3)/(n-k-3). \qquad \square
\end{aligned}$$

The essentials of the argument are much the same as they were in Section 8.1, namely, the set of O_n-invariant probabilities is a convex set with extreme points H_r, $r \geq 0$. Thus, to approximate $P^{(k)}$ well by a scale mixture of normals, it is sufficient to approximate πH_r well (in this case, uniformly) by scaled normals. This is what Theorem 8.8 together with Proposition 7.6 does.

The remarks concerning extendability made at the end of the previous section apply here. In particular, if $P^{(k)}$ on R^k is O_k-invariant and if $P^{(k)}$ is n-extendable (that is, $P^{(k)} = \pi P^{(n)}$ for some O_n-invariant $P^{(n)}$ on R^n), then $P^{(k)}$ is within $2(k+3)/(n-k-3)$ of some scale mixture of normals. This is just a restatement of Theorem 8.7.

8.3. Orthogonally invariant random matrices. Here, the results of the previous section are extended to the matrix case. First a bit of notation is needed. Fix a positive integer q and let $\mathscr{L}_{q,n}$ be the vector space of all real $n \times q$ matrices, $1 \le n \le +\infty$. Given a random matrix X in $\mathscr{L}_{q,\infty}$, let $X^{(n)}$: $n \times q$ for $1 \le n < +\infty$ denote the matrix in $\mathscr{L}_{q,n}$ consisting of the first n rows of X. If $P = \mathscr{L}(X)$ is the distribution of X, then $P^{(n)}$ denotes the distribution of $X^{(n)}$. The group O_n acts on $\mathscr{L}_{q,n}$ via matrix multiplication on the left:

$$x \to gx, \qquad x \in \mathscr{L}_{q,n}, \qquad g \in O_n.$$

A probability P on $\mathscr{L}_{q,\infty}$ is *left-orthogonally invariant* if for each finite n,

$$P^{(n)} = gP^{(n)}, \qquad g \in O_n.$$

Thus, if $\mathscr{L}(X) = P$ and P is left-orthogonally invariant, then

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

for each finite n.

As an example, consider Z in $\mathscr{L}_{q,\infty}$ whose rows $Z_1', Z_2', \dots$ are iid $N_q(0, \alpha^2)$ where α is a $q \times q$ positive semidefinite matrix. Then

$$\mathscr{L}\big(Z^{(n)}\big) = N\big(0, I_n \otimes \alpha^2\big)$$

with $\otimes$ denoting the Kronecker product. Let $P_\alpha = \mathscr{L}(Z)$ so P_α is obviously left-orthogonally invariant. Further, given any probability measure μ on the set $\mathbf{S}$ of $q \times q$ positive semidefinite matrices, the probability

$$P_\mu = \int P_\alpha \mu(d\alpha) \tag{8.17}$$

is also left-orthogonally invariant since

$$P_\mu^{(n)} = \int P_\alpha^{(n)} \mu(d\alpha). \tag{8.18}$$

The converse of this observation, established in Dawid (1977), is:

THEOREM 8.9. *Assume P on $\mathscr{L}_{q,\infty}$ is left-orthogonally invariant. Then P has the representation* (8.17). *Further, the representation is unique.*

This "infinite" theorem is usually stated as "P is left-orthogonally invariant iff P is a covariance mixture of normals." The uniqueness of μ is proved in the same way it is proved in the case $q = 1$. Theorem 8.9 can be proved using the finite version of this theorem to which we now turn.

As in the two previous sections, now fix a finite n and consider $P^{(n)}$ on $\mathscr{L}_{q,n}$ which is left-orthogonally invariant. Our first task is to apply Theorem 4.1 to the case at hand. The group O_n acts on $\mathscr{L}_{q,n}$. To specify a cross section in $\mathscr{L}_{q,n}$, let

$$\mathbf{Y} = \left\{ x \mid x \in \mathscr{L}_{q,n}, x = \begin{pmatrix} \alpha \\ 0 \end{pmatrix}, \alpha \in \mathbf{S} \right\},$$

and define τ on $\mathscr{L}_{q,n}$ to $\mathbf{Y}$ by

$$\tau(x) = \begin{pmatrix} (x'x)^{1/2} \\ 0 \end{pmatrix}.$$

Here, $(x'x)^{1/2}$ denotes the unique positive semidefinite square root of $x'x \in \mathbf{S}$. That $\mathbf{Y}$ is a measurable cross section (according to Definition 4.1) is easily checked. Theorem 4.3 yields:

THEOREM 8.10. *For $\alpha \in \mathbf{S}$, let H_α denote the distribution of*

$$U\begin{pmatrix}\alpha\\0\end{pmatrix},$$

where U is uniform on O_n and $\begin{pmatrix}\alpha\\0\end{pmatrix}$ is in $\mathbf{Y}$. Then $P^{(n)}$ on $\mathscr{L}_{q,n}$ is left-orthogonally invariant iff

$$P^{(n)} = \int H_\alpha \mu(d\alpha) \tag{8.19}$$

for some probability μ on $\mathbf{S}$.

PROOF. Apply Theorem 4.3 with $H_\alpha = \mu_y$ and $\mu = Q$. □

Now, let π denote the $k \times n$ matrix

$$\pi = (I_k \quad 0),$$

where $k < n$. If $P^{(n)} = \mathscr{L}(X^{(n)})$, then

$$P^{(k)} = \pi P^{(n)} = \mathscr{L}(\pi X^{(n)}).$$

To establish a finite theorem for $P^{(k)}$, we first establish a finite theorem for πH_α and then use (8.19). □

THEOREM 8.11. *For $k + q \le n - 3$, the variation distance between πH_α and the normal distribution $N(0, n^{-1}I_k \otimes \alpha^2)$ is bounded above by δ_n given in Proposition 7.7.*

PROOF. Recall that H_α is the distribution of

$$U\begin{pmatrix}\alpha\\0\end{pmatrix} = U\begin{bmatrix}I_q\\0\end{bmatrix}\alpha,$$

where U is uniform on O_n. Thus, πH_α is the distribution of

$$\pi U\begin{pmatrix}\alpha\\0\end{pmatrix} = (I_k \quad 0)U\begin{bmatrix}I_q\\0\end{bmatrix}\alpha = \Delta\alpha,$$

where Δ is the $k \times q$ upper left corner of U. Proposition 7.7 implies that

$$\left\| \mathscr{L}(\Delta) - N\left(0, n^{-1}I_k \otimes I_q\right) \right\| \le \delta_n.$$

Since $\pi H_\alpha = \mathscr{L}(\Delta\alpha)$, the result follows. □

The following finite theorem is from Diaconis, Eaton and Lauritzen (1987).

THEOREM 8.12. *Suppose $P^{(n)}$ on $\mathscr{L}_{q,n}$ is left-orthogonally invariant. If $k + q \le n - 3$, then*

$$\inf_\mu \left\| P^{(k)} - \int N\left(0, I_k \otimes \alpha^2\right)\mu(d\alpha) \right\| \le \delta_n, \tag{8.20}$$

where the inf *ranges over all probabilities on* $\mathbf{S}$ *and* δ_n *is given in Proposition* 7.7 (*with p replaced by* k).

PROOF. Since $P^{(n)}$ is left-orthogonally invariant, we can write

$$P^{(n)} = \int H_\alpha \mu_0(d\alpha)$$

for some probability μ_0 on $\mathbf{S}$. Therefore,

$$P^{(k)} = \pi P^{(n)} = \int \pi H_\alpha \mu_0(d\alpha).$$

Since variation distance is a convex function, Theorem 8.11 yields

$$\left\| P^{(k)} - \int N(0, n^{-1} I_k \otimes \alpha^2) \mu_0(d\alpha) \right\|$$

$$= \left\| \int \left[\pi H_\alpha - N(0, n^{-1} I_k \otimes \alpha^2) \right] \mu_0(d\alpha) \right\|$$

$$\leq \int \left\| \pi H_\alpha - N(0, n^{-1} I_k i \otimes \alpha^2) \right\| \mu_0(d\alpha) \leq \delta_n.$$

Hence (8.20) holds. □

The comments concerning extendability made at the end of the last section are valid here. Of course, extendability refers to increasing n with fixed k and q.

8.4. A linear model example. Some new considerations arise when we try to formulate a finite version of an "infinite" theorem described in Smith (1981). To describe the infinite result, let R^n, $1 \leq n \leq \infty$, denote n dimensional coordinate space and for each finite n, let

$$O_n(e) = \{g | g \in O_n, ge = e\},$$

where e is the vector of 1's in R^n. Let $Z = (Z_1, Z_2, \dots) \in R^\infty$ have coordinates which are iid $N(m, \sigma^2)$ where $m \in R^1$ and $\sigma \geq 0$. Thus the distribution of

$$Z^{(n)} = \begin{pmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_n \end{pmatrix}$$

is $N(me, \sigma^2 I_n)$. Clearly

$$\mathscr{L}(Z^{(n)}) = \mathscr{L}(gZ^{(n)}), \qquad g \in O_n(e).$$

$P_{m,\sigma}$ denotes the distribution of Z on R^∞ and

$$\mathscr{L}(Z^{(n)}) = P_{m,\sigma}^{(n)} = N(me, \sigma^2 I_n).$$

Given a probability μ on $R^1 \times [0, \infty)$, let

$$P_\mu = \int\int P_{m,\sigma} \mu(dm, d\sigma) \tag{8.21}$$

so P_μ is a translation-scale mixture of iid normals. Thus the projection of P_μ on R^n is given by

$$P_\mu^{(n)} = \int\int P_{m,\sigma}^{(n)} \mu(dm, d\sigma).$$

Clearly $gP_\mu^{(n)} = P_\mu^{(n)}$ for $g \in O_n(e)$.

THEOREM 8.13 [Smith (1981)]. *Let P be any probability on R^∞. Then the projection of P on R^n, say $P^{(n)}$, is $O_{(n)}(e)$-invariant for all $n = 1, 2, \dots$ iff P has the representation* (8.21).

To describe a finite version of this result, we first need a representation for $P^{(n)}$ defined on R^n (fixed n) which is $O_n(e)$-invariant. Here is a convenient cross section for this example. Fix the vector

$$x_0 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

and let

$$\mathbf{Y} = \left\{ x | x \in R^n,\ x = \sigma x_0 + me;\ \sigma \geq 0,\ m \in R^1 \right\}.$$

Define τ on R^n to $\mathbf{Y}$ by

$$\tau(x) = \|x - \bar{x}e\| x_0 + \bar{x}e,$$

where

$$\bar{x} = n^{-1} \sum_1^n x_i$$

and $\|\cdot\|$ denotes standard Euclidean distance. That $\mathbf{Y}$ is a measurable cross section is easily verified. Given $\sigma \geq 0$ and $m \in R^1$, let $H_{m,\sigma}$ be the distribution of

$$\sigma U x_0 + me,$$

where U is uniform on $O_n(e)$. Note that the random vector Ux_0 has a uniform distribution on

$$\{x | x \in R^n,\ \|x\| = 1,\ x'e = 0\}.$$

THEOREM 8.14. *Let $P^{(n)}$ be a probability on R^n. Then, $P^{(n)}$ is $O_n(e)$-invariant iff*

$$P^{(n)} = \int\int H_{m,\sigma} \mu(dm, d\sigma)$$

for some probability μ on $R^1 \times [0, \infty)$.

PROOF. This is an easy application of Theorem 4.3. □

As usual, we use

$$\pi = (I_k \quad 0) : k \times n$$

to project down from R^n to R^k with $k < n$. The next step in the argument is to approximate $\pi H_{m,\sigma}$ by some normal distribution. The approximation is based on the following:

LEMMA 8.1. *For U uniform on $O_n(e)$, $\pi U x_0$ is distributed as AV where:*

(i) *V is distributed as the first k coordinates of a random vector which has a uniform distribution on*

$$\{x | x \in R^{n-1}, \|x\| = 1\}.$$

(ii) *The $k \times k$ fixed matrix A is given by*

$$A = (\pi Q_0 \pi')^{1/2}$$

with

$$Q_0 = I_n - n^{-1} e e'.$$

PROOF. See Proposition A.1 in Diaconis, Eaton and Lauritzen (1987). □

THEOREM 8.15. *For $k \le n - 5$,*

$$\left\| \pi H_{m,\sigma} - N\left(m\pi e, (n-1)^{-1}\sigma^2 I_k\right) \right\| \le \beta_n, \tag{8.22}$$

where

$$\beta_n = 2\frac{k+3}{n-k-4} + 2\left[(\det A)^{-1} - 1\right]. \tag{8.23}$$

PROOF. The probability $\pi H_{m,\sigma}$ is the law of

$$\sigma \pi U x_0 + m\pi e,$$

which, according to Lemma 8.1, is the same as the law of

$$\sigma A V + m\pi e.$$

Here, A and V are as defined in Lemma 8.1. For notational convenience, let W be $N(0, I_k)$. Thus, the left side of (8.22) is

$$\begin{aligned}
&\left\| \mathscr{L}(\sigma A V + m\pi e) - \mathscr{L}\left((n-1)^{-1/2}\sigma W + m\pi e\right) \right\| \\
&\quad \le \left\| \mathscr{L}(AV) - \mathscr{L}\left((n-1)^{-1/2} W\right) \right\| \\
&\quad \le \left\| \mathscr{L}(AV) - \mathscr{L}\left((n-1)^{-1/2} A W\right) \right\| \\
&\qquad + \left\| \mathscr{L}\left((n-1)^{-1/2} A W\right) - \mathscr{L}\left((n-1)^{-1/2} W\right) \right\| \\
&\quad \le \left\| \mathscr{L}(V) - \mathscr{L}\left((n-1)^{-1/2} W\right) \right\| + \left\| \mathscr{L}(AW) - \mathscr{L}(W) \right\|.
\end{aligned}$$

But, Proposition 7.6 (with n replaced by $n-1$) yields

$$\left\| \mathscr{L}(V) - \mathscr{L}\left((n-1)^{-1/2}W\right)\right\| \le 2\frac{k+3}{n-k-4}$$

for $k \le n-5$. Because all the eigenvalues of A are less than or equal to 1, the easily established inequality

$$\| \mathscr{L}(AW) - \mathscr{L}(W)\| \le 2\left[(\det A)^{-1} - 1\right]$$

completes the proof. □

Finally, we come to the finite version of Theorem 8.13.

THEOREM 8.16. *Given $P^{(n)}$ on R^n which is $O_n(e)$-invariant and $k \le n-5$, let $P^{(k)} = \pi P^{(n)}$. Then*

$$\inf_{\mu}\left\| P^{(k)} - \int\int N\left(m\pi e, \sigma^2 I_k\right)\mu(dm, d\sigma)\right\| \le \beta_n, \tag{8.24}$$

where the inf *ranges over all probabilities on $R^1 \times [0,\infty)$ and β_n is given in* (8.23).

PROOF. Since $P^{(n)}$ is $O_n(e)$-invariant, Theorem 8.14 yields

$$P^{(k)} = \pi P^{(n)} = \int\int \pi H_{m,\sigma}\mu_0(dm, d\sigma)$$

for some μ_0. Thus

$$\begin{aligned}\inf_{\mu}&\left\| P^{(k)} - \int\int N\left(m\pi e, \sigma^2 I_k\right)\mu(dm, d\sigma)\right\| \\ &\le \left\| \int\int\left[\pi H_{m,\sigma} - N\left(m\pi e, (n-1)^{-1}\sigma^2 I_k\right)\right]\mu_0(dm, d\sigma)\right\| \\ &\le \int\int\left\|\pi H_{m,\sigma} - N\left(m\pi e, (n-1)^{-1}\sigma^2 I_k\right)\right\|\mu_0(dm, d\sigma) \le \beta_n.\end{aligned}$$

The final inequality follows from Theorem 8.15. □

The upper bound β_n in (8.23) and (8.24) consists of two parts. The argument used to prove Theorem 8.15 pinpoints the origin of the two pieces. The first piece is from a routine application of Proposition 7.6 which we understand fairly well. The second piece arises because the group in question leaves the subspace span $\{e\}$ fixed so that previous arguments must be modified by dropping down one dimension. The reduction in dimension introduces the $k \times k$ matrix A which appears in the bound as

$$2\left[(\det A)^{-1} - 1\right].$$

A routine calculation shows that

$$\det A^2 = 1 - \frac{k}{n}$$

so

$$(\det A)^{-1} - 1 \leq \left(1 - \frac{k}{n}\right)^{-1/2} - 1 \leq \frac{k}{n-k}$$

for $k \leq n - 5$. Thus β_n is bounded above by

$$4\frac{k+3}{n-k-4}$$

for $k \leq n - 5$. This bound is of the same type as obtained for the previous finite theorems (a constant times k/n for k/n bounded away from 1). From this, we conclude that the situation considered in this section is qualitatively the same as the situation in Section 8.2.

The finite result of this section is from Diaconis, Eaton and Lauritzen (1987) where a multivariate version of Theorem 8.16 is also proved. The previous remarks on extensions are of course valid for the situation of this section.