The Annals of Statistics
1995, Vol. 23, No. 1, 327–337

STATIONARY EXPONENTIAL FAMILIES

By I. H. Dinwoodie

Tulane University

A stationary exponential family is defined using transition densities which take the form of exponentiated symmetric k-linear forms on $\mathbf{R}^d$. Estimation is based on a mean value parametrization through a convex function on a finite-dimensional vector space. A consistency theorem and a central limit theorem are presented.

1. Introduction. We are concerned with defining and studying an exponential family for stationary sequences of random vectors $X_1, X_2, \ldots$ in the finite-dimensional vector space $F = \mathbf{R}^d$. The main results are a consistency theorem and a central limit theorem (Theorems 2.1 and 2.2) for an estimator $\hat{\theta}$ of the parameter θ which indexes the process. We begin with some definitions and a preliminary discussion of the exponential family, and we present the main results in Section 2 and examples in Section 3.

Let E be the vector space of symmetric k-linear forms on F^k. The dimension of E is $\binom{k+d-1}{d-1}$, since a basis corresponds to a nonnegative integer solution of $x_1 + \cdots + x_d = k$. Let μ be a reference probability measure on $\mathbf{R}^d$ such that $Z_\theta = \int \exp(\theta(x_1, \ldots, x_k))\mu^k(dx_1, \ldots, dx_k) < \infty$ precisely on an open set $\Theta \subset E$, $\theta \in \Theta$. The Borel field on $\mathbf{R}^d$ will be denoted B.

Construct a stationary process $\{X_i \in F,\ i \geq 1\}$ defined on the space F^∞ as follows. Fix $\theta \in \Theta$ and define

$$
\begin{aligned}
Z(x_1, \ldots, x_k) &= \exp(\theta(x_1, \ldots, x_k)),\\
Z(x_1, \ldots, x_i) &= \int \exp(\theta(x_1, \ldots, x_k))\mu^{k-i}(dx_k, dx_{k-1}, \ldots, dx_{i+1}), \qquad 1 \leq i \leq k, \qquad (1.1)\\
Z &= \int \exp(\theta(x_1, \ldots, x_k))\mu^k(dx_k, \ldots, dx_1).
\end{aligned}
$$

The Z will appear with a subscript θ occasionally to avoid ambiguity. Let the transition density p (p_θ occasionally) be given by

$$p(x_1, \ldots, x_k) = \frac{\exp(\theta(x_1, \ldots, x_k))}{Z(x_1, \ldots, x_{k-1})} = \frac{Z(x_1, \ldots, x_k)}{Z(x_1, \ldots, x_{k-1})}.$$

Let $\pi(x_1, \ldots, x_{k-1}, dx_k)$ denote the transition probability associated to the density p. The theorem of Ionescu Tulcea [see Ionescu Tulcea (1949) or

Received January 1992; revised December 1993.
AMS 1991 *subject classifications*. Primary 62M05; secondary 60G10.
Key words and phrases. Exponential family, stationary process, chain of finite order.

Shiryayev (1984)] assures us a measure $\mathbf{P}_\theta$ on (F^∞, B^∞) which satisfies

$$\begin{aligned}\mathbf{P}_\theta(A_1 \times \cdots \times A_n \times F \times \cdots) = &\int_{A_1} \frac{Z(x_1)}{Z} \mu(dx_1) \cdots \\ &\times \int_{A_{k-1}} \frac{Z(x_1, \dots, x_{k-1})}{Z(x_1, \dots, x_{k-2})} \mu(dx_{k-1}) \cdots \\ &\times \int_{A_k} \pi(x_1, \dots, x_{k-1}, dx_k) \cdots \\ &\times \int_{A_n} \pi(x_{n-k+1}, \dots, x_{n-1}, dx_n)\end{aligned}$$

for every positive integer n and all sets $A_i \in B$, $1 \le i \le n$. The coordinate process $X_1, X_2, \dots$ on $(F^\infty, B^\infty, \mathbf{P}_\theta)$ is a chain of finite order with transition probability π. One can verify that the process is strictly stationary and that the stationary measure on k coordinates for this process has density $\exp(\theta(x_1, \dots, x_k))/Z$.

Define the sequence of random elements $\{Y_n\}$ in the dual space E' of the vector space E by

$$Y_n(\alpha) = \alpha(X_n, \dots, X_{n+k-1})$$

for $\alpha \in E$. The vector Y_n is a k-fold tensor product defined via the symmetric k-linear forms.

Expectations and derivatives for the stationary exponential family of processes $\{\mathbf{P}_\theta\colon \theta \in \Theta\}$ are related as in the i.i.d. family, as we show in Proposition 1.1 below. For $\theta \in \Theta$, let $\mathbf{E}_\theta$ denote expectation with respect to the probability measure $\mathbf{P}_\theta$ on the measurable space F^∞. The symbol $\nabla \log Z$ denotes the derivative of the real-valued function Z, which is defined on E. Thus $\nabla \log Z\colon E \to E'$.

Proposition 1.1. $\mathbf{E}_{(\cdot)}(Y_n) = \nabla \log Z\colon \Theta \to E'$.

Proof. For each $\alpha \in E$,

$$\begin{aligned}\mathbf{E}_\theta(Y_n)(\alpha) &= \int \alpha(x_1, \dots, x_k) \frac{Z_\theta(x_1, \dots, x_k)}{Z_\theta} \mu^k(dx_1, \dots, dx_k) \\ &= \nabla_\theta \log Z(\alpha),\end{aligned}$$

which proves that $\mathbf{E}_\theta(Y_n) = \nabla_\theta \log Z$. □

In the following text we let $m_\theta = \mathbf{E}_\theta(Y_i) \in E'$. Consider now the problem of estimating the parameter θ. The ergodic theorem implies that the sequence $\overline{Y} = n^{-1}\Sigma^n Y_i \in E'$ has the property that $\overline{Y} \to \mathbf{E}_\theta(Y_i) = \nabla_\theta \log Z$. To simplify notation, let

$$f = \nabla \log Z\colon \Theta \to E'.$$

Proposition 1.1 suggests that we define our estimator $\hat{\theta}$ for the parameter $\theta \in E$ by

$$\hat{\theta} = f^{-1}(\overline{Y}) \tag{1.2}$$

since at least formally $\hat{\theta} = f^{-1}(\overline{Y}) \to f^{-1}(\mathbf{E}_\theta(Y_i)) = \theta$. If C is the closed convex hull of the support of the distribution of Y_i, then $\overline{Y} \in C$ and one needs to be sure that C is in the range of f and that f is bijective from E to C for (1.2) to make sense. These points will be addressed in the next section. We think of this estimation scheme very simply as choosing the parameter $\hat{\theta} \in E$ which makes the mean of Y_i under the law $\mathbf{P}_{\hat{\theta}}$ equal to the sample mean $\overline{Y} \in C$.

The results that follow in Section 2 justify the existence and continuity of f^{-1}, show asymptotic properties of $\overline{Y}$ and translate these properties to $\hat{\theta}$ through f^{-1}.

2. Two limit theorems.

Our study of $\hat{\theta}$ below begins with regularity properties of f^{-1} to obtain the consistency result Theorem 2.1. Then we apply a central limit theorem of Rosenblatt (1971) to the multivariate Markov process $(X_n, \dots, X_{n+k-1})$ in order to prove a central limit theorem for $\hat{\theta}$ at Theorem 2.2. Assume in what follows that:

$$\begin{aligned} &\text{(a) } \log Z_\theta \text{ is finite precisely for } \theta \in \Theta,\ \Theta \subset E \text{ open;} \\ &\text{(b) } Y_i \in \operatorname{int} C \quad \text{a.s.} \end{aligned} \tag{2.1}$$

The process can be constructed with less than (2.1a), but we use all of (2.1) to prove results about $\hat{\theta}$.

LEMMA 2.1. *Assume* (2.1). *Then the map* $f = \nabla \log Z\colon \Theta \subset E \to \operatorname{int} C$ *is bijective.*

PROOF. The function $\log Z$ is strictly convex on Θ by Theorem 7.1 of Barndorff-Nielsen (1978) and (2.1b). Now we need only apply Theorems 5.33 and 9.2 of Barndorff-Nielsen (1978). □

Lemma 2.1 makes sense of the estimation procedure (1.2), since now we have with assumption (2.1b) that $\hat{\theta}$ exists and is well defined with probability 1. Next we look for further regularity properties of f.

LEMMA 2.2. *Assume* (2.1). *Then* $f = \nabla \log Z$ *has differentiable inverse* f^{-1}: $\operatorname{int} C \to \Theta \subset E$.

PROOF. By Lemma 2.1, f^{-1} exists. To show it is differentiable, it is enough to show that f is differentiable without critical points. However, $D_\theta f$ is a linear map from E to E' corresponding in the standard way to a bilinear form on $E \times E$ which satisfies

$$D_\theta f(\alpha, \alpha) = \mathbf{E}_\theta\big(\langle \alpha, Y_1 - m_\theta \rangle^2\big)$$

for any $\alpha \in E$. $D_\theta f$ is, in fact, positive definite, since (2.1b) implies that the law of Y_i is not concentrated on an affine subspace of E'. Therefore, the matrix for $D_\theta f$ is invertible, and so f has no critical points. □

THEOREM 2.1. *Assume* (2.1). *Then the estimator* $\hat{\theta} \in E$ *converges a.s. to* θ.

PROOF. The map f^{-1} is continuous, so the formal argument following (1.2) is valid. □

We will prove below a central limit theorem for $\hat{\theta}$ from results of Rosenblatt (1971) for Markov chains. We start with the following observation.

LEMMA 2.3. *Assume* (2.1). *The process* $\{(X_n, \dots, X_{n+k-1}): n \geq 1\}$ *is a Markov chain on* F^k *with transition probabilities* π_k *given by*

$$\pi_k((x_0, \dots, x_{k-1}), (A_1, \dots, A_k))$$
$$= \delta_{(x_1, \dots, x_{k-1})}(A_1 \times \cdots \times A_{k-1}) \pi(x_1, \dots, x_{k-1}, A_k).$$

Let $L^{2,k}$ denote the set of real-valued measurable functions on the product space F^k such that

$$\|g\|_2^2 = \mathbf{E}_\theta(g^2(X_1, \dots, X_k)) < \infty.$$

This is the L^2 norm for the stationary probability measure on k-coordinates having density $Z(x_1, \dots, x_k)/Z$. Define the operator T on $L^{2,k}$ by

$$(Tg)(x_1, \dots, x_k) = \int \pi_\theta(x_2, \dots, x_k, dx_{k+1}) g(x_2, \dots, x_{k+1}).$$

We will say that $g \perp 1$ if $\mathbf{E}_\theta(g(X_1, \dots, X_k)) = 0$. Consider the following L^2 norm condition:

$$\sup_{g \perp 1} \frac{\|T^n g\|_2}{\|g\|_2} \to 0. \tag{2.2}$$

LEMMA 2.4. *Assume* (2.1) *and* (2.2). *Then the sequence* $\Sigma_1^n(Y_i - m_\theta)/\sqrt{n}$ *converges in distribution to the multivariate normal law* $\mathbf{N}(0, A)$, *where* A *is the symmetric bilinear form on* $E \times E$ *defined by*

$$A(\alpha, \alpha) = \lim \frac{1}{n} \mathbf{E}_\theta\left(\left(\sum_1^n \langle \alpha, Y_i - m_\theta \rangle\right)^2\right).$$

PROOF. Consider the usual operator for the multivariate Markov chain with the kernel π_k. Then (2.2) is the L^2 norm condition of Rosenblatt [(1971), page 206]. Apply Theorem 2 of Rosenblatt [(1971), page 217], to the sequence $S_n = \langle \alpha, \Sigma_1^n(Y_i - m_\theta) \rangle$ [where $f_{i,n} = (1/\sqrt{n})\alpha(X_i, \dots, X_{i+k-1})$ and $k_n = n$ in his notation] to see that $S_n/\sqrt{n} \to N(0, A(\alpha, \alpha))$. Then it is also true that for each $\alpha, \beta \in E$ the quantity

$$\frac{1}{n}\mathbf{E}_\theta\left(\sum_1^n \langle \alpha, Y_i - m_\theta \rangle \sum_1^n \langle \beta, Y_i - m_\theta \rangle\right)$$

converges to a number, say $A(\alpha, \beta) < \infty$, and A is necessarily symmetric and bilinear. Thus the covariance matrix A is of the stated form.

However, this implies that the sequence of vectors $\sum_1^n (Y_i - m_\theta)/\sqrt{n} \in E'$ converges in distribution to $\mathbf{N}(0, A)$. □

Consider the condition (2.2). If there exists $g \in L^1(\mu)$ such that

$$p(x_1, \dots, x_{k-1}, \cdot) \le g(\cdot), \tag{2.3}$$

then (2.2) and the above central limit theorem hold. The condition (2.3) gives uniform integrability of the kernels π_k, which is stronger than the Doeblin condition and has been used extensively for central limit theorems [Doob (1953)]. Rosenblatt [(1971), Theorem 1, page 211] shows that the Doeblin condition implies the L^2 norm condition, and generally (2.3) is easier to check. Condition (2.2) means in a precise way that the multivariate chain $(X_n, \dots, X_{n+k-1})$ is asymptotically uncorrelated [Rosenblatt (1971), page 207].

For $k = 2$, the process (X_n) is Markov with ordinary transition operator π. Now π is self-adjoint on the space L^2 of functions on F with the stationary distribution, since

$$\begin{aligned}\langle g, \pi h \rangle &= \int \mu(dx_1) Z(x_1)/Z\, g(x_1) \int \mu(dx_2) Z(x_1, x_2)/Z(x_1) h(x_2) \\ &= \int \mu(dx_2) Z(x_2)/Z\, h(x_2) \int \mu(dx_1) Z(x_1, x_2)/Z(x_2) g(x_1) \\ &= \langle \pi g, h \rangle\end{aligned}$$

using the symmetry of $Z(\cdot, \cdot)$. Note that $T^{n+1}g(x_1, x_2) = \pi^n T g(x_2)$, and if $g \perp 1$, then $h = Tg$ is orthogonal to 1 in L^2 and thus (2.2) is satisfied provided

$$\sup_{h \perp 1} \frac{\|\pi^n h\|_2}{\|h\|_2} \to 0.$$

In particular, if π has a complete set of eigenfunctions (necessarily orthogonal), then for $h \perp 1$,

$$\|\pi^n h\|_2 \le |\lambda|^n \|h\|_2,$$

where λ is the second largest eigenvalue of π in absolute value, or the largest eigenvalue of π when restricted to the orthogonal complement of the function 1. The condition (2.2) is then satisfied if $|\lambda| < 1$, which will be the case in Examples 3.3 and 3.5.

Theorem 2.2 below is the central limit theorem for $\hat{\theta}$. Recall that as a bilinear map on $E \times E$, $D_\theta f$ satisfies $D_\theta f(\alpha, \alpha) = \mathbf{E}_\theta(\langle \alpha, Y_1 - m_\theta \rangle^2)$ and can thus be interpreted as the covariance matrix for the vector Y_1.

THEOREM 2.2. *Assume* (2.1) *and* (2.2). *Then the vector* $\sqrt{n}(\hat{\theta} - \theta)$ *converges in distribution to* $\mathbf{N}(0, B)$, *where* B *is the symmetric bilinear form on* $E' \times E'$ *defined by*

$$B(v,v) = A\big([D_\theta f]^{-1}(v), [D_\theta f]^{-1}(v)\big).$$

PROOF. Since $\hat{\theta} = f^{-1}(\overline{Y})$ and f^{-1} is differentiable at $f(\theta)$, it follows that

$$\hat{\theta} - \theta = D_{f(\theta)} f^{-1}\big(\overline{Y} - f(\theta)\big) + o\big(\overline{Y} - f(\theta)\big),$$

where $D_{f(\theta)} f^{-1}$ is a linear map from E' to E and $o(\cdot)$ is a function from E' to E such that $o(x)/\|x\| \to 0$ as $\|x\| \to 0$. We can use the standard argument for transferring central limit theorems through differentiable maps to see that for each $v \in E'$,

$$\big\langle \sqrt{n}(\hat{\theta} - \theta), v \big\rangle \to N\Big(0, A\big([D_\theta f]^{-1}(v), [D_\theta f]^{-1}(v)\big)\Big).$$

However, this proves the multivariate assertion that $\sqrt{n}(\hat{\theta} - \theta)$ converges in distribution to $\mathbf{N}(0, B)$. □

3. Examples.

EXAMPLE 3.1. Let $d = 1$. Then E and E' have dimension $\binom{k+d-1}{d-1} = 1$, regardless of k. The parameter θ is simply an element of $\mathbf{R}^1$, and the process Y_n takes values in $\mathbf{R}^1$ as well as is given by $Y_n = \prod_0^{k-1} X_{n+i}$. Thus

$$m_\theta = \int_{\mathbf{R}^k} x_1 \times \cdots \times x_k \frac{\exp(\theta x_1 \times \cdots \times x_k)}{Z_\theta} \, d\mu^k(x_1, \dots, x_k).$$

If condition (2.2) holds, let $\sigma^2 = \lim(1/n)\mathbf{E}_\theta(\Sigma^n(Y_i - m_\theta))^2$. Then $\Sigma^n Y_i/\sqrt{n}$ converges in distribution to $N(m_\theta, \sigma^2)$. Since $[D_\theta f]^{-1} = 1/\mathbf{E}_\theta(Y_1 - m_\theta)^2$, it follows that $\sqrt{n}(\hat{\theta} - \theta)$ converges in distribution to $N(0, \sigma^2/[\mathbf{E}_\theta(Y_1 - m_\theta)^2]^2)$.

EXAMPLE 3.2. The special case where $d = 1$ and $k = 2$ corresponds to a Markov chain on the real line. The stationary density is given by $Z(x)/Z$. The chain is reversible since $Z(x)p(x, y)/Z = Z(y)p(y, x)/Z$, which leads to a self-adjoint transition operator π, but not symmetric.

Let the reference measure μ be Lebesgue measure on $(0, 1)$. Then we can take $\Theta = \mathbf{R} = E$ and the transition density $p(x, y)$ is given by

$$p(x,y) = \frac{\theta x e^{\theta x y}}{e^{\theta x} - 1},$$

which is bounded uniformly in x on $(0, 1)$ and satisfies conditions (2.1) and (2.2) (in fact, the chain satisfies the Doeblin condition with uniformly μ-integrable transition functions). The map f is given by

$$f(\theta) = \int_0^1 \int_0^1 xy e^{\theta x y} \, dx \, dy \Big/ \int_0^1 \int_0^1 e^{\theta x y} \, dx \, dy$$

$$= \frac{1}{\theta}\left(\frac{e^\theta - 1}{c_\theta} - 1\right),$$

where $c_\theta = \int_0^\theta (e^x - 1)/x\,dx$. The function f is monotone increasing from the real line $\Theta = (-\infty, \infty)$ onto the interval $(0, 1) = \operatorname{int}(C)$ (cf. Lemma 2.1). Now $\bar{Y} = \Sigma^n X_i X_{i+1}/n$ and the estimator $\hat{\theta}$ is the unique solution to

$$f(\hat{\theta}) = \bar{Y} \in (0, 1).$$

EXAMPLE 3.3. Suppose we are interested in a standard Gaussian reference measure. Then the transition density $p(x, y)$ with respect to Lebesgue measure on $\mathbf{R}$ becomes

$$p(x, y) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(y - \theta x)^2}{2} \right),$$

and when $\theta \in (-1, 1)$ this represents the transition probability for a process in discrete time analogous to the Ornstein–Uhlenbeck process. The sequence $X_0, X_1, \dots$ can also be represented as an autoregressive series with $X_n = \xi_n + \theta X_{n-1} = \Sigma_0^\infty \theta^i \xi_{n-i}$, $\xi_i \sim$ i.i.d. $N(0, 1)$.

Set $\Theta = (-1, 1)$. Conditions (2.1) are satisfied and the invariant marginal distribution on X_n is $N(0, 1/(1 - \theta^2))$. Condition (2.2) is also satisfied, since the self-adjoint operator π on L^2 takes the form

$$\pi h(x) = \int h(y) \frac{\exp(-(y - \theta x)^2/2)}{\sqrt{2\pi}}\, dy.$$

One can check that the spectrum of π is $1, \theta, \theta^2, \dots$ and that these numbers correspond to Hermite polynomials for the stationary distribution $N(0, 1/(1 - \theta^2))$. By the remarks following Lemma 2.4, the L^2 norm condition is satisfied.

The map f: $\Theta \to \mathbf{R}$ is given by

$$f(\theta) = \frac{\theta}{1 - \theta^2},$$

which is an increasing map from $\Theta = (-1, 1)$ onto $E = (-\infty, \infty)$, with asymptotes at ± 1. This is of course the covariance of X_n and X_{n+1} with joint bivariate normal distribution having mean 0 and covariance matrix R given by

$$R^{-1} = \begin{pmatrix} 1 & -\theta \\ -\theta & 1 \end{pmatrix}.$$

One can solve explicitly $f(\hat{\theta}) = \bar{Y}$ to find the estimator for θ. With some straightforward calculation it is seen that $\operatorname{Cov}(Y_i, Y_j) = \theta^2 \operatorname{Cov}(Y_i, Y_{j-1})$ for

$j \geq i + 2$, and

$$\sigma^2 = \lim \frac{1}{n} \operatorname{Var}_\theta\left(\sum^n Y_i\right) = \frac{1 - \theta^4 + 4\theta^2}{(1 - \theta^2)^3},$$

$$\operatorname{Var}(Y_i) = f'(\theta) = \frac{1 + \theta^2}{(1 - \theta^2)^2},$$

$$\sqrt{n}\,(\hat{\theta} - \theta) \to N\left(0, 1 + \theta^2 - \frac{8\theta^4}{(1 + \theta^2)^2}\right).$$

The information $I_n(\theta)$ in the sample $\{X_0, X_1, \dots, X_n\}$, given by

$$I_n(\theta) = \mathbf{E}_\theta\left(\left(\partial_\theta \log(1 - \theta^2) - X_0^2(1 - \theta^2) - \sum_1^n (X_i - \theta X_{i-1})^2\right)^2\right)$$

satisfies $I_n(\theta)/n \to 1/(1 - \theta^2)$. The maximum likelihood estimator $\hat{\theta}_{\mathrm{ML}}$ is the value of θ which maximizes the quantity $\log(1 - \theta^2) - X_0^2(1 - \theta^2) - \sum_1^n (X_i - \theta X_{i-1})^2$. It is known that $\hat{\theta}_{\mathrm{ML}}$ satisfies $\sqrt{n}\,(\hat{\theta}_{\mathrm{ML}} - \theta) \to N(0, 1 - \theta^2)$ [Box and Jenkins (1970), pages 280–281] and that this asymptotic variance is also attained by the sample correlation r_1. Thus the asymptotic variance for $\hat{\theta}$ in this example is slightly greater than the variance of these estimators.

EXAMPLE 3.4. Let $d = 2$ and let $k = 2$. Then E and E' have dimension 3, and E consists of symmetric bilinear forms on $\mathbf{R}^2 \times \mathbf{R}^2$. Elements of E can be identified with symmetric 2×2 matrices and a basis consists of the three vectors

$$(e_1, e_2, e_3) = \left(\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1/\sqrt{2} \\ 1/\sqrt{2} & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}\right),$$

which are orthogonal for the entrywise inner product. E is then a three-dimensional subspace of the vector space of all 2×2 matrices. A corresponding dual basis (e_1', e_2', e_3') for E' has the same matrix representation and $\langle e_i, e_j' \rangle = \delta_{ij}$ for the entrywise inner product. For example, if we think of $\mathbf{x}, \mathbf{y}$ in the state space $F = \mathbf{R}^2$ as column vectors, then

$$e_1(\mathbf{x}, \mathbf{y}) = \mathbf{x}^T \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \mathbf{y}.$$

With this identification, $f(\theta)$ is an element of E' determined by coordinates

$$f(\theta)(e_i) = \mathbf{E}_\theta[e_i(X_1, X_2)] = \int_{F \times F} \mathbf{x}^T e_i \mathbf{y} \frac{\exp(\mathbf{x}^T \theta \mathbf{y})}{Z_\theta} \mu \times \mu(d\mathbf{x}, d\mathbf{y}).$$

Also $D_\theta f$ is a linear map from E to E' given by a 3×3 matrix (d_{ij}) with columns $D_\theta f(e_i)$, the entries of which are determined by the effect of $D_\theta f(e_i)$ on each basis element:

$$d_{ij} = [D_\theta f(e_i)](e_j) = \mathbf{E}_\theta\big[(e_i(X_1, X_2) - \mathbf{E}_\theta[e_i(X_1, X_2)]) \times (e_j(X_1 X_2) - \mathbf{E}_\theta[e_j(X_1, X_2)])\big].$$

The bilinear form A on $E \times E$ can also be represented as a 3×3 symmetric matrix (a_{ij}), where

$$a_{ij} = \lim \frac{1}{n} \mathbf{E}_\theta \left(\sum_{m=1}^{n} \langle e_i, Y_m - m_\theta \rangle \sum_{m=1}^{n} \langle e_j, Y_m - m_\theta \rangle \right).$$

EXAMPLE 3.5. Take $d = 2$ and $k = 2$ as above and let the reference measure μ on $\mathbf{R}^2$ be the product of two independent $N(0,1)$ distributions. Let $\Theta \subset E$ consist of those symmetric matrices

$$\theta = \begin{pmatrix} a & b \\ b & c \end{pmatrix}$$

such that the matrix

$$M_\theta = \begin{pmatrix} 1 - a^2 - b^2 & -b(a+c) \\ -b(a+c) & 1 - b^2 - c^2 \end{pmatrix} = id - \theta^2$$

is positive definite. This means that the eigenvalues of θ are in $(-1, 1)$. One can easily show that conditions (2.1) are satisfied. One can show that condition (2.2) is satisfied, as we did in Example 3.3, by showing that the self-adjoint operator π on L^2 has largest eigenvalue equal to the largest eigenvalue θ_1 of the matrix θ, when π is restricted to the subspace of L^2 orthogonal to the function 1.

One finds that $Z_\theta^2 = 1/\det M_\theta$, and the transition density $p(\mathbf{x}, \mathbf{y})$ with respect to Lebesgue measure takes a form analogous to the one-dimensional situation,

$$p(\mathbf{x}, \mathbf{y}) = \frac{1}{2\pi} \exp\left(\frac{-\|\mathbf{y} - \theta \mathbf{x}\|^2}{2} \right),$$

so the step has the Gaussian distribution with mean $\theta\mathbf{x}$. The stationary marginal distribution on $F = \mathbf{R}^2$ is bivariate normal $N(0, M_\theta^{-1})$ and the stationary distribution on $F \times F = \{(x_1, x_2, y_1, y_2)\}$ is Gaussian with mean 0 and covariance matrix R such that

$$R^{-1} = \begin{pmatrix} 1 & 0 & -a & -b \\ 0 & 1 & -b & -c \\ -a & -b & 1 & 0 \\ -b & -c & 0 & 1 \end{pmatrix}.$$

In particular,

$$f(\theta)(e_1) = \frac{1}{Z_\theta} \int x_1 y_1 \exp(\mathbf{x}^T \theta \mathbf{y}) \mu \times \mu(d\mathbf{x}, d\mathbf{y}) = \frac{a + cb^2 - ac^2}{\det M_\theta},$$

$$f(\theta)(e_2) = \frac{1}{Z_\theta} \int \frac{y_1 x_2 + x_1 y_2}{\sqrt{2}} \exp(\mathbf{x}^T \theta \mathbf{y}) \mu \times \mu(d\mathbf{x}, d\mathbf{y}) = \frac{\sqrt{2}\, b(1 + ac - b^2)}{\det M_\theta},$$

$$f(\theta)(e_3) = \frac{1}{Z_\theta} \int x_2 y_2 \exp(\mathbf{x}^T \theta \mathbf{y}) \mu \times \mu(d\mathbf{x}, d\mathbf{y}) = \frac{1 - a^2 - b^2}{\det M_\theta}.$$

Thus we can represent $f(\theta) \in E'$ as the symmetric 2×2 matrix

$$f(\theta) = \frac{1}{\det M_\theta} \begin{pmatrix} a + cb^2 - ac^2 & b(1 + ac - b^2) \\ b(1 + ac - b^2) & 1 - a^2 - b^2 \end{pmatrix}$$

$$= \theta(id - \theta^2)^{-1}.$$

Finally, the statistic $\overline{Y} \in E'$ takes the following form. Let $X_i = (X_{i,1}, X_{i,2}) \in F = \mathbf{R}^2$. Then

$$\overline{Y}(e_i) = e_i \frac{1}{n} \sum_1^n (X_i, X_{i+1}),$$

and this gives the formulas

$$\overline{Y}(e_1) = \frac{1}{n^2} \Sigma X_{i,1} \Sigma X_{i+1,1},$$

$$\overline{Y}(e_2) = \frac{1}{\sqrt{2}\, n^2} [\Sigma X_{i,2} \Sigma X_{i+1,1} + \Sigma X_{i,1} \Sigma X_{i+1,2}],$$

$$\overline{Y}(e_3) = \frac{1}{n^2} \Sigma X_{i,2} \Sigma X_{i+1,2},$$

which in matrix form becomes

$$\overline{Y} = \begin{pmatrix} \overline{Y}(e_1) & \overline{Y}(e_2)/\sqrt{2} \\ \overline{Y}(e_2)/\sqrt{2} & \overline{Y}(e_3) \end{pmatrix}.$$

We remark that the exponential family presented here is a restrictive parametric model for a stationary process that is different from typical time series models. The advantage of this is a very simple estimation scheme which is based on studying a sample mean in a finite-dimensional vector space. The sample mean itself is an easy candidate for an ergodic theorem and a central limit theorem. Then a smooth map transfers its statistical properties to the estimator $\hat{\theta}$.

One can ask about the efficiency of the estimator $\hat{\theta}$ which we have introduced based on natural convexity considerations. Example 3.3 indicates that $\hat{\theta}$ does not always have minimum asymptotic variance, although it is close. In the i.i.d. model ($k = 1$), $\hat{\theta}$ is of course the MLE and an optimal large deviation property for the estimator was recently proved by Kester and Kallenberg (1986). A similar result for $k \geq 2$ will require first a careful look at the large deviation properties of $\overline{Y}$.

Acknowledgments. We thank the referees for indicating another approach to exponential families in Küchler and Sorensen (1989), and also for suggesting that assumptions in the first version be weakened.

REFERENCES

BARNDORFF-NIELSEN, O. (1978). *Information and Exponential Families*. Wiley, New York.

BERK, R. H. (1972). Consistency and asymptotic normality of MLE's for exponential models. *Ann. Math. Statist.* **43** 193–204.

BOX, G. E. P. and JENKINS, G. M. (1970). *Time Series Analysis, Forecasting and Control.* Holden-Day, San Francisco.

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

IONESCU TULCEA, C. T. (1949). Mesures dans les espaces produits. *Rend. Accad. Naz. Lincei* **7** 208–211.

KESTER, A. D. M. and KALLENBERG, W. C. M. (1986). Large deviations of estimators. *Ann. Statist.* **14** 648–664.

KÜCHLER, U. and SORENSEN, M. (1989). Exponential families of stochastic processes: a unifying semimartingale approach. *Internat. Statist. Rev.* **57** 123–144.

ROCKAFELLAR, R. T. (1970). *Convex Analysis*. Princeton Univ. Press.

ROSENBLATT, M. (1971). *Markov Processes. Structure and Asymptotic Behaviour*. Springer, New York.

SHIRYAYEV, A. N. (1984). *Probability*. Springer, New York.

DEPARTMENT OF MATHEMATICS
TULANE UNIVERSITY
NEW ORLEANS, LOUISIANA 70118