The Annals of Statistics
1987, Vol. 15, No. 3, 1257–1268

BAYES PROCEDURES FOR ROTATIONALLY SYMMETRIC MODELS ON THE SPHERE

BY ALBERT Y. LO[1] AND JAVIER CABRERA

SUNY at Buffalo and Rutgers University

Consistency of Bayes procedures on the sphere is studied. Equivalent conditions for the consistency of Bayes procedures for a rotationally symmetric model are given. Equivalent conditions for that of an antipodally symmetric model are also provided.

1. Introduction. The problem of estimating the direction of symmetry for a rotationally symmetric model on a sphere has been studied extensively from a frequentist viewpoint [for a good list of references, see Watson (1983a)]. A popular estimate is the unit vector at the direction of the sample average [Watson (1983a), Chapter 4]. A great advantage of this estimate is that the central limit theorem can be applied to give the large-sample distribution and, hence, large-sample frequentist inferential methods based on the sample mean are possible. However, the following example shows that this method of estimating the direction of symmetry breaks down easily for a model identifiable in terms of the direction of symmetry.

EXAMPLE 1.1. *The zero-mean model.* Let $X_1, \dots, X_n$ be an i.i.d. sample from a rotationally symmetric distribution (with respect to a point μ on the sphere) on the sphere in R^k, $k \geq 2$. Let $T_n = n^{-1}\Sigma X_i$ be the sample mean and $T_n/|T_n|$ ($|a|$ denotes the length of the vector a) be the sample mean direction, i.e., the unit vector at the direction of the sample mean. Suppose the mean of X_1 is the zero vector and the covariance matrix of X_1 is Σ with rank $\rho \geq 1$, then $T_n/|T_n|$ does not converge to μ in probability. To see this, it suffices to note that according to the central limit theorem and the continuous mapping theorem, $\sqrt{n}\, T_n/|\sqrt{n}\, T_n| = T_n/|T_n|$ converges in distribution to $Z/|Z|$ which has a nondegenerate angular normal distribution on the sphere in R^ρ [i.e., Z is an $N(0, \Sigma)$ random vector; see Watson (1983a)]. For example, let $k = 2$ and let the model distribution be the mixture of a von Mises distribution with mean direction v and a point mass distribution at $-v$. Assume there are appropriate mixing weights such that the center of mass of the model distribution is at the origin; in this case $Z/|Z|$ has a uniform distribution on the two point set $\{-v, v\}$. Nevertheless, this mixture model is (classically) identifiable in the sense that different v's correspond to different model distributions.

The estimation of the axis of symmetry for an antipodally symmetric model (i.e., X and $-X$ have the same distribution, defined in Section 4) is rather

Received October 1985; revised October 1986.

[1]The research of this author was supported in part by NSF Grant MCS81-02523-01.

AMS 1980 *subject classification*. Primary 62099.

Key words and phrases. Bayes procedures, rotational symmetry, antipodal symmetry, strong consistency.

similar. A popular estimate, the sample mean projector, is studied in great detail in Chapter 5 of Watson (1983a). Again, the central limit theorem is readily applicable to give nice large-sample frequentist inferential methods based on the sample mean projector. However, if the model has the identity matrix as its mean projector, the sample mean projector invariably converges to the identity matrix, which is not the "true" projector defining the axis of symmetry. The following example illustrates a case in point.

EXAMPLE 1.2. *The identity mean projector model.* Assume i.i.d. sampling from a mixture of a two-point-mass distribution (giving 0.5 mass to v and $-v$) and a Scheiddegger–Watson distribution [Watson (1983a)] such that the two axes of symmetry are orthogonal. The mixing weights are such that the mean of $X_1 X_1^*$ is the identity matrix. The resulting model is an antipodally symmetric model on the sphere. The sample average is useless since the mean is zero. The sample mean projector is not helpful since its limit is the identity matrix. Note that this model is identifiable with respect to the axis going through v and $-v$, i.e., the projector defined by vv^*.

These estimation problems were also studied by Kim (1978). Kim proposed the estimation of the direction and the axis of symmetry using the method of moments and obtained the asymptotic distributions of his moment estimates and the corresponding eigenvalues using the perturbation method; these large-sample results imply $\sqrt{n}$-consistency of his moment estimates. However, for rotationally symmetric distributions with center of mass at the origin, his method is applicable only to antipodally symmetric distributions and, hence, the zero-mean model given in Example 1.1 eludes his theory. On the other hand, for an antipodally symmetric model, the success of Kim's moment method of estimating the axis of symmetry depends on the fact that the sample mean projector does not converge to the identity matrix and, hence, his method is not applicable to the identity mean projector model in Example 1.2.

The question, then, is: Are there any estimates that are free of the erratic large-sample behavior of the above-mentioned frequentist's estimates under the minimum condition that the rotationally symmetric model is identifiable? Recently, Watson (1983b) suggested the posterior means (with respect to the uniform prior) as estimates of the direction and the projector defining the axis of symmetry. He showed that these estimates are equivariant and are optimal in the sense that they are Pitman estimates with respect to the group of rotations. Other small-sample optimalities (minimaxity and admissibility) of his estimates are also obtained, whereas large-sample properties remain unknown. This paper is the result of an investigation into the question of whether these Pitman estimates, as well as other Bayes procedures with respect to smooth priors, are consistent for identifiable rotationally symmetric models. The answer is affirmative; our study reveals the strong consistency of Bayes procedures for all parameter values in this model and find that identifiability is indeed a necessary and sufficient condition for consistency.

The usual methods for proving consistency of Bayes procedures for all parameter points [Le Cam (1953), Schwartz (1965), and Strasser (1981)] typically require the existence of densities and do not apply since the model considered here is not assumed to be dominated, nor do the observations need to be independent. Instead, the consistency phenomenon exemplifies a rule on invariant statistical models mentioned by Lo (1984): Subject to identifiability, Bayes procedures are consistent for almost all parameter points [Doob (1949)] and, hence, consistency has to hold for at least one parameter point. By invariance, it must then hold for all parameter points.

Section 2 defines a direction of symmetry and gives the notation and some technical preliminaries. In Section 3, we show the equivalence of identifiability in the direction of symmetry, the consistency of the posterior distributions and the consistency of the Bayes estimates (posterior means) for smooth priors; we then give sufficient conditions for identifiability in the mean direction. If the model is identifiable in terms of the axis of rotational symmetry instead of the direction of symmetry, similar results for the consistency of Bayes procedures for the projector defining the axis of rotational symmetry are obtained in Section 4.

2. Notation and preliminaries. Let $\Omega = \{x\colon x^*x = 1,\ x \in R^q\}$ be the unit sphere in R^q and $\mathscr{B}$ the Borel σ-field on Ω; x^* is the transpose of x. Denote the group of real orthogonal transformations from R^q to R^q by $\mathbf{H}$, and note that $H \in \mathbf{H}$ if and only if $H^*H = I$, where I is the identity matrix. Denote points on Ω by x, y, μ, v and λ. For any n, let $H\mathbf{x} = (Hx_1, \dots, Hx_n)$ for $\mathbf{x} \in \Omega^n$ and $HA = \{H\mathbf{x}\colon \mathbf{x} \in A\}$ for $A \in \mathscr{B}^n$. Let $X_1, \dots, X_n, \dots$ be a sequence of random vectors taking values on Ω with joint distribution P_μ, where $\mu \in \Omega$ is a direction of symmetry in the following sense: For each n, the joint distribution of $\mathbf{X} = (X_1, \dots, X_n)$ given μ, denoted by $P_\mu(d\mathbf{x})$, satisfies

$$(2.1) \qquad P_\mu(A) = P_{H\mu}(HA), \quad \text{for all } A \in \mathscr{B}^n, \text{ all } H \in \mathbf{H} \text{ and all } \mu \in \Omega.$$

For each $A \in \mathscr{B}^n$, $P_\mu(A)$ is assumed to be $\mathscr{B}$-measurable in μ. Note that (2.1) is equivalent to

$$(2.1') \qquad \mathbf{X} \sim P_\mu(d\mathbf{x}), \quad \text{if and only if } \mathbf{Y} = H\mathbf{X} \sim P_{H\mu}(d\mathbf{y}) \text{ for all } H \in \mathbf{H}.$$

Let $\mu \sim \pi_q(d\mu)$, where π_q is the uniform distribution on Ω. Define the marginal probability Q on $\mathscr{B}^n$ by $Q(A) = \int_\Omega P_\mu(A)\pi_q(d\mu)$ for all $A \in \mathscr{B}^n$. Q satisfies $Q(HA) = Q(A)$ for all $H \in \mathbf{H}$ and all $A \in \mathscr{B}^n$. Let $P^n(B|\mathbf{x})$ be a function defined by $\mathscr{B} \otimes \Omega^k$ such that

$$(2.2) \qquad \text{for each } \mathbf{x},\ P^n(\cdot\,|\mathbf{x}) \text{ is a probability on } \mathscr{B},$$

$$(2.3) \qquad \text{for each } B,\ P^n(B|\,\cdot) \text{ is } \mathscr{B}^n\text{-measurable},$$

$$(2.4) \qquad \iint h(\mathbf{x}, \mu)P_\mu(d\mathbf{x})\pi_q(d\mu) = \iint h(\mathbf{x}, \mu)P^n(d\mu|\mathbf{x})Q(d\mathbf{x})$$

for all nonnegative $\mathscr{B}^{n+1}$-measurable functions h

and

$$(2.5) \qquad P^n(HB|H\mathbf{x}) = P^n(B|\mathbf{x}), \quad \text{for all } B \in \mathscr{B}, \text{ all } H \in \mathbf{H} \text{ and all } \mathbf{x} \in \Omega^n.$$

A P^n satisfying (2.2), (2.3) and (2.4) is called a posterior distribution of μ given $\mathbf{X} = \mathbf{x}$ with respect to the uniform prior. It is called a Pitman distribution if, in addition, it satisfies (2.5). The existence of a Pitman distribution has been proved in great generality by Le Cam (1972), Proposition 10, using the lifting theorem. The choice of a posterior distribution, which is Pitman, is crucial in generating consistent Bayes procedures for all parameter points. In fact, Example 3.1 in Section 3 illustrates a case where a posterior distribution which is not Pitman fails to be consistent at one parameter point.

In the rest of this section, we specialize the above model to the i.i.d. and dominated case and derive the posterior distribution given by Watson (1983b). For a dominated family of rotationally symmetric distributions, it is customary to compute the model density with respect to the uniform distribution [Watson (1983a, b)]. The following lemma provides a justification of this computation [a similar phenomenon for a location family on the line has been noted by Ferguson (1962)].

LEMMA 2.1. *If $P_\mu(dx_1)$ is dominated by some σ-finite measure $\alpha(dx_1)$ for all μ, then it is dominated by $\pi_q(dx_1)$ for all μ.*

PROOF. Since α is σ-finite, there is a finite measure equivalent to α. Thus, it suffices to prove the lemma for a finite α. Assume then α is a finite measure. Suppose there is a $B \in \mathscr{B}$ such that $\pi_q(B) = 0$, but yet $P_e(B) > 0$ for some $e \in \Omega$. Then, $P_{He}(HB) > 0$ for all $H \in \mathbf{H}$, implying $\alpha(HB) > 0$ for all $H \in \mathbf{H}$.

On the other hand, denoting the Haar probability on the orthogonal group by $\partial(dH)$, we have

$$\begin{aligned}\int_{\mathbf{H}} \alpha(HB)\, \partial(dH) &= \int_{\mathbf{H}}\int_{\Omega} I_{[x \in HB]}\alpha(dx)\, \partial(dH)\\ &= \int_{\Omega}\int_{\mathbf{H}} I_{[Hx \in B]}\, \partial(dH)\, \alpha(dx),\end{aligned} \tag{2.6}$$

by Fubini's theorem and the H-invariance of $\partial(dH)$. Denote the inner integral $\int_{\mathbf{H}} I_{[Hx \in B]}\, \partial(dH)$ by $\pi^x(B)$. Notice that for each x, $\pi^x(HB) = \pi^x(B)$ for all H. Thus, π^x is equal to π_q since π_q is the unique H-invariant probability on Ω. Hence, $\int_H \alpha(HB)\, \partial(dH) = \pi_q(B)\int_\Omega \alpha(dx)$. Clearly, $\pi_q(B) = 0$ entails $\alpha(HB) = 0$ almost surely $[\partial]$, contradicting $\alpha(HB) > 0$ for all $H \in \mathbf{H}$. □

Suppose then $P_\mu(dx_1)$ is dominated by a σ-finite measure for all μ. Denote the density of $P_\mu(dx_1)$ with respect to $\pi_q(dx_1)$ by $g(x_1, \mu)$. Furthermore, $g(x_1, \mu)$ is measurable with respect to (x_1, μ). Note that for each $\mu \in \Omega$ and each $H \in \mathbf{H}$, $g(x, \omega) = g(Hx, H\mu)$ almost surely $[\pi_q]$ [see page 254 in Eaton (1983)]. Routine arguments [Lehmann (1959), page 225] show that there exists a version $g(x|\mu)$ of $g(x, \mu)$ such that $g(x|\mu) = g(Hx|H\mu)$ for all H, all x and all μ, i.e., $g(\cdot|\mu)$ is rotationally symmetric with respect to a direction specified by μ. Hence, $g(x|\mu) = f(x^*\mu)$ for some nonnegative function f on the line [Saw (1978) and Watson (1983a)].

The natural posterior distribution of μ given $\mathbf{X} = \mathbf{x}$ is defined by

$$P^n(B|\mathbf{x}) = \frac{\int_B \Pi f(x_i^* \mu)\pi_q(d\mu)}{\int_\Omega \Pi f(x_i^* \mu)\pi_q(d\mu)}. \tag{2.7}$$

Clearly, P^n satisfies (2.5) and, hence, is a Pitman distribution. The posterior mean $\int \mu P^n(d\mu|\mathbf{x})$ is the Pitman estimate considered by Watson (1983b).

3. The identifiable direction of symmetry. We say that the model is identifiable in μ if μ can be recovered (i.e., computed measurably) from the data $X_1, \dots, X_n, \dots$. Formally, this property is defined by $\mu \in \mathscr{F}_\infty$ = P-completion of the σ-field generated by $X_1, \dots, X_n, \dots$ and P is the joint distribution of $\omega = (\mu, X_1, \dots, X_n, \dots)$ defined in Section 2. The main result of this section (Theorem 3.1) is that $\mu \in \mathscr{F}_\infty$ is a necessary and sufficient condition for establishing consistency of Bayes procedures of μ for all μ and is also equivalent to the existence of a sequence of weakly consistent estimates of μ. In the i.i.d. situation, $\mu \in \mathscr{F}_\infty$ is also equivalent to the classical identifiability condition (Proposition 3.2). First assume a prior $\pi(d\mu)$ on Ω which is dominated by $\pi_q(d\mu)$ with density $\pi'(\mu)$. A posterior distribution of μ with respect to this prior can be defined in terms of $\pi'(\mu)$ and $P^n(d\mu|\mathbf{X})$ as

$$\pi^n(B|\mathbf{X}) = \frac{\int_B \pi'(\mu)P^n(d\mu|\mathbf{X})}{\int_\Omega \pi'(\mu)P^n(d\mu|\mathbf{X})}. \tag{3.1}$$

Thus, for priors differentiable with respect to the uniform distribution the posterior distribution can be represented as a ratio. This representation is important: It allows us to study the large-sample behaviors of the numerator and the denominator separately, each of which depends only on the large-sample behavior of the Pitman distribution and the smoothness of the prior density. In this sense, the study of the posteriors is reduced to that of the Pitman distribution. The next theorem gives the main result. Denote the posterior mean $\int_\Omega \mu \pi^n(d\mu|\mathbf{X})$ by $m_\pi(\mathbf{X})$ and let $m(\mathbf{X})$ be the Pitman estimate $\int_\Omega \mu P^n(d\mu|\mathbf{X})$. Denote a point mass probability at λ by δ_λ. We say that μ is estimable if for each n there is a Borel function (estimate) T_n of $X_1, \dots, X_n$ such that for each μ, $T_n \to \mu$ in P_μ-probability. The main result of this paper (Theorem 3.1) is that the following six statements are essentially equivalent:

(S1) μ is estimable.
(S2) $\mu \in \mathscr{F}_\infty$.
(S3) For each λ, $P_\lambda\{P^n(\cdot|\mathbf{X}) \Rightarrow \delta_\lambda\} = 1$.
(S4) For each λ, $P_\lambda\{m(\mathbf{X}) \to \lambda\} = 1$.
(S5) For each λ, $P_\lambda\{\pi^n(\cdot|\mathbf{X}) \Rightarrow \delta_\lambda\} = 1$.
(S6) for each λ, $P_\lambda\{m_\pi(\mathbf{X}) \to \lambda\} = 1$.

Let us comment briefly on these statements. (S1) states that there exists a weakly consistent estimate of the mean direction whereas (S2) is the key identifiability condition. (S4) states that the Pitman estimate of Watson (1983b) is a strongly consistent estimate of the mean direction and (S3) depicts the

Pitman distribution merging to a degenerate distribution at the mean direction. (S5) and (S6) give analogous convergence statements for the posterior distribution and posterior mean with respect to the prior distribution π.

THEOREM 3.1. *Statements* (S1)—(S4) *are equivalent and, if* π' *is continuous and positive,* (S1)-(S6) *are equivalent.*

PROOF. Assume (S1). Let $\{T_n\}$ be a sequence of weakly consistent estimates of μ. There is a subsequence $n(k)\uparrow\infty$ such that $T_{n(k)}\to\mu$ almost surely $[P_\mu]$. By Fubini's theorem, $P\{\omega: T_{n(k)}\to\mu\}=1$. Since $T_{n(k)}\in\mathscr{F}\{X_1,\dots,X_{n(k)}\}$, $\mu\in\mathscr{F}_\infty$.

Next, assume (S2). By the forward martingale convergence theorem [Doob (1953)], the posterior characteristic function of μ given $X_1,\dots,X_n$, defined by $E[e^{is^*\mu}|X_1,\dots,X_n]$, converges to $E[e^{is^*\mu}|X_1,X_2,\dots]$ almost surely $[P]$. Since $\mu\in\mathscr{F}_\infty$,

$$(3.2)\qquad E\left[e^{is^*\mu}|X_1,\dots,X_n\right]\to e^{is^*\mu}\quad\text{almost surely }[P].$$

An application of Fubini's theorem entails

$$(3.3)\qquad P_\mu\left\{E\left[e^{is^*\mu}|X_1,\dots,X_n\right]\to e^{is^*\mu}\right\}=1\quad\text{almost surely }[\pi_q].$$

Hence, there is an v such that

$$(3.4)\qquad P_v\left\{E\left[e^{is^*\mu}|X_1,\dots,X_n\right]\to e^{is^*v}\right\}=1.$$

By (2.5), for all $H\in\mathbf{H}$

$$(3.5)\qquad P_v\left\{E\left[e^{is^*H^*\mu}|HX_1,\dots,HX_n\right]\to e^{is^*v}\right\}=1.$$

For each λ, let H be an orthogonal matrix such that $Hv=\lambda$. Since $\mathbf{X}\sim P_v$ implies $\mathbf{Y}=H\mathbf{X}\sim P_{Hv}=P_\lambda$, we conclude that for each $s\in R^q$

$$(3.6)\qquad P_\lambda\left\{E\left[e^{i(Hs)^*\mu}|X_1,\dots,X_n\right]\to e^{i(Hs)^*\lambda}\right\}=1.$$

Hence, $P_\lambda\{P^n(\cdot|\mathbf{X})\Rightarrow\delta_\lambda\}=1$, proving (S3). Statement (S4) follows from (S3) since the identity map on Ω is bounded and continuous. If (S4) is assumed, we can let $T_n=m(\mathbf{X})$ and hence (S1).

Suppose π' is continuous and positive. The posterior characteristic function of μ given $X_1,\dots,X_n$ with respect to the prior $\pi(d\mu)$ is given by

$$(3.7)\qquad E_\pi\left[e^{is^*\mu}|X_1,\dots,X_n\right]=\frac{\int_\Omega e^{is^*\mu}\pi'(\mu)P^n(d\mu|\mathbf{X})}{\int_\Omega\pi'(\mu)P^n(d\mu|\mathbf{X})}.$$

Assume (S3). For each λ the numerator converges almost surely $[P_\lambda]$ to $e^{is^*\lambda}\pi'(\lambda)$ whereas the denominator converges almost surely $[P_\lambda]$ to $\pi'(\lambda)>0$. Hence, $E_\pi[e^{is^*\mu}|X_1,\dots,X_n]\to e^{is^*\lambda}$ almost surely $[P_\lambda]$, proving (S5). Again, statement (S6) follows from (S5) since the identity map on Ω is bounded and continuous and, if (S6) holds, Fubini's theorem implies that $P\{m_\pi(\mathbf{X})\to\mu\}=1$ and, hence, $\mu\in\mathscr{F}_\infty$. □

According to Theorem 3.1, for a smooth prior if the posterior mean does not converge to the "true" direction of symmetry λ (or if the posterior distribution does not converge to a point mass at λ), no weakly consistent estimate of λ exists. This phenomenon may not be true if the posterior distribution is not generated from the Pitman distribution as in (3.1). We illustrate this phenomenon by the following example which also demonstrates what may happen if the invariant condition (2.5) is violated.

EXAMPLE 3.1. Assume a uniform prior. A posterior distribution generated from a Pitman distribution as in (3.1) will itself be a Pitman distribution and, hence, enjoys nice large-sample properties according to Theorem 3.1. On the other hand, a posterior distribution which is not a Pitman distribution can be inconsistent at a particular parameter point. Take for example, P_μ, a point mass at $(\mu, \mu, \dots)$. Apparently, $P_\mu\{n^{-1}\Sigma X_1 = X_1 = \mu \text{ for all } n\} = 1$ for each μ and, hence, μ is estimable (by X_1) and $\mu \in \mathscr{F}_\infty$. For $X_1 \neq (1; 0, \dots, 0)$, let $P^n(\cdot|\mathbf{X})$ be a point mass at X_1; otherwise, let it be a point mass at $(0, \dots, 0, 1)$. Clearly, $P^n(\cdot|\mathbf{X})$ fails to converge to a point mass at λ for $\lambda = (1, 0, \dots, 0)$.

The condition to check for the validity of Theorem 3.1 is $\mu \in \mathscr{F}_\infty$ or, perhaps, (S1). One easy sufficient condition is the following:

(A) There is a unit vector e such that $n^{-1}\Sigma X_i \to te$ in P_e-probability for some nonzero $t \in [-1, 1]$.

Condition (A) is far from being necessary for consistency; it is not applicable to any zero-mean model (Example 1.1). Yet, it has the advantage of being easily applicable to dependent observations in the presence of moment inequalities. Condition (A) is equivalent to the fact that the sequence $X_1, \dots, X_n, \dots$ obeys a weak law of large numbers in the sense that $n^{-1}\Sigma X_i \to C$ in P_e-probability, where e is some unit vector and C is some nonzero constant. Indeed, in this case $HC = C$ for all $H \in \mathbf{H}$ such that $He = e$. Hence, C must be equal to te for some nonzero $t \in [-1, 1]$.

PROPOSITION 3.1. *Assume* (A). *Then* $\mu \in \mathscr{F}_\infty$.

PROOF. Assume (A) and let $T_n = n^{-1}\Sigma X_i$. There is a subsequence $n(k) \uparrow \infty$ such that $T_{n(k)} \to te$ almost surely $[P_e]$. In view of (2.1), for each μ, $T_{n(k)} \to t\mu$ almost surely $[P_\mu]$, where $He = \mu$. By Fubini's theorem, $P\{\omega\colon T_{n(k)} \to t\mu\} = 1$. Since $T_{n(k)} \in \mathscr{F}\{X_1, \dots, X_{n(k)}\}$ and $t \neq 0$, $\mu \in \mathscr{F}_\infty$. □

Theorem 3.1 states that Bayes procedures enjoy nice large-sample properties if one presupposes the existence of a consistent estimate of μ. However, in the case that we do not know of the existence of a consistent estimate of μ, this theorem is useless (see, however, Proposition 3.3 which follows). This difficulty does not arise in the usual i.i.d. case. In fact, it will be shown in Proposition 3.2 that in the case of i.i.d. sampling, $\mu \in \mathscr{F}_\infty$ is equivalent to the following classical identifiability condition:

(B) $X_1, \dots, X_n, \dots$ are i.i.d. F_μ and $\mu \neq v$ implies that there is a $B \in \mathscr{B}$ such that $F_\mu(B) \neq F_v(B)$, where F_μ denotes the distribution of X_1.

Condition (B) states that different parameters correspond to different model distributions. By the law of large numbers, it also implies that $n^{-1}\Sigma X_i \to te$ for some $t \in [-1, 1]$, yet t can be zero. Hence, (A) is not more general than (B).

PROPOSITION 3.2. *Assume that* $X_1, \dots, X_n, \dots$ *are i.i.d.* F_μ. *Then* (B) *is equivalent to* $\mu \in \mathscr{F}_\infty$.

PROOF. Assume (B). Then $\mu \in \mathscr{F}_\infty$ follows from the arguments of Doob (1949) [see also Schwartz (1965) and Breiman, Le Cam and Schwartz (1964)].

Conversely, $\mu \in \mathscr{F}_\infty$ implies that (S3) holds by Theorem 3.1. That is, $P_\mu\{m(\mathbf{X}) \to \mu\} = 1$ and $P_\lambda\{m(\mathbf{X}) \to \lambda\} = 1$. If $\mu \neq \lambda$, P_μ and P_λ are then singular on $\mathscr{F}\{X_1, \dots, X_n, \dots\}$. The proof is completed by noting that for i.i.d. sampling, P_μ and P_λ are singular on $\mathscr{F}\{X_1, \dots, X_n, \dots\}$ for $\mu \neq \lambda$ is equivalent to (B). □

The following result is an immediate consequence of this proposition and Theorem 3.1.

THEOREM 3.2. *Assume that* $X_1, \dots, X_n, \dots$ *are i.i.d.* F_μ. *Then* (B), (S1)–(S4) *are equivalent and, if the prior density* π' *is positive and continuous,* (*B*), (*S*1)–(*S*6) *are all equivalent.*

Theorem 3.2 can be applied to the zero-mean model in Example 1.1. In particular, the Bayes estimate (posterior mean) of μ in the mixture model is consistent.

REMARK 3.1. For i.i.d. sampling, the classical identifiability condition (B) holds if and only if $\{P_\mu\}$ is a pairwise orthogonal family (i.e., $\mu \neq \lambda$ implies P_μ and P_λ are orthogonal); hence, pairwise orthogonality of the joint distributons is a necessary and sufficient condition for $\mu \in \mathscr{F}_\infty$. For dependent observations, characterizations of $\mu \in \mathscr{F}_\infty$ in terms of orthogonal properties of the joint distributions P_μ can be obtained by appealing to some recent work of Mauldin, Preiss and Weizsacker (1983) on orthogonal Markov kernels. The family $\{P_\mu\}$ is orthogonality preserving, if for any pair of mutually orthogonal prior distributions π_1 and π_2, the two marginal distributions $\int P_\mu \pi_1(d\mu)$ and $\int P_\mu \pi_2(d\mu)$ are also mutually orthogonal. The family $\{P_\mu\}$ is completely orthogonal if there is a Borel set B of $(\mu, \mathbf{x})$, i.e., $B \in \mathscr{B} \otimes \mathscr{B}^\infty$, such that for each μ, $P_\mu(B_\mu) = 1$ and if $\mu \neq \lambda$, then $B_\mu \cap B_\lambda = \varnothing$ (B_μ denotes the μ-section of B). The family $\{P_\mu\}$ is completely orthogonal implies that it is orthogonality preserving [Mauldin, Preiss and Weizsacker (1983), Theorem 1.8]; the latter implies $\mu \in \mathscr{F}_\infty$ [Mauldin, Preiss and Weizsacker (1983), Theorem 4.1]. Moreover for a rotationally symmetric model, our Theorem 3.1 and Theorem 1.8 in Mauldin, Preiss and Weizsacker (1983) can be applied to prove that $\mu \in \mathscr{F}_\infty$ implies that the family is completely orthogonal. We collect these results in the following proposition:

PROPOSITION 3.3. *The family* $\{P_\mu\}$ *is orthogonality preserving if and only if it is completely orthogonal if and only if* $\mu \in \mathscr{F}_\infty$.

4. The identifiable projector defining the axis of symmetry. In this section we assume the model specified by (2.1) is also antipodally symmetric in the sense that for each n,

$$P_\mu(A) = P_\mu(-A), \quad \text{for all } A \in \mathscr{B}^n \text{ and all } \mu \in \Omega. \tag{4.1}$$

Roughly speaking, (4.1) means the direction of symmetry defined by μ and that defined by $-\mu$ are the same, i.e., μ and $-\mu$ cannot be distinguished. In this case, assumption (A) fails to hold, because, if $n^{-1}\Sigma X_i$ converges, the limit must be the zero vector. In the i.i.d. case, condition (B) is not satisfied since both μ and $-\mu$ correspond to the same distribution. In fact, the next result states that weakly consistent estimates of μ do not exist for an antipodally symmetric model on the sphere.

PROPOSITION 4.1. *Assume* (4.1). *μ is not measurable with respect to $\mathscr{F}_\infty$.*

PROOF. The proof is based on a contrapositive argument. Suppose $\mu \in \mathscr{F}_\infty$. By Theorem 3.1, (S4) holds. Then there is a $\mu \in \Omega$ such that $P_\mu\{m(\mathbf{X}) \to \mu\} = 1$. Note that $m(H\mathbf{X}) = Hm(\mathbf{X})$ for all $H \in \mathbf{H}$ and, in particular, $m(-\mathbf{X}) = -m(\mathbf{X})$. Hence, $P_\mu\{m(-\mathbf{X}) \to -\mu\} = 1$. According to (4.1), $P_\mu\{m(\mathbf{X}) \to -\mu\} = 1$, implying $\mu = -\mu$, i.e., $\mu = (0,\dots,0)$ contradicting $\mu \in \Omega$. Hence, μ is not measurable with respect to $\mathscr{F}_\infty$. □

Even though one cannot recover μ from the data, the data may still provide information about some functions of μ. A natural function to look at is the axis going through μ and $-\mu$, or equivalently, the projector $\mu\mu^*$ into the direction of μ. In what follows, we discuss consistent Bayes procedures for the projector $\mu\mu^*$ defining the axis of μ [Watson (1983a, b)]. A necessary and sufficient condition in establishing consistency then is $\mu\mu^* \in \mathscr{F}_\infty$, i.e., the model is identifiable in $\mu\mu^*$. As in the last section, this identifiability is guaranteed by one of the following two conditions:

(A′) There is a unit vector e such that $n^{-1}\Sigma X_i X_i^* \to ee^*$ in P_e-probability.

(B′) $X_1,\dots,X_n,\dots$ are i.i.d. F_μ and $\mu\mu^* \neq vv^*$ implies that there is a $B \in \mathscr{B}$ such that $F_\mu(B) \neq F_v(B)$.

PROPOSITION 4.2.

(i) (A′) *implies* $\mu\mu^* \in \mathscr{F}_\infty$.

(ii) *If* $X_1,\dots,X_n,\dots$ *are i.i.d.* P_μ, *then* (B′) *and* $\mu\mu^* \in \mathscr{F}_\infty$ *are equivalent.*

PROOF. Assume (A′) and let $T_n = n^{-1}\Sigma X_i X_i^*$. There is a subsequence $n(k)\uparrow\infty$ such that $T_{n(k)} \to ee^*$ almost surely $[P_e]$. In view of (2.1), for each μ, $T_{n(k)} \to Hee^*H^* = \mu\mu^*$ almost surely $[P_\mu]$, where $He = \mu$. By Fubini's theorem, $P\{\omega\colon T_{n(k)} \to \mu\mu^*\} = 1$. Since $T_{n(k)} \in \mathscr{F}\{X_1,\dots,X_{n(k)}\}$, $\mu\mu^* \in \mathscr{F}_\infty$.

Assume (B′). Then $F_\mu(B)$ is a function of $\mu\mu^*$; it then follows from Theorem 3 in Blackwell (1956) that for each B, $F_\mu(B)$ is a Borel function of $\mu\mu^*$. Hence, the arguments of Doob (1949) can be applied to conclude $\mu\mu^* \in \mathscr{F}_\infty$. The arguments of Proposition 3.2 can also be applied to prove the converse (with Theorem 4.1, to follow, instead of Theorem 3.1). □

Condition (A′) is not difficult to check for dependent observations. In general, $n^{-1}\Sigma X_i X_i^*$ may not converge and, if it converges, the limit can be a projector not depending on μ (see Example 2.1). This last phenomenon is particularly irritating since in this case and for i.i.d. observations, neither the eigenvector method of Watson (1983a), Chapter 5, or the moment method of Kim (1978) is applicable; it is comforting that in this situation the Bayes procedures remain consistent, subject only to the (necessary and sufficient) identifiability condition.

The next result gives the analogy of Theorem 3.1 for the projector $\mu\mu^*$ defining the axis of rotational symmetry. If the posterior distribution of $\mu|\mathbf{X}$ is $P^n(\cdot|\mathbf{X})$, we denote the posterior distribution of $\mu\mu^*|\mathbf{X}$ by $\mathscr{L}(\cdot|\mathbf{X})$. Define $\mathscr{L}_\pi(\cdot|\mathbf{X})$ similarly. Since π is absolutely continuous with respect to π_q, its image under the map $\mu \to \mu\mu^*$ is also absolutely continuous with respect to that of π_q. Denote the density of the image measure by π''. Consider the following six statements that correspond to those for Theorem 3.1:

(S1′) $\mu\mu^*$ is estimable.
(S2′) $\mu\mu^* \in \mathscr{F}_\infty$.
(S3′) For each λ, $P_\lambda\{\mathscr{L}(\cdot|\mathbf{X}) \Rightarrow \delta_{\lambda\lambda^*}\} = 1$.
(S4′) For each λ, $P_\lambda\{E[\mu\mu^*|\mathbf{X}] \to \lambda\lambda^*\} = 1$.
(S5′) For each λ, $P_\lambda\{\mathscr{L}_\pi(\cdot|\mathbf{X}) \Rightarrow \delta_{\lambda\lambda^*}\} = 1$.
(S6′) For each λ, $P_\lambda\{E_\pi[\mu\mu^*|\mathbf{X}] \to \lambda\lambda^*\} = 1$.

THEOREM 4.1.

(i) *Statements* (S1′)–(S4′) *are equivalent and, if π'' is continuous and positive,* (S1′)–(S6′) *are equivalent.*

(ii) *If the observations are i.i.d.,* (B*), (S1*)–(S4′) *are equivalent and, for a continuous and positive π'',* (B′), (S1*)–(S6′) *are all equivalent.*

PROOF. Statement (S1′) implies (S2′) is clear from the proof of the analogous part in Theorem 3.1. Assume (S2′). The posterior characteristic function of $\mu\mu^*$ given $X_1, \dots, X_n$, defined by $E[\operatorname{etr}(iM\mu\mu^*)|X_1, \dots, X_n]$ for any symmetric matrix M, converges to $\operatorname{etr}(iM\mu\mu^*)$ almost surely $[P]$. Hence, there is an v such that

$$P_v\{E[\operatorname{etr}(iM\mu\mu^*)|X_1, \dots, X_n] \to \operatorname{etr}(iMvv^*)\} = 1. \tag{4.2}$$

By (2.5), for all $H \in \mathbf{H}$,

$$P_v\{E[\operatorname{etr}(iMH^*\mu\mu^*H)|HX_1, \dots, HX_n] \to \operatorname{etr}(iMvv^*)\} = 1. \tag{4.3}$$

For each λ, let H be an orthogonal matrix such that $Hv = \lambda$. Then (4.3) becomes

$$P_\lambda\{E[\operatorname{etr}(iMH^*\mu\mu^*H)|X_1, \dots, X_n] \to \operatorname{etr}(iMH^*\lambda\lambda^*H)\} = 1. \tag{4.4}$$

That is, for each λ there is an orthogonal matrix H such that for all symmetric matrices M,

$$(4.5) \qquad P_\lambda\{E[\operatorname{etr}(iHMH^*\mu\mu^*)|X_1,\dots,X_n] \to \operatorname{etr}(iHMH^*\lambda\lambda^*)\} = 1.$$

Denote HMH^* by N. Note that N is a symmetric matrix and the map $M \to N$ is one-to-one. This concludes the proof of (S3′).

The proofs for other implications in (i) are similar to that of Theorem 3.1 with obvious modifications and will not be given. Part (ii) is a direct consequence of (i) and Proposition 4.2. □

Theorem 4.1 can be applied to the mixture model in Example 1.2 to conclude that the Bayes procedures for the mean projector are consistent.

REMARK 4.1. Watson (1983b) also suggested estimating a projector Π projecting into an s-dimensional subspace of R^q by the Pitman estimate with respect to the uniform distribution on the Grassmann manifold [James (1954) and Eaton (1983)]. By selecting an invariant posterior distribution, i.e., a Pitman distribution satisfying an analogy of (2.5), one also obtains consistent Bayes procedures for all Π (subject to the identifiability condition). That is, Theorem 4.1 remains valid in this case. The proof is the same and will not be reproduced.

REFERENCES

BLACKWELL, D. (1956). On a class of probability spaces. *Proc. Third Berkeley Symp. Math. Statist. Probab.* **2** 1–6. Univ. of California Press.

BREIMAN, L., LE CAM, L. and SCHWARTZ, L. (1964). Consistent estimates and zero-one sets. *Ann. Math. Statist.* **35** 157–161.

DOOB, J. (1949). An application of the theory of martingales. *Colloq. Internat. du CNRS, Paris* 23–27.

DOOB, J. (1953). *Stochastic Processes*. Wiley, New York.

EATON, M. L. (1983). *Multivariate Statistics*. Wiley, New York.

FERGUSON, T. S. (1962). Location and scale parameters in exponential families of distributions. *Ann. Math. Statist.* **33** 986–1001.

JAMES, A. T. (1954). Normal multivariate analysis and the orthogonal group. *Ann. Math. Statist.* **25** 40–75.

KIM, K. M. (1978). Orientation-shift model on the sphere. Ph.D. thesis. Dept. of Statistics, Univ. of California, Berkeley.

LE CAM, L. (1953). On some asymptotic properties of maximum likelihood estimates and related Bayes estimates. *Univ. California Publ. Statist.* **1** 277–330.

LE CAM, L. (1972). Limits of experiments. *Proc. Sixth Berkeley Symp. Math. Statist. Probab.* **1** 245–261. Univ. of California Press.

LEHMANN, E. L. (1959). *Testing Statistical Hypotheses*. Wiley, New York.

LO, A. Y. (1984). Consistency in the location model: The undominated case. *Ann. Statist.* **12** 1584–1587.

MAULDIN, R. D., PREISS, D. and WEIZSACKER, H. V. (1983). Orthogonal transition kernels. *Ann. Probab.* **11** 970–988.

SAW, J. G. (1978). A family of distributions on the m-sphere and some hypothesis tests. *Biometrika* **65** 69–73.

SCHWARTZ, L. (1965). On Bayes procedures. *Z. Wahrsch. verw. Gebiete* **4** 10–26.

STRASSER, H. (1981). Consistency of maximum likelihood and Bayes estimates. *Ann. Statist.* **9** 1107–1113.

WATSON, G. S. (1983a). *Statistics on Spheres*. Wiley, New York.

WATSON, G. S. (1983b). Optimal and robust estimation on the sphere. *Bull. Internat. Statist. Inst.* **44** 816–818.

DEPARTMENT OF STATISTICS
STATE UNIVERSITY OF NEW YORK
BUFFALO, NEW YORK 14260

DEPARTMENT OF STATISTICS
HILL CENTER FOR THE MATHEMATICAL SCIENCES
RUTGERS UNIVERSITY
NEW BRUNSWICK, NEW JERSEY 08903