The Annals of Probability
1979, Vol. 7, No. 5, 874–881

LARGE DEVIATIONS OF FUNCTIONS OF MARKOVIAN TRANSITIONS AND MATHEMATICAL PROGRAMMING DUALITY[1]

BY GEUNG-HO KIM AND H. T. DAVID

Iowa State University

The duality-related point of view of Hoeffding for i.i.d. large deviations is extended to the Markov case. This reconciles Koopmans' "function-analytic" view of Markovian large deviations with Boza's "information theoretic" view, and extends the validity of the latter to the not-necessarily-finite case.

1. Introduction. Koopmans [20] has considered the rates of decay of probabilities of "nonlocal" (in the sense of Chernoff [8]) errors of sequences of likelihood ratio tests discriminating between two Markov processes P and Q. In Koopmans' setting, where P and Q are discrete-time and stationary, these rates of decay are expressible in terms of a certain extremal dominant root. Koopmans' work in effect determines the rate of decay of the probabilities of large deviations of the sample averages of the function $\ln[q(x_i|x_{i-1})/p(x_i/x_{i-1})]$ of observed transitions; i.e., of the probabilities of events

$$E_n: \{\Sigma_{i=1}^n \ln[q(X_i|X_{i-1})/p(X_i|X_{i-1})] \geqslant 0\}.$$

Slight modification of the argument in [20] verifies that the rate of decay of the probabilities of events

$$F_n: \{\Sigma_{i=1}^n a(X_{i-1}, X_i) \geqslant 0\},$$

$a(x,y)$ not necessarily of the form $\ln[q(y|x)/p(y|x)]$, similarly is expressible in terms of an analogous extremal dominant root.[2]

A related investigation, by Boza [6], dealing with the comparison of tests of hypotheses concerning finite state space stationary Markov chains, involves the rate of decay of probabilities of certain events G_n defined in terms of transition counts; specializations of these events G_n are of the form F_n. Boza found the rate of decay of the probabilities of the events G_n (and hence of the events F_n) to be given by a certain extremal information functional.

Thus two expressions, one function-analytic or spectral [20] and the other information theoretic [6], are available for the rate of decay of the probabilities of events F_n, with the second restricted to finite state space.

Received March 9, 1978.

[1]Research partially supported by AFOSR grant #78-3518.

AMS 1970 *subject classifications.* Primary, 60F19; secondary, 60J05.

Key words and phrases. Large deviations, Markov process, information, mathematical programming, duality.

[2]R. R. Bahadur has pointed out to us the asymptotic equivalence, in analogy to the i.i.d. case (cf. [2]), of the forms $a(x,y)$ and $\ln[q(y|x)/p(y|x)]$, based on the transformation of our Remark 3. Thus Koopmans' L^2 setting [20] would be at least as appropriate for us as the setting of [15] adopted below, provided it could be shown to allow part (iii) of Theorem 1.

Also available in the literature are corresponding "function-analytic" (Chernoff [8], Bahadur and Rao [3]) and information theoretic (Sanov [23]) rates of decay in the analogous i.i.d. case (with the validity of neither restricted to the finite case), and reconciliations of the two, essentially using mathematical programming duality, by Hoeffding [16] and Whittle [24]. See also general expositions (e.g., Barndorff-Nielsen [4], Borovkov and Rogozin [5], and Efron and Truax [12]) which contain at least implicitly related duality ideas.

This paper, expanding on [19], brings the duality approach to the Markov case. Our objective here is not only to extend to this case the "direct" duality point of view of [16] and [24]; it is, in addition, to extend the work in [6] to the not-necessarily-finite case. This roundabout approach to extending [6] (by way of [20] plus duality) may not be entirely unnatural, since the argument in [6] is combinatoric.

Assumptions are laid out in Section 2. In Section 3, some properties of the relevant dominant root λ_t are examined, in a manner analogous to Koopmans'. Among these is an expression for the derivative λ_t' of λ_t, of use in the later development. We also compute the rate of decay of the probabilities of the event F_n in terms of λ_t, in a manner entirely analogous to that in [20]. Section 4 implements the program of the previous paragraph; the point of view here is essentially Hoeffding's, extended to the Markov case and viewed in the light of mathematical programming duality.

2. Assumptions. Let $(S, \mathfrak{B}, \mu)$ be a finite measure space. Following Harris [15] consider a transition probability density kernel $p(y|x)$ that (a) is $\mathfrak{B} \times \mathfrak{B}$-measurable, and (b) has the property that there are real numbers c and d, and a positive integer N, such that the Nth iterate $p^{(N)}(y|x)$ of $p(y|x)$ satisfies

$$\text{(A1)} \qquad 0 < c \leqslant p^{(N)}(y|x) \leqslant d < +\infty.$$

By Theorem 10.1 of [15], the kernel $p(y|x)$ has a bounded left eigenfunction $\phi(x)$ corresponding to the dominant root 1.

Now consider a $\mathfrak{B} \times \mathfrak{B}$-measurable function $a(x, y)$ satisfying

$$\text{(A2)} \qquad |a(x,y)| < M < +\infty \quad \text{on} \quad S \times S,$$

$$\text{(A3)} \qquad \int\int a(x,y)p(y|x)\phi(x)\,d\mu^2 < 0,$$

and, for some $\varepsilon > 0$,

$$\text{(A4)} \qquad \operatorname{ess\,inf}_x\left[\int_{\{a(x,y)>\varepsilon\}} p(y|x)\,d\mu(y)\right] > 0.$$

Also define kernels $K_t(x,y) = e^{ta(x,y)}p(y|x)$, $-\infty < t < +\infty$. In view of (A1) and (A2), K_t satisfies as well the assumptions of Theorem 10.1 of [15] for all t, so that K_t possesses a positive eigenvalue λ_t, and corresponding left and right eigenfunctions ϕ_t and ψ_t that are bounded and uniformly positive. Furthermore, if ϕ_t and ψ_t are normalized so that $\int \phi_t \psi_t \, d\mu = 1$, which will henceforth be assumed, then, for $(x, y) \in S \times S$,

$$\text{(2.1)} \qquad K_t^{(n)}(x,y) = \lambda_t^n \psi_t(x)\phi_t(y)\left[1 + g_t(x,y;n)\right], \qquad \forall t,$$

where $|g_t(x,y;n)| \leqslant \Delta_t^n$, with $0 < \Delta_t < 1$.

Some remarks on assumptions (A3) and (A4) also are in order; (A3) is used below in conjunction with (A2) to guarantee that λ_t is decreasing near zero, while (A4) is used to guarantee that λ_t tends to $+\infty$ with t. In the i.i.d. case, analogous conditions have been used by Bahadur [1], to ensure the "standard condition."

3. The decay rate and the dominant root λ_t. Let $\{X_i\}_{i=0}^{\infty}$ form a Markov process over state space S with initial density $p_0(x)$ and transition density $p(y|x)$. Let $M_n(t)$ be the moment generating function of $S_n \equiv \Sigma_{i=1}^n a(X_{i-1}, X_i)$.

THEOREM 1. *Under assumptions* (A1) *and* (A2),

(i) $\lim_n [M_n(t)]^{1/n} = \lambda_t$, $\forall t$.

(ii) λ_t *is convex and analytic*, $\forall t$.

(iii) $\lambda_t' = \int\int a(x, y) K_t(x, y) \phi_t(x) \psi_t(y) d\mu^2$, $\forall t$.

Under assumptions (A1)–(A4),

(iv) λ_t *achieves a unique minimum at* t^*, $0 < t^* < +\infty$.

PROOF. (i) We shall in fact prove the slightly stronger

$$\lim_n (M_n(t)/\lambda_t^n) = A_t, \quad \forall t, \tag{3.1}$$

where $A_t = \int\int \psi_t(x)\phi_t(y)p_0(x)d\mu^2 > 0$. Now $M_n(t) = \int\int K_t^{(n)}(x, y)p_0(x)d\mu^2 = \lambda_t^n(A_t + B_{t,n})$, with $B_{t,n} = \int\int \psi_t(x)\phi_t(y)g_t(x, y; n)p_0(x)d\mu^2$, where the first equality is by definition, and the second holds in view of (2.1), which also entails $\lim_n B_{t,n} = 0$.

(ii) Convexity follows from (i), and the fact that $M_n(t)^{1/n}$ is convex for all n. As for analyticity, following Koopmans [20] consider the bilateral Laplace transform $H_n(z)$ of $F_n(x) \equiv \Pr\{S_n \leqslant x\}$, and, for any $r > 0$, let $T_r = \{t: |t| \leqslant r\}$. In view of (A2), $|[H_n(z)]^{1/n}| \leqslant e^{Mr}$ for every z with real part in T_r, and for every $n = 1, 2, \cdots$. Hence each $[H_n(z)]^{1/n}$ is analytic in the infinite strip $\mathfrak{T}_r$: $\{z = t + iu, t \in T_r\}$. Thus (i), the uniform bound e^{Mr} for $|[H_n(z)]^{1/n}|$, and Vitali's theorem imply that $\lim_n [H_n(z)]^{1/n}$ is analytic in the interior of $\mathfrak{T}_r$; hence that $\lim_n [H_n(t)]^{1/n} = \lim_n [M_n(t)]^{1/n} = \lambda_t$ is analytic in the interior of T_r.

(iii) Let $\delta > 0$, and consider

$$(\lambda_{t+\delta} - \lambda_t)\psi_{t+\delta}(x) + \lambda_t[\psi_{t+\delta}(x) - \psi_t(x)] \tag{3.2}$$
$$= \lambda_{t+\delta}\psi_{t+\delta}(x) - \lambda_t\psi_t(x)$$

$$= \int K_{t+\delta}(x, y)[\psi_{t+\delta}(y) - \psi_t(y)]d\mu(y) + \int[e^{\delta a(x,y)} - 1]K_t(x, y)\psi_t(y)d\mu(y), \tag{3.3}$$

where the second equality follows from subtracting and adding $\int K_{t+\delta}(x, y) \cdot \psi_t(y)d\mu(y)$. Now, multiplying (3.2) and (3.3) by $\phi_{t+\delta}(x)$, integrating w.r.t. x (using Fubini's theorem for the first addend of (3.3)), and rearranging, $(\lambda_{t+\delta} - \lambda_t) - (\lambda_{t+\delta} - \lambda_t)\{\int(\psi_{t+\delta}(x) - \psi_t(x))\phi_{t+\delta}(x)d\mu(x)\} = \int\int(e^{\delta a(x,y)} - 1)K_t(x, y)\phi_{t+\delta}(x)\psi_t(y)d\mu^2$. Rearranging again and dividing by δ,

$$(\lambda_{t+\delta} - \lambda_t)\delta^{-1} = \int\int(e^{\delta a(x,y)} - 1)\delta^{-1}k_\delta(x, y)d\mu^2, \tag{3.4}$$

where $k_\delta(x,y) \equiv K_t(x,y)\phi_{t+\delta}(x)\psi_t(y)[\int\phi_{t+\delta}(x)\psi_t(x)d\mu(x)]^{-1}$. Similarly,

$$(\lambda_t - \lambda_{t-\delta})\delta^{-1} = \int\int(1 - e^{-\delta a(x,y)})\delta^{-1}k_{-\delta}(x,y)d\mu^2. \tag{3.5}$$

Now $\lambda_t' = \lim_{\delta\to 0}(\lambda_{t+\delta} - \lambda_t)\delta^{-1} \geqslant \lim_{\delta\to 0}\int\int a(x,y)k_\delta(x,y)d\mu^2 = \int\int a(x,y) k_0(x,y)d\mu^2$.

Here the first equality is due to the fact that λ_t is analytic.

The second equality follows from the uniform (in (x,y)) convergence of k_δ to k_0, since $a(x,y)$ is bounded and $\mu(S)$ is finite; this uniform convergence of k_δ in turn follows from the uniform (in x) convergence of $\phi_{t+\delta}$ to ϕ_t, since the other components of k_δ do not depend on δ; finally, the uniform convergence of $\phi_{t+\delta}$ follows from a straightforward application of the arguments in Theorem 2.4.2 of [10] or Theorem 5.2 of [21], with $(K_t^{(N)}, K_{t+\delta}^{(N)})$, $|\delta| < 1/M$, replacing (M_n, M_{n+1}); cf. [18] for the details.

The inequality follows from (3.4), from the fact that the limit exists, and the fact that $e^{\delta a} - 1 \geqslant \delta a$.

Similarly, using (3.5) and the comparison $1 - e^{-\delta a} \leqslant \delta a$,

$$\lambda_t' \leqslant \int\int a(x,y)k_0(x,y)d\mu^2.$$

(iv) (A3) and part (iii) imply that

$$\lambda_0' = \int\int a(x,y)p(y|x)\phi_0(x)d\mu^2 < 0. \tag{3.6}$$

Now, by (A4), $\exists\eta > 0 \ni \text{ess inf}_x[\int_{\{a(x,y)>\varepsilon\}}p(y|x)d\mu(y)] \geqslant \eta$. Hence, assuming $\int\phi_t d\mu = 1$,

$$\begin{aligned}\lambda_t &= \int\int e^{ta(x,y)}p(y|\,x)\phi_t(x)d\mu^2 \\ &\geqslant \int\Big[\int_{\{a(x,y)>\varepsilon\}}e^{ta(x,y)}p(y|x)d\mu(y)\Big]\phi_t(x)d\mu(x) \\ &\geqslant e^{t\varepsilon}.\eta \to +\infty \text{ as } t\to+\infty.\end{aligned} \tag{3.7}$$

(3.6), (3.7) and convexity assure that λ_t achieves its minimum in $(0, +\infty)$, the uniqueness of which is guaranteed by the fact that λ_t, being analytic for all t, cannot be flat in a proper subinterval without being flat everywhere.

The remaining part of this section is to verify that

$$\lim_n \Pr\{S_n \geqslant 0\}^{1/n} = \lambda_{t^*}. \tag{3.8}$$

To begin with, analogously to the argument in Lemma 5.1 of [20], Theorem 1(i) yields $\limsup_n \Pr\{S_n \geqslant 0\}^{1/n} \leqslant \lambda_{t^*}$. Next, again analogously to the argument in [20], note that, in view of (ii) and (iv) of Theorem 1, for any $t_0 > t^*$ and any $b > 0$, there is an $s > 0$ such that

$$\lambda_{t_0-s}/\lambda_{t_0} < 1 \quad \text{and} \quad \lambda_{t_0+s}/\lambda_{t_0} < e^{bs}, \tag{3.9}$$

and also, as in Lemmata 5.2 and 5.3 of [20],

(3.10)

$$\Pr\{S_n \geqslant 0\}^{1/n} \geqslant e^{-bt}M_n(t)\big[1 - M_n(t-s)/M_n(t) - e^{-nbs}M_n(t+s)/M_n(t)\big]^{1/n},$$

$$\text{for all} \quad b > 0, s > 0 \quad \text{and} \quad t > 0.$$

Now, in view of (3.1) and (3.9), the right-hand side of (3.10) with $t = t_0$ tends to $e^{-bt_0}\lambda_{t_0}$ with n, so that $\liminf_n \Pr\{S_n \geqslant 0\}^{1/n} \geqslant e^{-bt_0}\lambda_{t_0}$. But $b > 0$ was arbitrary, so that $\liminf_n \Pr\{S_n \geqslant 0\}^{1/n} \geqslant \lambda_{t_0} \geqslant \lambda_{t^*}$.

4. The two mutually weakly dual problems. As indicated at the end of Section 1, this section brings the direct duality-related point of view to our Markov setting, by equating the decay rate $\ln \lambda_{t^*}$ of the previous section to a not-necessarily-finite analogue of Boza's information theoretic decay rate. This equality, combined with (3.8), extends the validity of Boza's rate expression.

For a bivariate density $f(x, y)$ on $S \times S$, let $h(x)$ and $g(y|x)$ be the marginal density on S and the essentially unique conditional density kernel on $S \times S$ respectively. Now let the set $\mathcal{F}_a$ of densities f satisfy

$$\int f(x, y)\,d\mu(y) = \int f(y, x)\,d\mu(y), \qquad \text{a.a. } x \in S \tag{4.1}$$

$$\int\int a(x, y) f(x, y)\,d\mu^2 \geqslant 0, \tag{4.2}$$

and

$$\int\int \ln[g(y|x)/p(y|x)] f(x, y)\,d\mu^2 < +\infty. \tag{4.3}$$

Define

$$I(f, p) = \int\int \ln[g(y|x)/p(y|x)] f(x, y)\,d\mu^2, \quad f \in \mathcal{F}_a.$$

THEOREM 2. *Under* (A1)–(A4),

$$I(f, p) \geqslant -\ln \lambda_t, \quad \textit{for } f \in \mathcal{F}_a \textit{ and } t \in (0, +\infty). \tag{4.4}$$

PROOF. Note first that (4.1) implies that, for any bounded measurable function $s(\cdot)$,

$$\int\int s(x) f(x, y)\,d\mu^2 = \int\int s(y) f(x, y)\,d\mu^2. \tag{4.5}$$

Now write

$$
\begin{aligned}
&(4.6a) \quad I(f, p) \geqslant \int\int \ln[g(y|x)/p(y|x)]\, g(y|x) h(x)\,d\mu^2 \\
&(4.6b) \quad\quad - \int\int \ln[\psi_t(y)/\psi_t(x)]\, g(y|x) h(x)\,d\mu^2 \\
&(4.6c) \quad\quad - t\int\int a(x, y) g(y|x) h(x)\,d\mu^2 \\
&\quad\quad = \int\int -\ln[p(y|x)/g(y|x).\psi_t(y)/\psi_t(x).e^{ta(x,y)}]\, g(y|x) h(x)\,d\mu^2 \\
&(4.6d) \quad\quad \geqslant -\ln \int\int \psi_t(y)/\psi_t(x).e^{ta(x,y)} p(y|x) h(x)\,d\mu^2 \\
&\quad\quad = -\ln \lambda_t.
\end{aligned}
$$

The first inequality is due to the fact that (4.6b) is zero in view of (4.5), and (4.6c) is nonpositive in view of (4.2). The second inequality is Jensen's, and the last equality follows by integrating first w.r.t. y and then w.r.t. x, and appealing to definitions of ψ_t and λ_t given in Section 2.

Now let t^* be the unique minimizer of λ_t identified in Theorem 1(iv), and define

$$h^*(x) = \psi_{t^*}(x)\phi_{t^*}(x)$$
$$(4.7)\qquad g^*(y|x) = K_{t^*}(x,y)\psi_{t^*}(y)/\lambda_{t^*}\psi_{t^*}(x)$$
$$f^*(x,y) = h^*(x)g^*(y|x).$$

COROLLARY 1. $\min_{f\in\mathcal{F}_a} I(f,p) = I(f^*,p) = -\ln\lambda_{t^*} = \max_{t\in(0,\infty)}[-\ln\lambda_t]$.

PROOF. In v956 66 Theorem 2, it is sufficient to verify that $f^* \in \mathcal{F}_a$ and that $I(f^*,p) = -\ln\lambda_{t^*}$, i.e., that equality holds in the statement of Theorem 2 for $f = f^*$ and $t = t^*$ (cf. Remark 1 below). That $f^* \in \mathcal{F}_a$ is easily verified. That equality holds in Theorem 2 for $f = f^*$ and $t = t^*$ follows from verifying equality in (4.6a) and (4.6d). Regarding (4.6a), one needs to verify only that expression (4.6c) equals 0 when $f = f^*$, which follows from Theorem 1(iii), (iv), which prescribe that $(-t^*\lambda_{t^*}^{-1})\int\int a(x,y)K_{t^*}(x,y)\phi_{t^*}(x)\psi_{t^*}(y)d\mu^2 = (-t^*\lambda_{t^*}^{-1})\lambda'_{t^*} = 0$. Regarding (4.6d), the argument of $\ln(\cdot)$, namely $e^{t^*a(x,y)}p(y|x)\psi_{t^*}(y)/g^*(y|x)\psi_{t^*}(x)$, equals the constant λ_{t^*}, in view of (4.7).

REMARK 1. The argument used in fact exploits the concept of weak duality of mathematical programming in the following way: if U is the problem of minimizing $u(q)$, w.r.t. q over Q, and V is the problem of maximizing $v(r)$, w.r.t. r over R, then the pair (U, V) is said to be mutually weakly dual if $u(q) \geqslant v(r)$ for $q \in Q$ and $r \in R$. Now, if in addition a pair (q^*, r^*) satisfies $u(q^*) = v(r^*)$, then (q^*, r^*) are optimal solutions, solving respectively U and V.

REMARK 2. As suggested by [16], mutually dual problem pairs typically admit certain irregular parametric cases for which V is "unbounded" and U is "infeasible." This feature is present also here, if (A4) is replaced by (A4)$'$: $\int_{\{a(x,y)<-\varepsilon\}} p(y|x)d\mu(y) = 1$, for all $x \in S$. In this case, (a) $\mathcal{F}_a$ is empty, and (b) $\sup_{t\in(0,\infty)}[-\ln\lambda_t] = +\infty$.

To see (a), note that (A4)$'$ implies that $\int\int a(x,y)p(y|x)h(x)d\mu^2 < 0$, for any $h(x)$, which, together with (4.3), implies that $\int\int a(x,y)g(y|x)h(x)d\mu^2 < 0$, for any g. To show (b), observe that (A4)$'$ implies that $\lambda_t = \int\int e^{ta(x,y)}p(y|x)\phi_t(x)d\mu^2 < e^{-t\varepsilon}$, $t > 0$.

The symmetric alternative, with unbounded U and infeasible V, cannot be exhibited in view of nonnegativity of $I(f,p)$ (i.e., U admits a natural lower bound). However, the parametric case where minimization of $I(f,p)$ achieves the lower bound 0 can be tied to the case where $\lambda'_0 \geqslant 0$. For suppose that (A3) is replaced by (A3)$'$: $\int\int a(x,y)p(y|x)\phi(x)d\mu^2 \geqslant 0$. Then $\min_{f\in\mathcal{F}_a} I(f,p) = 0$, trivially with $f(x,y) = p(y|x)\phi_0(x)$, and $\sup_{t>0}[-\ln\lambda_t] = -\ln\lambda_0 = 0$.

REMARK 3. The parametrized family $\{f_t(x,y),\ t \geqslant 0\}$, $f_t(x,y) = K_t(x,y)\phi_t(x)\cdot\psi_t(y)\cdot\lambda_t^{-1}$, is a Markov analogue of a construct that has repeatedly been used for the i.i.d. case, for example by Cramér [11], Khinchin [17], Chernoff [8], [9], Bahadur [1], [2] and Feller [14].

REMARK 4. The restriction (4.1) is in fact an ergodicity-related condition, since it is equivalent to the condition that h is an eigenfunction of g corresponding to the root 1.

Acknowledgment. We are indebted to L. H. Koopmans, for remarks concerning the applicability of Vitali's theorem. We are equally indebted to R. R. Bahadur and Malay Ghosh, for helpful comments and references.

REFERENCES

[1] BAHADUR, R. R. (1971). *Some Limit Theorems in Statistics*. SIAM, Philadelphia.

[2] BAHADUR, R. R. and RAGHAVACHARI, M. (1971). Some asymptotic properties of likelihood ratios on general sample spaces. *Proc. Sixth Berkeley Symp. Math. Statist. Prob.* **1** 129–152. Univ. of California Press.

[3] BAHADUR, R. R. and RAO, R. R. (1960). On deviation of the sample mean. *Ann. Math. Statist.* **31** 1015–1027.

[4] BARNDORFF-NIELSON, O. (1970). *Exponential Families: Exact Theory*. Aarhus University, Various Publication Series, Vol. 19.

[5] BOROVKOV, A. A. and ROGOZIN, B. A. (1965). On the central limit theorem in multidimensional case. *Theor. Probability Appl.* **10** 55–62.

[6] BOZA, L. B. (1971). Asymptotically optimal tests for finite Markov chains. *Ann. Math. Statist.* **42** 1992–2007.

[7] CHANDA, K. S. (1972). On estimation of tail end probabilities of the sample mean for linear stochastic processes. *Ann. Math. Statist.* **43** 1680–1686.

[8] CHERNOFF, H. (1952). A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Ann. Math. Statist.* **23** 493–507.

[9] CHERNOFF, H. (1972). *Sequential Analysis and Optimal Design*. SIAM, Philadelphia.

[10] CONN, P. S. (1969). Asymptotic properties of sequences of positive kernels. Unpublished Ph.D. dissertation, Iowa State Univ.

[11] CRAMÉR, H. (1938). Sur un nouveau théorème-limite de la théorie des probabilités. *Actualités Sci. Indust.* **736** 5–23.

[12] EFRON, B. and TRUAX, D. (1968). Large deviations theory in exponential families. *Ann. Math. Statist.* **39** 1402–1424.

[13] EL-SABBAGH, M. F. A. (1973). Aspects of the convergence of Bayes policies and posterior distributions for Markov decision processes. Unpublished Ph.D. dissertation, Iowa State Univ.

[14] FELLER, W. (1971). *An Introduction to Probability Theory and its Applications*, **2**, 2nd ed. Wiley, New York.

[15] HARRIS, T. E. (1963). *Theory of Branching Processes*. Springer-Verlag, Berlin.

[16] HOEFFDING, W. (1967). On probabilities of large deviations. *Proc. Fifth Berkeley Symp. Math. Statist. Prob.* **1** 203–219. Univ. California Press.

[17] KHINCHIN, A. I. (1949). *Mathematical Foundations of Statistical Mechanics*. Dover, New York.

[18] KIM, G.-H. (1978). Constructive optimization for infinite-dimensional problems in probability and statistics. Unpublished Ph.D. dissertation, Iowa State Univ.

[19] KIM, G.-H., EL-SABBAGH, M. F. A. and DAVID, H. T. (1977). Lagrangian duality and large deviations for Markov chains (preliminary report). *IMS Bulletin* **6** 140.

[20] KOOPMANS, L. H. (1960). Asymptotic rate of discrimination for Markov processes. *Ann. Math. Statist.* **31** 982–994.

[21] MADSEN, R. W. and CONN, P. S. (1973). Ergodic behavior for nonnegative kernels. *Ann. Probability* **1** 995–1013.

[22] MONTROLL, E. W. (1947). On the theory of Markov chains. *Ann. Math. Statist.* **18** 18–36.

[23] SANOV, I. N. (1957). On the probability of large deviations of random variables. *Mat. Sb.* **42** 11–44 (in Russian). (English translation in *Selected Transl. Math. Statist. Prob.* (1961) **1** 213–244.)

[24] WHITTLE, P. (1971). *Optimization Under Constraints; Theory and Applications of Nonlinear Programming*. Wiley-Interscience, London.

DEPARTMENT OF STATISTICS
STATE UNIVERSITY OF NEW YORK, BUFFALO
AMHERST, NEW YORK 14226

DEPARTMENT OF STATISTICS
IOWA STATE UNIVERSITY
AMES, IOWA 50010