The Annals of Probability
1993, Vol. 21, No. 2, 721–758

I-PROJECTION AND CONDITIONAL LIMIT THEOREMS FOR DISCRETE PARAMETER MARKOV PROCESSES

BY CAROLYN SCHROEDER

University of Massachusetts, Lowell

Let $(X, \mathscr{B})$ be a compact metric space with $\mathscr{B}$ the σ-field of Borel sets. Suppose this is the state space of a discrete parameter Markov process. Let C be a closed convex set of probability measures on X. Known results on the asymptotic behavior of the probability that the empirical distributions $\hat{P}_n$ belong to C and new results on the Markov process distribution of $\omega_0, \dots, \omega_{n-1}$ under the condition $\hat{P}_n \in C$ are obtained simultaneously through a large deviations estimate. In particular, the Markov process distribution under the condition $\hat{P}_n \in C$ is shown to have an asymptotic quasi-Markov property, generalizing a concept of Csiszár.

1. Introduction. Suppose $X_1, X_2, \dots$ is a sequence of independent random variables taking values in an arbitrary measure space $(S, \mathscr{B})$ with common distribution P_X. The empirical distribution of a sample $s = (s_1, \dots, s_n) \in S^n$ is the discrete probability measure defined by

$$\hat{P}_n(s, B) = \frac{1}{n} \sum_{i=1}^{n} \chi_B(s_i).$$

If P_X^n is the nth Cartesian power of P_X, the probability that the empirical distribution $\hat{P}_n$ of $(X_1, \dots, X_n)$ belongs to a set C of probability measure on $(S, \mathscr{B})$ is given by

$$P\{\hat{P}_n \in C\} = P_X^n(A_n), \qquad A_n = \{s \colon \hat{P}_n(s, \cdot) \in C\}.$$

This last probability is well defined if $A_n \in \mathscr{B}^n$. Csiszár (1984) defines a set C of probability measures as having the Sanov property if

$$\text{(1.1)} \qquad \lim_{n \to \infty} \frac{1}{n} \log P\{\hat{P}_n \in C\} = -h(C, P_X),$$

where $h(C, P_X) = \inf_{Q \in C} h(Q, P_X)$ and

$$\text{(1.2)} \qquad h(Q, P_X) = \begin{cases} \int \log(dQ/dP_X)\, dQ, & \text{if } Q \ll P_X, \\ +\infty, & \text{otherwise.} \end{cases}$$

In the event $A_n \notin \mathscr{B}^n$, the Sanov property is interpreted to mean that the limit relation holds for both the upper and lower probabilities $\overline{P}\{\hat{P}_n \in C\}$ and

Received April 1990, revised October 1991.
AMS 1991 *subject classifications*. Primary 60F10, 60J05, 60G10; secondary 62B10, 94A17.
Key words and phrases. *I*-projection, large deviations in abstract space, asymptotically quasi-Markov.

$\underline{P}\{\hat{P}_n \in C\}$. Here

$$\overline{P}\{\hat{P}_n \in C\} = P_X^n(\overline{A_n}), \qquad \underline{P}\{\hat{P}_n \in C\} = P_X^n(\underline{A_n}),$$

where $\overline{A_n} \supset A_n$ and $\underline{A_n} \subset A_n$ respectively are sets in $\mathscr{B}^n$ having minimum, respectively maximum, P_X^n measure among all such sets. The limit relation (1.1) is often referred to as Sanov's theorem due to the importance of Sanov (1957).

An alternative definition to (1.2) is

$$(1.3) \quad h(Q, P_X) = \sup_{\mathscr{P}} h_{\mathscr{P}}(Q, P_X), \qquad h_{\mathscr{P}}(Q, P_X) = \sum_{i=1}^{k} Q(B_i) \log \frac{Q(B_i)}{P_X(B_i)},$$

where $\mathscr{P} = (B_1, \dots, B_k)$ ranges over all finite measurable partitions. Here the conventions $0 \log 0 = 0 \log 0/0 = 0$ and $a \log a/0 = +\infty$ if $a > 0$ apply. A proof of the equivalence of (1.2) and (1.3) is given in Pinsker (1964), Theorem 2.4.2.

A set of probability measure Π on $(S, \mathscr{B})$ is completely convex if for every probability space $(\Omega, \mathscr{A}, \mu)$ and $\mathscr{A}$-measurable mapping $\omega \to \nu(\omega, \cdot) \in \Pi$, the probability measure $\mu\nu$ defined by

$$\mu\nu(B) = \int_{\Omega} \nu(\cdot, B)\, d\mu, \qquad B \in \mathscr{B},$$

also belongs to Π. A convex set of probability measures Π is almost completely convex if there exist completely convex subsets $\Pi_1 \subset \Pi_2 \subset \cdots$ of Π such that $\bigcup_{k=1}^{\infty} \Pi_k \supset \Pi \cap \Lambda_f$, Λ_f the set of probability measures on $(S, \mathscr{B})$ whose support is a finite subset of S. Csiszár (1984) shows that the Sanov property for an almost completely convex set C of probability measures implies that the $X_1, \dots, X_n$ are asymptotically quasi-independent under the condition $\hat{P}_n \in C$. To describe this result, a probability measure P^* is called the *I-projection* of P_X on C if $h(P^*, P_X) = h(C, P_X)$. A probability measure P^* is called the *generalized I-projection* if any sequence of $P_n \in C$ with $h(P_n, P_X) \to h(C, P_X)$ converges to P^* in variation. For C a convex set of probability measures, the generalized I-projection exists [Csiszár (1975), Theorem 2.1 and Remark]. If $X^n = (X_1, \dots, X_n)$ and $P_{X^n|\hat{P}_n \in C}$ denotes the conditional P_X^n distribution of X^n under the condition $\hat{P}_n \in C$, a completely convex set, the asymptotic quasi-independence shown by Csiszár (1984), Theorem 1, is

$$(1.4) \qquad \lim_{n \to \infty} \frac{1}{n} h\left(P_{X^n|\hat{P}_n \in C}, (P^*)^n\right) = 0,$$

where P^* is the generalized I-projection of P_X on C.

Here an analogous result is formulated for a discrete parameter Markov process with state space a compact metric space X with its σ-field of Borel sets. We assume the Markov process has stationary transition probability function $\pi(dy|x)$. In addition, we assume for $\lambda(dx)$ a probability measure on X

that:

1. $\pi(dy|x) = \pi(y|x)\lambda(dy)$. Then $\pi(y|x)$ may be chosen jointly measurable in x and y.
2. There exist constants a and A such that $0 < a \le \pi(y|x) \le A < \infty$ for all $x \in X$ and almost all λ (measure) $y \in X$.
3. For any continuous function $f(y)$

$$\int_X \pi(dy|x) f(y)$$

is a continuous function of x. Under assumptions 1 and 2, the same will hold for any $f(y) \in L^1(\lambda)$.

Let $(\Omega_x, \mathscr{B})$ denote the measure space of all sequences $(\omega_0, \omega_1, \omega_2, \dots)$ with $\omega_0 = x \in X$, $\omega_j \in X$ and $\mathscr{B}$ the Borel sets of Ω_x. Then $(\Omega_x, \mathscr{B}) = \prod_{i=0}^{\infty}(X_i, \mathscr{B}_i)$, where $X_i = X$ and $\mathscr{B}_i$ the Borel sets on X, $i = 1, 2, \dots$, and $X_0 = x$, $\mathscr{B}_0 = \{x\}$. The transition function $\pi(dy|x)$ induces a probability measure on Ω_x; call it P_x. Impose the weak topology on the space $\mathscr{M}(X)$ of probability measures on X. Let $\hat{P}_n(\omega, \cdot)$ be the empirical distribution of $(\omega_0, \dots, \omega_{n-1})$, $\omega \in \Omega_x$. Then for each n, $\hat{P}_n(\omega, \cdot)$: $\Omega_x \to \mathscr{M}(X)$ is a continuous map on $(\omega_0, \dots, \omega_{n-1})$, so for any measurable $S \in \mathscr{M}(X)$, $\{\omega\colon \hat{P}_n(\omega, \cdot) \in S\}$ is measurable.

Donsker and Varadhan (1975, 1976) described the asymptotic probabilities that $\hat{P}_n(\omega, \cdot)$ lies in closed and open sets of $\mathscr{M}(X)$.

For any open set $G \subset \mathscr{M}(X)$,

$$\liminf_{n \to \infty} \frac{1}{n} \log P_x\big\{\hat{P}_n(\omega, \cdot) \in G\big\} \ge -\inf_{\mu \in G} I(\mu), \tag{1.5}$$

uniformly for $x \in X$ [Donsker and Varadhan (1976), Corollary 3.4]. Here

$$I(\mu) = -\inf_{u \in \mathscr{U}_1} \int_X \log\left(\frac{\pi u}{u}\right)(x)\mu(dx), \tag{1.6}$$

where $\mathscr{U}_1$ is the set of continuous positive functions on X and

$$\pi u(x) = \int_X u(y)\pi(dy|x).$$

Also for any closed set $C \in \mathscr{M}(X)$,

$$\limsup_{n \to \infty} \frac{1}{n} \log \sup_{x \in X} P_x\big\{\hat{P}_n(\omega, \cdot) \in C\big\} \le -\inf_{\mu \in C} I(u) \tag{1.7}$$

[Donsker and Varadhan (1976), Theorem 4.4].

We first define an I-projection appropriate to this context. In the literature, an I-projection is a minimizing element of a divergence or a measure of entropy [Csiszár (1975), (1984) and Csiszár, Cover and Choi (1987)]. Under assumption 3 on $\pi(dy|x)$, $I(\mu)$ defined by (1.6) is a lower semicontinuous function of μ so that if C is a closed set in $\mathscr{M}(X)$,

$$I(C) = \inf_{\mu \in C} I(\mu) = I(\mu^*) \tag{1.8}$$

for some $\mu^* \in C$. Let Λ_0 be the set of probability measures on $X \times X$ whose marginals are equal. A theorem of Donsker and Varadhan (1976), which is precisely stated as part of Theorem 2.2, is that there is some element P^* of Λ_0 with marginals equal to μ^* which is naturally associated with $I(\mu^*)$. We define such a P^* to be an I-projection of π onto C. We establish the uniqueness of P^* in Theorem 2.3 under the additional assumptions that C is convex, C^0 is nonempty and $I(C^0) < \infty$.

The I-projection thus defined stands in clear relation to that of Csiszár, Cover and Choi (1987) in their study of second-order empirical distributions of a finite-state Markov chain. Theorem 2.9 of this paper, which identifies P^* for a convex set of C of interest, is a partial generalization of one of their examples (cf. Theorem 4 and the remarks following it). It is related to earlier results for finite-state Markov chains obtained by Justesen and Hoholdt (1984) and Spitzer (1972).

Let $\Omega = \prod_{i=-\infty}^{\infty} X_i$, $X_i = X$ for all i, and let $\mathscr{B}$ be the σ-field of Borel sets of Ω. As before, $(\Omega, \mathscr{B}) = \prod_{i=-\infty}^{\infty}(X_i, \mathscr{B}_i)$, where for each i, $\mathscr{B}_i$ is the σ-field of Borel sets on X. Ω is a compact space with metric

$$\rho(\omega, \omega') = \sum_{i=-\infty}^{\infty} \frac{1}{2^{|i|}} \frac{d(\omega_i, \omega'_i)}{1 + d(\omega_i, \omega'_i)},$$

where $d(\cdot, \cdot)$ is the metric on X.

Now for $\omega \in \Omega$, define ω_n by

$$\begin{aligned} \omega_n(i) &= \omega(i), && 0 \le i \le n-1, \\ \omega_n(i+n) &= \omega_n(i) && \text{for all } i,\ -\infty < i < \infty. \end{aligned}$$

Let $(\theta_i \omega_n)(j) = \omega_n(i+j)$, $0 \le i \le n-1$, and for a Borel set A in Ω define

$$R_{n,\omega}(A) = \frac{1}{n} \sum_{i=0}^{n-1} \chi_A(\theta_i \omega_n).$$

$R_{n,\omega}$ is the nth-order empirical distribution of ω_n. For each $\omega \in \Omega$ and $n > 0$, $R_{n,\omega}(\cdot)$ is a stationary process. Impose the weak topology on the set $M_S(\Omega)$ of stationary processes on $(\Omega, \mathscr{B})$. For each n, $R_{n,\omega}(\cdot)$: $\Omega \to M_S(\Omega)$ is a continuous map of $(\omega_0, \dots, \omega_{n-1})$. Let C be a closed convex set with nonempty interior satisfying $I(C^0) < \infty$. Let P^* be the I-projection of π onto C. Then P^* defines a stationary Markov process on $(\Omega, \mathscr{B})$ which we again denote by P^*.

We now add the assumption that $I(C) = I(C^0)$ so that we have the Markov process analog of the Sanov property. Lemma 3.1 proves that in terms of the metric for the weak topology on $M_S(\Omega)$, $R_{n,w}$ converges to P^* in conditional P_x-probability given $\hat{P}_n(\omega, \cdot) \in C$, uniformly for $x \in X$. Also for each A a Borel set in $\mathscr{B}$ and each $n > 0$, $R_{n,\omega}(A)$: $\Omega \to \mathbb{R}$ is a measurable function of $\omega_0, \dots, \omega_{n-1}$. Then it is possible to define stationary processes

$$R^C_{n,x}(\cdot) = E^{P_x}\{R_{n,\omega}(\cdot) | \hat{P}_n(\omega, \cdot) \in C\}.$$

Theorem 3.2 shows that the processes $R^C_{n,x}$ converge weakly to P^*.

Let $u(x)$ be a probability density function with respect to $\lambda(dx)$. Let P_u be the Markov process on $\prod_{i=0}^{\infty}(X_i, \mathscr{B}_i)$, $X_i = X$, $\mathscr{B}_i$ the σ-field of Borel sets of X, with initial distribution $u(x)\lambda(dx)$ and probability transition function $\pi(dy|x)$. The results of Section 3 imply that the measures

$$R_{n,u}^{C}(\cdot) = E^{P_u}\{R_{n,\omega}(\cdot)|\hat{P}_n(\omega,\cdot) \in C\} \tag{1.9}$$

converge weakly to P^* (cf. the remarks prior to Corollary 5.2). Suppose that each $(\omega_0, \omega_1, \omega_2, \dots)$ is a sequence of independent, identically distributed random variables with the common distribution $\lambda(dx)$, that is, $\pi(y|x) = 1$. Let $\mathscr{F}_m^n$ denote the sub-σ-field of $\mathscr{B}$ generated by ω_i, $n \le i \le m$. Suppose $B \in F_{n-1}^{0}$. Setting $u(x) = 1$,

$$E^{P_1}\{R_{n,\omega}(B)|\hat{P}_n(\omega,\cdot) \in C\} = P^{\lambda^n}\{B|\hat{P}_n(\omega,\cdot) \in C\},$$

where λ^n is the nth Cartesian power of $\lambda(dx)$ (cf. Lemma 4.3). Csiszár, Cover and Choi (1987), Theorem 1, show that for sequences of independent, identically distributed random variables on a finite set X, the joint distribution of $\omega_0, \omega_1, \dots, \omega_m$ under the condition $\hat{P}_n \in C$ converges to $(P^*)^m$ as $n \to \infty$, P^* the I-projection of λ on C. Thus the weak convergence established here in Section 3 is a generalization of this result to discrete parameter Markov processes on a compact state space.

We introduce the new definition *asymptotically quasi-Markov* as follows. A sequence of measures P_n on $(X_0, \dots, X_{n-1})$, $n = 1, 2, \dots$, is said to be asymptotically quasi-Markov if there exists a stationary transition probability function $Q(dy|x)$ such that

$$\lim_{n \to \infty} \frac{1}{n} h(P_n, \overline{Q_n}) = 0, \tag{1.10}$$

where $\overline{Q_n}$ is the probability measure on $(X_0, \dots, X_{n-1})$ defined by the transition probability function $Q(dy|x)$ with initial distribution given by the first marginal of P_n. In Lemma 4.4 a large deviations estimate is proved which establishes the asymptotic quasi-Markov property for certain sequences of measures. Suppose that the probability density function $u(x)$ is bounded from above. Then Theorem 4.5 establishes that the measures $R_{n,u}^{C}$ defined by (1.9) on $\prod_{i=0}^{n}(X_i \mathscr{B}_i)$ give a sequence which is asymptotically quasi-Markov with respect to the transition probability function of P^* which is uniquely defined a.e. λ. When the sequence of measures P_n in the definition comes from the restriction of stationary processes R_n on Ω to $\prod_{i=0}^{n-1}(X_i, \mathscr{B}_i)$, as is the case with the measures $R_{n,u}^{C}$, the asymptotic quasi-Markov property with respect to the transition probability function $P^*(dy|x)$ is a stronger property than the weak convergence of R_n to P^*. The sense of this is made precise in Corollary 5.2.

Under the additional assumption that $u(x)$ is bounded away from 0, Corollary 5.3 shows that the conditional P_u distribution of $X_0, \dots, X_{n-1}$ under the condition $\hat{P}_n(\omega, \cdot) \in C$ is asymptotically quasi-Markov with respect to the probability transition function $P^*(dy|x)$. This is a generalization of Csiszár's limit (1.4) for independent, identically distributed random variables to Markov processes on a compact metric space. These conditional measures do not enjoy the same properties as the measures $R^C_{n,u}$ described above, so the implications of the asymptotic quasi-Markov property are less significant. However, one consequence is as follows. Let

$$P_n(\cdot) = P_u\{(\cdot)|\hat{P}_n(\omega,\cdot) \in C\}. \tag{1.11}$$

Let $\overline{Q_n}$ be the Markov process with transition probability function $P^*(dy|x)$ and initial distribution given by the first marginal of P_n. Then if $B_n \in \mathscr{F}^0_{n-1}$,

$$\lim_{n\to\infty} P_n(B_n) = 0$$

if

$$\overline{Q_n}(B_n) \le \exp(-\alpha n), \qquad n = 1, 2, \dots,$$

for some $\alpha > 0$. This follows from (1.10) since (1.3) implies that

$$P_n(B_n)\log\frac{P_n(B_n)}{\overline{Q_n}(B_n)} + (1 - P_n(B_n))\log\frac{1 - P_n(B_n)}{1 - \overline{Q_n}(B_n)} \le h(P_n, \overline{Q_n}).$$

2. *I*-Projection of π onto *C*. Let C be a closed set of $\mathscr{M}(X)$. Let M_C be the subset of Λ_0 whose marginals are equal to an element of C. For $Q \in \Lambda_0$, define

$$h^1(Q|\pi) = h(Q, \overline{P}),$$

where $\overline{P}$ is the measure $q(dx)\pi(dy|x)$, $q(dx)$ the marginal of Q. An I-projection P^* of π onto C is defined as an element of M_C for which

$$\inf_{Q\in M_C} h^1(Q|\pi) = h^1(P^*|\pi).$$

Donsker and Varadhan (1975), Lemma 2.1, show that for P and Q probability measures on a Polish space X with σ-field given by the Borel sets, then $h(Q, P)$ defined by (1.2) can alternatively be defined as

$$h(Q,P) = \sup_{\mu\in\mathscr{U}_1}\left[\int_X \log u(x)Q(dx) - \log\int_X u(x)P(dx)\right], \tag{2.1}$$

where $\mathscr{U}_1$ is the set of continuous functions on X for which there exist constants c_1 and c_2 such that $0 < c_1 \le u(x) \le c_2 < \infty$. In particular, for fixed P, $h(Q, P)$ is a lower semicontinuous function of Q in the weak topology on $\mathscr{M}(X)$ and for fixed Q, $h(Q, P)$ is a lower semicontinuous function of P.

Lemma 2.1. *Under assumption* 3 *on* $\pi(dy|x)$, $h^1(Q|\pi)$ *is a lower semicontinuous, convex function of* Q.

Proof. Let $q(dx)$ denote the marginal of Q. Then from (2.1),

$$h^1(Q|\pi) = \sup_{u\in\mathscr{U}_1}\left[\iint_{X\times X}\log u(x,y)Q(dx,dy) - \log\iint_{X\times X} u(x,y)q(dx)\pi(dy|x)\right],$$

where $\mathscr{U}_1$ is the set of continuous functions $u(x, y) > 0$ on $X \times X$. For each $u \in \mathscr{U}_1$,

$$\iint_{X\times X} \log u(x,y)Q(dx,dy) - \log \iint_{X\times X} u(x,y)q(dx)\pi(dy|x)$$

is a continuous, convex function of Q. The lemma follows. □

For C a closed set, M_C is a compact set of probability distributions on $X \times X$, and the existence of an I-projection of π onto M_C follows from Lemma 2.1. To relate an I-projection as defined above to $I(C)$ defined by (1.8) requires a result of Donsker and Varadhan (1976), Theorem 2.1. This and a further result of theirs that will be required [Donsker and Varadhan (1976), Lemma 2.5] are stated as the following theorem.

THEOREM 2.2. *Let $(X, \mathscr{B})$ be a Polish space with $\mathscr{B}$ the σ-field of Borel sets. Let M_μ be the set of probability measures on $X \times X$ having both marginals μ. If $\pi(dy|x)$ is the transition probability function of a discrete parameter Markov process with $(X, \mathscr{B})$ as state space, then*

$$I(\mu) = \inf_{P \in M_\mu} h^1(P|\pi).$$

Suppose that $\pi(dy|x)$ satisfies assumption 3, so that there is a $P \in M_\mu$ for which the infinimum is actually achieved. If there exists a reference measure λ on X such that $\pi(dy|x) = \pi(y|x)\lambda(dy)$, if $I(\mu) < \infty$ and $\pi(y|x) > 0$ a.e. $\mu \times \mu$, then there are measurable functions $a(x)$ and $b(y)$ such that

$$P(dx, dy) = \frac{a(x)}{b(y)}\pi(y|x)u(dx)\lambda(dy),$$

where $0 \le a(x) < \infty$ a.e. μ and $0 < b(y)$ a.e. λ.

Now

$$\begin{aligned}\inf_{Q\in M_C} h^1(Q|\pi) &= \inf_{\mu\in C}\inf_{Q\in M_\mu} h^1(Q|\pi)\\ &= \inf_{\mu\in C} I(\mu)\\ &= I(C).\end{aligned}$$

For C a closed set, an I-projection P^* of π onto C is an element of M_C satisfying

$$I(C) = h^1(P^*|\pi).$$

The marginals of P^* minimize $I(\mu)$ for $\mu \in C$.

In this section we establish the following theorem.

THEOREM 2.3. *Let C be a closed convex set with nonempty interior C^0. Suppose that $I(C^0) < \infty$. Then an I-projection of π onto C is unique. It is a measure P^* having probability density $P^*(x, y)$ with respect to $\lambda \times \lambda$ which is* *positive a.e.* $\lambda \times \lambda$. *Further, for any* $Q \in M_C$,

$$h^1(Q|\pi) \geq h^1(Q|P^*(\cdot|\cdot)) + h^1(P^*|\pi).$$

For the proof of Theorem 2.3, we establish the following lemmas.

LEMMA 2.4. *Let* $Q \in M_q$. *If* $h^1(Q|\pi) < \infty$, *then* $Q(dx, dy) \ll \lambda \times \lambda$ *and*

$$h^1(Q|\pi) = \iint_{X\times X} Q(x,y)\log\frac{Q(x,y)}{q(x)\pi(y|x)}\lambda(dx)\lambda(dy),$$

where $Q(x, y)$ *is the density* Q *with respect to* $\lambda \times \lambda$ *and* $q(x)$ *is the density of* q *with respect to* λ.

PROOF. Since $h^1(Q|\pi) < \infty$, it follows that $Q(dx, dy) \ll q(dx)\pi(y|x)\lambda(dy)$. Further from Theorem 2.2, $I(q) < \infty$. It can be shown [cf. the proof of Lemma 4.1 in Donsker and Varadhan (1975)] that $I(q) < \infty$ implies that $q \ll \lambda$. Then $Q(dx, dy) \ll \lambda \times \lambda$ and the rest of the lemma follows from (1.2). □

LEMMA 2.5. *Let* C *be a measurable set in* $\mathscr{M}(X)$ *such that* C^0 *is nonempty and* $I(C^0) < \infty$. *Then there is a measure* Q *in* M_C *with a positive density with respect to* $\lambda \times \lambda$ *satisfying* $h^1(Q|\pi) < \infty$.

PROOF. Let $\mu \in C^0$ satisfy $I(\mu) < \infty$. Then $\mu \ll \lambda$. Let

$$\mu_n = (1 - 1/n)\mu + (1/n)\lambda.$$

Then $d\mu_n/d\lambda \geq 1/n$ λ-a.e. The sequence μ_n converges in variation to μ so for sufficiently large n, $\mu_n \in C^0$. Let $\bar{\mu}$ be such a μ_n and suppose $d\bar{\mu}/d\lambda = m(x) \geq \eta$. Let

$$m_n(x) = \frac{m(x) \wedge n}{\int_X[m(x) \wedge n]\lambda(dx)},$$

where n is chosen greater than or equal to η and so large that $\int_X[m(x) \wedge n]\lambda(dx) \geq 1/2$. Let $\bar{\mu}_n(dx) = m_n(x)\lambda(dx)$. Then $\bar{\mu}_n$ converges in variation to $\bar{\mu}$ so for sufficiently large n, $\bar{\mu}_n \in C^0$. By construction, $\eta \leq m_n(x) \leq 2n$. Let $\nu(dx)$ be such an element $\bar{\mu}_n$ and let $d\nu/d\lambda = u(x)$. Define $Q(dx, dy) = u(x)u(y)\lambda(dx)\lambda(dy)$. By the bounds on $u(x)$ and $\pi(y|x)$, it follows that

$$h^1(Q|\pi) = \iint_{X\times X} u(x)u(y)\log\frac{u(y)}{\pi(y|x)}\lambda(dx)\lambda(dy) < \infty. \qquad \square$$

LEMMA 2.6. *Let* C *be a closed convex set with nonempty interior* C^0 *satisfying* $I(C^0) < \infty$. *Suppose* $P^* \in M_C$ *is such that*

$$I(C) = h^1(P^*|\pi).$$

Then P^* *has a density* $P^*(x, y)$ *with respect to* $\lambda \times \lambda$ *which is positive a.e.* $\lambda \times \lambda$ *and for any* $Q \in M_C$,

$$h^1(Q|\pi) \geq h^1(Q|P^*(\cdot|\cdot)) + h^1(P^*|\pi).$$

PROOF. Suppose $P \in M_C$ satisfies $h^1(P|\pi) < \infty$. Consider $h^1(\varepsilon P + (1-\varepsilon)P^*|\pi)$, $0 \le \varepsilon \le 1$. By the convexity of $h^1(Q|\pi)$ as a function of Q, this is a convex function of ε. Since $h^1(P^*|\pi) \le h^1(\varepsilon P + (1-\varepsilon)P^*|\pi)$, $h^1(\varepsilon P + (1-\varepsilon)P^*|\pi)$ is a nondecreasing function of ε. Then

$$\lim_{\varepsilon \to 0} \frac{d}{d\varepsilon} h^1(\varepsilon P + (1-\varepsilon)P^*|\pi) \ge 0, \tag{2.2}$$

provided the derivatives exist.

For $Q \in M_C$, let q denote its marginal. Then $I(q) \le h^1(Q|\pi)$ so that if $h^1(Q|\pi)$ is finite, so is $I(q)$. It is an easy consequence of the bounds on $\pi(\cdot|\cdot)$ and (1.6) that

$$h(q,\lambda) - \log A \le I(q) \le h(q,\lambda) - \log a$$

[Donsker and Varadhan (1975), Lemma 2.8] so that if $I(q) < \infty$, $h(q,\lambda) < \infty$.

Writing $P_\varepsilon(x,y) = \varepsilon P(x,y) + (1-\varepsilon)P^*(x,y)$ and using the same notation for the marginals,

$$\begin{aligned} h^1(\varepsilon P + (1-\varepsilon)P^*|\pi) &= \iint_{X\times X} P_\varepsilon(x,y) \log \frac{P_\varepsilon(x,y)}{p_\varepsilon(x)\pi(y|x)} \lambda(dx)\lambda(dy) \\ &= \iint_{X\times X} P_\varepsilon(x,y)\log\frac{P_\varepsilon(x,y)}{\pi(y|x)}\lambda(dx)\lambda(dy) \\ &\quad - \int_X p_\varepsilon(x)\log p_\varepsilon(x)\lambda(dx). \end{aligned} \tag{2.3}$$

Further, for $0 \le \varepsilon \le 1$,

$$\iint_{X\times X} P_\varepsilon(x,y)\left|\log\frac{P_\varepsilon(x,y)}{\pi(y|x)}\right|\lambda(dx)\lambda(dy) < \infty, \tag{2.4)(i}$$

$$\int_X p_\varepsilon(x)|\log p_\varepsilon(x)|\lambda(dx) < \infty. \tag{2.4)(ii}$$

For each integral in (2.3) a derivative exists for $0 < \varepsilon < 1$. Consider the first integral. Here

$$\begin{aligned} \frac{d}{d\varepsilon} P_\varepsilon(x,y)\log\frac{P_\varepsilon(x,y)}{\pi(y|x)} &= (P(x,y) - P^*(x,y))\log\frac{P_\varepsilon(x,y)}{\pi(y|x)} \\ &\quad + P(x,y) - P^*(x,y). \end{aligned}$$

Using the bounds

$$\begin{aligned} \log^+ \frac{P_\varepsilon(x,y)}{\pi(y|x)} &= \log^+\left[\varepsilon\frac{P(x,y)}{\pi(y|x)} + (1-\varepsilon)\frac{P^*(x,y)}{\pi(y|x)}\right] \\ &\le \log^+\left[\frac{P(x,y)}{\pi(y|x)} + \frac{P^*(x,y)}{\pi(y|x)}\right] \\ &\le \log 2 + \log^+\left[\frac{P_{1/2}(x,y)}{\pi(y|x)}\right] \end{aligned} \tag{2.5)(i}$$

and

$$\log^{-}\frac{P_\varepsilon(x,y)}{\pi(y|x)} = \log^{-}\left[\varepsilon\frac{P(x,y)}{\pi(y|x)} + (1-\varepsilon)\frac{P^*(x,y)}{\pi(y|x)}\right]$$

$$(2.5)(\text{ii}) \qquad \le \log^{-}\left[\min(\varepsilon, 1-\varepsilon)\left(\frac{P(x,y)}{\pi(y|x)} + \frac{P^*(x,y)}{\pi(y|x)}\right)\right]$$

$$\le \log^{-}(2\min(\varepsilon,1-\varepsilon)) + \log^{-}\left[\frac{P_{1/2}(x,y)}{\pi(y|x)}\right]$$

combined with (2.4)(i) shows by dominated convergence that the derivative can be taken inside the integral sign and is $L^1(\lambda\times\lambda)$. Thus

$$\frac{d}{d\varepsilon}\int\int_{X\times X}P_\varepsilon(x,y)\log\frac{P_\varepsilon(x,y)}{\pi(y|x)}\lambda(dx)\lambda(dy)$$
$$= \int\int_{X\times X}(P(x,y)-P^*(x,y))\log\frac{P_\varepsilon(x,y)}{\pi(y|x)}\lambda(dx)\lambda(dy).$$

Arguing similarly with the second integral in (2.3) shows that

$$\frac{d}{d\varepsilon}h^1(\varepsilon P + (1-\varepsilon)P^*|\pi)$$

$$(2.6) \qquad = \int\int_{X\times X}(P(x,y)-P^*(x,y))\log\frac{P_\varepsilon(x,y)}{\pi(y|x)}\lambda(dx)\lambda(dy)$$
$$-\int_X(p(x)-p^*(x))\log p_\varepsilon(x)\lambda(dx).$$

Now using the bound (2.5)(i) and the bound

$$\log^{-}\frac{P_\varepsilon(x,y)}{\pi(y|x)} \le \log^{-}\left[(1-\varepsilon)\frac{P^*(x,y)}{\pi(y|x)}\right]$$
$$\le \log^{-}(1-\varepsilon) + \log^{-}\left(\frac{P^*(x,y)}{\pi(y|x)}\right)$$

shows by dominated convergence that

$$\lim_{\varepsilon\to 0}\int\int_{X\times X}P^*(x,y)\log\frac{P_\varepsilon(x,y)}{\pi(y|x)}\lambda(dx)\lambda(dy)$$
$$= \int\int_{X\times X}P^*(x,y)\log\frac{P^*(x,y)}{\pi(y|x)}\lambda(dx)\lambda(dy).$$

Similarly,

$$\lim_{\varepsilon\to 0}\int_X p^*(x)\log p_\varepsilon(x)\lambda(dx) = \int_X p^*(x)\log p^*(x)\lambda(dx).$$

It follows from (2.6) and (2.2) that

$$\lim_{\varepsilon\to 0}\iint_{X\times X} P(x,y)\log\frac{P_\varepsilon(x,y)}{p_\varepsilon(x)\pi(y|x)}\lambda(dx)\lambda(dy)$$
$$\geq \iint_{X\times X} P^*(x,y)\log\frac{P^*(x,y)}{p^*(x)\pi(y|x)}\lambda(dx)\lambda(dy).$$

Rewriting the integral on the left-hand side as

$$\int_{X\times X} P(x,y)\log\frac{P(x,y)}{p(x)\pi(y|x)}\lambda(dx)\lambda(dy)$$
$$-\iint_{X\times X} P(x,y)\log\frac{P(x,y)}{p(x)P_\varepsilon(y|x)}\lambda(dx)\lambda(dy)$$

shows that

$$\lim_{\varepsilon\to 0}\int_{X\times X} P(x,y)\log\frac{P(x,y)}{p(x)P_\varepsilon(y|x)}\lambda(dx)\lambda(dy) \tag{2.7}$$
$$\leq h^1(P|\pi) - h^1(P^*|\pi).$$

We can write the integral on the left-hand side of (2.7) as

$$\int_X p(x)\lambda(dx)h(P(dy|x),P_\varepsilon(dy|x)),$$

where $P_\varepsilon(dy|x) = P_\varepsilon(y|x)\lambda(dy)$. Clearly $h(P(dy|x), P_\varepsilon(dy|x))$ is defined for $p(dx)$ a.e. x.

By Fatou's lemma,

$$\int_X p(x)\lambda(dx)\liminf_{\varepsilon\to 0}(h(P(dy|x),P_\varepsilon(dy|x))) \tag{2.8}$$
$$\leq \lim_{\varepsilon\to 0}\iint_{X\times X} P(x,y)\log\frac{P(x,y)}{p(x)P_\varepsilon(y|x)}\lambda(dx)\lambda(dy).$$

Now on the set of $p(dx)$ measure 1 where $P_\varepsilon(dy|x)$ is defined, we see that as $\varepsilon\to 0$, $P_\varepsilon(dy|x)$ converges in variation to $P^*(dy|x)$ when $p^*(x) > 0$ and $P_\varepsilon(dy|x)$ is $P(dy|x)$ when $p^*(x) = 0$. By the lower semicontinuity of $h(Q,P)$ as a function of P for fixed Q, it follows that

$$h\Big(P(dy|x), \lim_{\varepsilon\to 0} P_\varepsilon(dy|x)\Big)$$
$$= \begin{cases} h(P(dy|x),P^*(dy|x)), & p^*(x) > 0,\\ h(P(dy|x),P(dy|x)) = 0, & p^*(x) = 0,\end{cases} \tag{2.9}$$
$$\leq \liminf_{\varepsilon\to 0}(h(P(dy|x),P_\varepsilon(dy|x))).$$

Let E be the set in X, where $p^*(x) > 0$. In view of (2.7) and (2.8) it follows that $P(dy|x) \ll P^*(dy|x)$ for $p(dx)$ a.e. x in E. Since by Lemma 2.5 there is a

Q in M_C with a positive density with respect to $\lambda \times \lambda$ satisfying $h^1(Q|\pi) < \infty$, it follows that $P^*(y|x) > 0$ for $\lambda \times \lambda$ a.e. (x, y) in $E \times X$. Then $\int_E p^*(x)P^*(y|x)\lambda(dx) = p^*(y)$ for λ-a.e. y so that $p^*(y) > 0$ for λ-a.e. y. It follows that $E = X/N$ where N is a $\lambda(dx)$ null set. This establishes the positivity of $P^*(x, y)$ a.e. $\lambda \times \lambda$.

To conclude the proof of Lemma 2.6, it follows from (2.7), (2.8) and (2.9) that

$$h^1(P|P^*(\cdot|\cdot)) \le h^1(P|\pi) - h^1(P^*|\pi)$$

or

$$h^1(P|\pi) \ge h^1(P|P^*(\cdot|\cdot)) + h^1(P^*|\pi) \tag{2.10}$$

for $P \in M_C$ satisfying $h^1(P|\pi) < \infty$. For the general case, $P^*(dy|x)$ is only defined λ-a.e. x. Extend it arbitrarily to make it a transition probability satisfying $P^*(dy|x) = P^*(y|x)\lambda(dy)$. Then $h^1(P|P^*(\cdot|\cdot)) < \infty$ for $P \in M_C$ implies the marginal p of P satisfies $p \ll \lambda$ so $h^1(P|P^*(\cdot|\cdot))$ is well defined. Since (2.10) is obviously true if $h^1(P|\pi) = \infty$, the lemma is established. □

COROLLARY 2.7. *Let C and P^* be as in Lemma 2.4. Then an I-projection P^* of π onto C is unique.*

PROOF. Suppose that P_1^* and P_2^* both satisfy

$$I(C) = h^1(P_1^*|\pi) = h^1(P_2^*|\pi).$$

It follows from (2.10) that

$$h^1(P_2^*|\pi) \ge h^1(P_2^*|P_1^*(\cdot|\cdot)) + h^1(P_1^*|\pi).$$

Then $h^1(P_2^*|P_1^*(\cdot|\cdot)) = 0$ or $h^1(P_1^*|\pi)$ would be strictly less than $h^1(P_2^*|\pi)$. Since

$$h^1(P_2^*|P_1^*(\cdot|\cdot)) = \int_X p_2^*(dx)\int P_2^*(dy|x)\log\frac{P_2^*(y|x)}{P_1^*(y|x)},$$

it follows, using the fact that $p_2^*(x) > 0$ a.e. $\lambda(dx)$ that

$$P_2^*(dy|x) = P_1^*(dy|x), \quad \lambda \text{ a.e. } x. \tag{2.11}$$

Extend $P_1^*(dy|x)$ arbitrarily to make it a transition probability satisfying $P_1^*(dy|x) = P_1^*(y|x)\lambda(dy)$. Let I^* be the I-function with transition probability $P_1^*(dy|x)$. Then from Theorem 2.2,

$$I^*(p_2^*) \le h^1(P_1^*|P_1^*(\cdot|\cdot)) = 0,$$

so that $p_2^*(dx)$ is an invariant measure for the transition probability $P_1^*(dy|x)$ [Donsker and Varadhan (1975), Lemma 4.1]. P_1^* defines a stationary Markov process on $(\Omega, \mathscr{B})$ with transition probability function $P_1^*(dy|x)$ and invariant measure $p_1^*(dx)$. Using the positivity of $P_1^*(y|x)$ $\lambda \times \lambda$-a.e., the P_1^* process is ergodic. Then the ergodic theorem and the positivity of $p_1^*(x)$ a.e. λ ensure that the transition probability function $P^*(dy|x)$ has a unique invariant measure [Harris (1956), Theorem 1]. Then $p_1^*(dx) = p_2^*(dx)$ which in view of (2.11) implies $P_1^* = P_2^*$. □

COROLLARY 2.8. *Let P^* be the I-projection of π onto C as above. Then $P^*(dy|x)$ can be chosen to have the following property: There is a function $b(y)$, $0 < b(y) < \infty$ λ-a.e. y such that if $h(y) \in L^1(1/b(y)\lambda(dy))$, then*

$$\int_X P^*(dy|x)h(y) \tag{2.12}$$

is a continuous function of x.

PROOF. It follows from Theorem 2.2 and Lemma 2.6 that there are measurable functions $a(x)$ and $b(y)$ such that

$$P(x,y) = p(x)\frac{a(x)}{b(y)}\pi(y|x) \quad \text{a.e. } \lambda \times \lambda,$$

where $0 < a(x) < \infty$ a.e. $\lambda(dx)$ and $0 < b(y) < \infty$ a.e. $\lambda(dy)$. Then

$$a(x)p(x)\int_X \frac{\pi(y|x)}{b(y)}\lambda(dy) = p(x) \quad \text{a.e. } \lambda(dx).$$

Since $p(x) > 0$ a.e. $\lambda(dx)$,

$$\int \frac{\pi(y|x)}{b(y)}\lambda(dy) = \frac{1}{a(x)} \quad \text{a.e. } \lambda(dx).$$

By altering $a(x)$ on a set of measure 0, it is possible to assume that this equation holds for all x. Since $a \le \pi(y|x)$ for all x and λ-a.e. y and since $P(x,y) \in L^1(\lambda \times \lambda)$, it follows from Fubini's theorem that $1/b(y) \in L^1(\lambda(dy))$. It then follows from assumption 3 on $\pi(y|x)$ that $a(x)$ is continuous. Now define

$$P^*(y|x) = \frac{a(x)}{b(y)}\pi(y|x).$$

Then if $h(y) \in L^1(1/b(y)\lambda(dy))$,

$$\int_X P^*(dy|x)h(y) = a(x)\int_X \pi(y|x)h(y)\frac{1}{b(y)}\lambda(dy).$$

Again using assumption 3 on $\pi(y|x)$, this is a continuous function of x. □

Consider a somewhat stronger continuity assumption on $\pi(dy|x)$ than assumption 3:

4. $\pi(y|x)$ as a map from $x \to L_1(\lambda)$ is continuous.

Under assumption 4, Theorem 2.9 sharpens the results of Corollary 2.8 in a case of interest. It is a partial generalization of an example of Csiszár, Cover and Choi (1987), as explained in Section 1.

THEOREM 2.9. *Let* $C = \{\mu \in \mathscr{M}(X): \int_X f_i d_\mu \geq \gamma_i,\ i = 1, \dots, n\}$ *for continuous, real-valued functions* $f_1, f_2, \dots, f_n$ *on* X. *Suppose there is some* $\mu \in C$ *satisfying*

$$\text{(2.13)}\qquad \begin{aligned} &\text{(a)}\quad \int_X f_i d_\mu > \gamma_i \qquad i = 1, \dots, n, \\ &\text{(b)}\quad I(\mu) < \infty. \end{aligned}$$

Assume the transition probability density $\pi(y|x)$ *satisfies assumption* 4.

Let $\mathbb{R}_n^+$ *denote* $\{\zeta \in \mathbb{R}^n,\ \zeta_i \geq 0,\ i = 1, \dots, n\}$. *Let* T_ζ *be the mapping of the set of continuous functions on* X, $C(X)$, *onto itself given by*

$$T_\zeta g(x) = e^{\sum_{i=1}^n \zeta_i f_i(x)} \int_X g(y) \pi(y|x) \lambda(dy).$$

Let V_ζ *be the unique positive eigenvector for* T_ζ *and let* $\psi_\zeta \in L^1(\lambda)$ *be the unique almost everywhere positive eigenvector for* T_ζ^*, *the adjoint of* T_ζ, *corresponding to the same positive eigenvalue* p_ζ, *which is greatest in modulus of all the eigenvalues of* T_ζ. *Assume* V_ζ *and* ψ_ζ *have been normalized so that*

$$\text{(2.14)}\qquad \int_X V_\zeta(x) \psi_\zeta(x) \lambda(dx) = 1.$$

Then

$$\text{(2.15)}\qquad I(C) = \max_{\zeta \in \mathbb{R}_n^+} \left(\sum_{i=1}^n \zeta_i \gamma_i - \log \rho_\zeta \right)$$

and

$$P^*(x, y) = \left(V_\zeta(y) \pi(y|x) e^{\sum_{i=1}^n \zeta_i f_i(x)} \psi_\zeta(x) \right) / \rho_\zeta$$

for ζ *attaining the maximum in* (2.15). *Further* $I(C) = I(C^0)$, *so the analog of the Sanov property holds.*

Given that $\pi(y|x)$ is a transition probability density, assumption 4 is a necessary and sufficient condition for T_ζ to be a compact operator [Edwards (1965), Proposition 9.5.17]. The lower bound (assumption 2) on $\pi(y|x)$ and the continuity of f_i, $i = 1, \dots, n$, ensure that T_ζ is a strongly positive operator, so a theorem of Krein and Rutman (1948), Theorem 6.3, proves the existence of V_ζ, ψ_ζ and ρ_ζ as in the statement of the theorem.

The set C described in the theorem is weakly closed and convex. Any measure μ satisfying (2.13)(a) is an element of the interior C^0. In particular, the hypotheses of Theorem 2.3 are satisfied, so an I-projection of π onto C exists and is unique. To see $I(C^0) = I(C)$, suppose μ satisfies (2.13)(a) and (b) and let ν_1 be any element of C. Then $\nu_\alpha = (1 - \alpha)\mu + \alpha\nu_1 \in C^0$ and by convexity,

$$\begin{aligned} \limsup_{\alpha \to 1} I(\nu_\alpha) &\leq \limsup_{\alpha \to 1} \left[(1 - \alpha) I(\mu) + \alpha I(\nu_1) \right] \\ &= I(\nu_1). \end{aligned}$$

The remainder of the theorem is proved in a sequence of four lemmas.

LEMMA 2.10. *Let P^* be the unique I-projection of π onto C. Then*

$$I(C) = h^1(P^*|\pi) = \inf_{Q \in M_C} h(Q, p^*(dx)\pi(dy|x)), \tag{2.16}$$

where p^* is the marginal distribution of P^*.

PROOF. Equation (2.16) means that P^* is the I-projection onto M_C of the two-dimensional measure $p^*(dx)\pi(dy|x)$ as defined by Csiszár (1975). To establish (2.16), suppose $Q \in M_C$ satisfies $h(Q, p^*(dx)\pi(dy|x)) < \infty$. Equation (2.1) implies the one-dimensional divergence $h(q, p^*) < \infty$ for the marginal q of Q. From (1.2) there follows

$$h(Q, p^*(dx)\pi(dy|x)) = h^1(Q|\pi) + h(q, p^*).$$

From (2.10),

$$h(Q|\pi) \geq h^1(Q|P^*(\cdot\mid\cdot)) + h^1(P^*|\pi).$$

Adding $h(q, p^*)$ to both sides and using $h(Q, P^*) = h^1(Q|P^*(\cdot\mid\cdot)) + h(q, p^*)$ shows

$$\begin{aligned} h(Q, p^*(dx)\pi(dy|x)) &\geq h(Q, P^*) + h^1(P^*|\pi) \\ &\geq h(Q, P^*) + \inf_{Q \in M_C} h(Q, p^*(dx)\pi(dy|x)). \end{aligned}$$

This last equation determines P^* as the unique I-projection of $p^*(dx)\pi(dy|x)$ onto M_C [Csiszár (1975), Theorem 2.2]. □

LEMMA 2.11. *Let $f_i(x)$, $i = 1, 2, \ldots, n$, be real-valued measurable functions on a measure space $(X, \mathscr{B}, \lambda)$. Then the convex cone $K = \sum_{i=1}^n \alpha_i f_i(x)$, $\alpha_i \geq 0$, is closed in the topology of pointwise sequential convergence on the space of real-valued measurable functions.*

PROOF. First suppose that the functions $f_i(x)$, $i = 1, 2, \ldots, n$, are linearly independent λ-a.e. Suppose there exists a sequence $\sum_{i=1}^n \alpha_{m_i} f_i(x)$ converging pointwise λ-a.e. to a real-valued function $g(x)$. Let c_{m_i} be the sequence

$$\begin{aligned} (\alpha_{1_i} - \alpha_{2_i}, \alpha_{1_i} - \alpha_{3_i}, \alpha_{1_i} - \alpha_{4_i}, \ldots, \alpha_{2_i} - \alpha_{3_i}, \alpha_{2_i} - \alpha_{4_i}, \ldots, \\ \alpha_{n_i} - \alpha_{n+1_i}, \alpha_{n_i} - \alpha_{n+2_i}, \ldots). \end{aligned}$$

Then $\lim_{m\to\infty} \sum_{i=1}^n c_{m_i} f_i(x) = 0$. Let

$$c'_{m_i} = c_{m_i} / \max(|c_{m_1}|, |c_{m_2}|, \ldots, |c_{m_n}|, 1). \tag{2.17}$$

Then $\lim_{m\to\infty} \sum_{i=1}^n c'_{m_i} f_i(x) = 0$ and $|c'_{m_i}| \leq 1$ for $i = 1, \ldots, n$. However, any subsequential limit of the vector-valued sequence $\{\mathbf{c}'_m\}$ is 0 from the linear independence of the functions $f_i(x)$, $i = 1, \ldots, n$, λ-a.e. It follows that the sequence $\{\mathbf{c}'_m\}$ converges to 0. From the definition (2.17), it follows that the sequence $\{\mathbf{c}_m\}$ converges to 0. But then $\{\alpha_{m_i}\}$ is Cauchy for each $i = 1, \ldots, n$, which concludes the proof in this case.

For the general case, suppose that k is the dimension of the real linear subspace spanned by $\{f_i\}_{i=1}^n$. Then any element h of the convex cone K can be written as a nonnegative linear combination of some subcollection of k linearly independent functions of $\{f_1, f_2, \dots, f_n\}$. This follows exactly as in the proof of Carathéodory's theorem [Rockafellar (1970), Theorem 17.1]. Thus K is a finite union of sets each of which is closed in the topology of pointwise sequential convergence. □

LEMMA 2.12. *Let V_ζ, ψ_ζ and ρ_ζ be defined as in the statement of Theorem 2.9. Then the I-projection P^* of π onto C has density*

$$P^*(x,y) = \left(V_\zeta(y)\pi(y|x)e^{\sum_{i=1}^n \zeta_i f_i(x)}\psi_\zeta(x)\right)/\rho_\zeta$$

with respect to $\lambda \times \lambda$ for some $\zeta \in \mathbb{R}_n^+$.

PROOF. Lemma 2.10 shows that P^* is the I-projection on $p^*(dx)\pi(dy|x)$ onto M_C. Let $\{g_k\}_{k=1}^\infty$ be a countable dense collection of continuous functions on X. Then M_C can be described as the set of all measures on $X \times X$:

$$\left\{P: \int\int_{X\times X} f\,dP \geq 0,\ f \in \mathscr{F}\right\},$$

where $\mathscr{F}$ is the convex cone generated by nonnegative finite linear combinations of

$$\{h_j(x,y)\} = \{f_i(x) - \gamma_i\}_{i=1}^n \cup \{\pm(g_k(x) - g_k(y))\}_{k=1}^\infty.$$

It now follows from Csiszár (1984), Lemma 3.4, that $\log(P^*(y|x)/\pi(y|x)) - I(C)$ belongs to the $L^1(P^*)$-closure of $\mathscr{F}$. Since $P^* \sim \lambda \times \lambda$, there exist functions

$$\sum_{i=1}^n \alpha_{m_i}(f_i(x) - \gamma_i) + \sum_{k=1}^{M_m} \beta_{m_k}(g_k(x) - g_k(y)),$$

$\alpha_{m_i} \geq 0$ and $\beta_{m_k} \in \mathbb{R}$ which converge in $\lambda \times \lambda$ measure to $\log P^*(y|x)/\pi(y|x) - I(C)$. It follows from Donsker and Varadhan (1975), Lemma 2.3, that there is a subsequence $(\overline{m})$ and a sequence of constants $(a_{\overline{m}})$ so that

$$\text{(2.18)(i)} \quad \lim_{\overline{m}\to\infty}\left(\sum_{i=1}^n \alpha_{\overline{m}_i}(f_i(x) - \gamma_i) + \sum_{k=1}^{M_{\overline{m}}} \beta_{\overline{m}_k} g_k(x) - a_{\overline{m}}\right) = f(x)$$

exists for λ-a.e. x and

$$\text{(2.18)(ii)} \quad \lim_{\overline{m}\to\infty}\left(-\sum_{k=1}^{M_{\overline{m}}} \beta_{\overline{m}_k} g_k(y) + a_{\overline{m}}\right) = g(y)$$

exists for λ-a.e. y and $\log(P^*(y|x)/\pi(y|x)) - I(C) = f(x) + g(y)$ for $\lambda \times \lambda$-a.e. (x, y). Comparing (2.18)(i) and (2.18)(ii), it follows that for λ-a.e. x,

$$\lim_{\overline{m}\to\infty} \sum_{i=1}^n \alpha_{\overline{m}_i}(f_i(x) - \gamma_i) = f(x) + g(x).$$

In view of Lemma 2.11, there exist constants ζ_i, $i = 1, \dots, n$, $\zeta_i \geq 0$ such that

$$\sum_{i=1}^{n} \zeta_i (f_i(x) - \gamma_i) = f(x) + g(x), \quad \lambda\text{-a.e.}$$

Thus the conditional density satisfies

$$P^*(y|x) = e^{I(C)} e^{\sum_{i=1}^n \zeta_i (f_i(x) - \gamma_i)} e^{-g(x)} e^{g(y)} \pi(y|x), \quad \lambda \times \lambda\text{-a.e.}$$

Since $P^*(y|x)$ is a transition probability density function, there follows

$$e^{\sum_{i=1}^n \zeta_i f_i(x)} \int_X e^{g(y)} \pi(y|x) \lambda(dy) = e^{-I(C)} e^{\sum_{i=1}^n \zeta_i \gamma_i} e^{g(x)}, \quad \lambda\text{-a.e.}$$

Redefine $g(x)$ on a set of measure 0 so that the equation is valid for all x. Using assumption 3 on $\pi(y|x)$, $g(x)$ is continuous. Then $e^{g(x)}$ is the unique positive eigenvector for T_ζ^* with positive eigenvalue

$$(2.19) \qquad \rho_\zeta = \exp\left(\sum_{i=1}^{n} \zeta_i \gamma_i - I(C) \right).$$

By definition of $P^* \in M_C$, the *I*-projection P^* has identical marginals. Letting $p^*(x)$ be the density of the marginal with respect to λ, there follows

$$\int p^*(x) e^{-g(x)} e^{\sum_{i=1}^n \zeta_i f_i(x)} \pi(y|x) \lambda(dx) = e^{-I(C)} e^{\sum_{i=1}^n \zeta_i \gamma_i} p^*(y) e^{-g(y)}.$$

Then $p^*(y)e^{-g(y)} \in L^1(\lambda)$ is the unique positive eigenvector for T_ζ^* corresponding to the same eigenvalue. Since the product $V_\zeta(x)\psi_\zeta(x) = p^*(x)$, (2.14) holds. The conclusion of the lemma follows. □

LEMMA 2.13. *Under the assumption of Theorem* 2.9,

$$I(C) = \max_{\zeta \in \mathbb{R}_n^+} \left(\sum_{i=1}^{n} \zeta_i \gamma_i - \log \rho_\zeta \right),$$

where ρ_ζ is the (positive) eigenvalue of greatest modulus for the operator T_ζ.

PROOF. For any vector $\boldsymbol{\zeta} \in \mathbb{R}_n^+$, let

$$P_\zeta(x,y) = \left(V_\zeta(y) \pi(y|x) e^{\sum_{i=1}^n \zeta_i f_i(x)} \psi_\zeta(x) \right) / \rho_\zeta,$$

where V_ζ, ψ_ζ and ρ_ζ are as defined in Theorem 2.9. From Lemma 2.12, the *I*-projection P^* of π onto C has $\lambda \times \lambda$ density $P^*(x,y) = P_{\zeta^*}(x,y)$ for some $\zeta^* \in \mathbb{R}_n^+$. It follows from (2.19) that

$$(2.20) \qquad I(C) = \sum_{i=1}^{n} \zeta_i^* \gamma_i - \log \rho_{\zeta^*}.$$

Let $\Lambda = \{P' \in \Lambda_0,\ P' \sim \lambda \times \lambda\}$. Arguing exactly as in the proof of Corollary 2.7, (2.10) for all $Q \in M_C$ uniquely determines P^* among the set of $P' \in \Lambda$.

Thus if $P' \in \Lambda$, $P' \neq P^*$, there exists some $Q \in M_C$, $h^1(Q|\pi) < \infty$ such that

$$h^1(Q|\pi) < h^1(Q|P'(\cdot|\cdot)) + I(C).$$

From Lemma 2.4 it follows that

$$I(C) > \iint_{X\times X} Q(x,y)\log\frac{P'(y|x)}{\pi(y|x)}\lambda(dx)\lambda(dy).$$

Let $\tilde{M}_C$ denote the set of $Q \in M_C$ satisfying $h^1(Q|\pi) < \infty$. The argument above and (2.10) imply that for any $P' \in \Lambda$,

$$I(C) \geq \inf_{Q\in\tilde{M}_C}\iint_{X\times X} Q(x,y)\log\frac{P'(y|x)}{\pi(y|x)}\lambda(dx)\lambda(dy),$$

with strict inequality if $P' \neq P^*$. Applying this to $P' \in \Lambda$ with density P_ζ, observing that the marginal of P_ζ is

$$p_\zeta(dx) = \psi_\zeta(x)V_\zeta(x)\lambda(dx),$$

one obtains

$$\begin{aligned} I(C) &\geq \inf_{Q\in\tilde{M}_C}\iint_{X\times X} Q(x,y)\log\frac{V_\zeta(y)e^{\sum_{i=1}^n \zeta_i f_i(x)}}{V_\zeta(x)\rho_\zeta} \\ &= \inf_{Q\in\tilde{M}_C}\left(\iint_{X\times X} Q(x,y)\sum_{i=1}^n \zeta_i f_i(x) - \log\rho_\zeta\right) \qquad (2.21) \\ &\geq \left(\sum_{i=1}^n \zeta_i\gamma_i - \log\rho_\zeta\right) \quad \text{for } \zeta \in \mathbb{R}_n^+, \end{aligned}$$

where the inequality is strict if $\zeta \neq \zeta^*$. The conclusion of the lemma follows from (2.20) and (2.21). □

3. Convergence of $R_{n,\omega}$ in conditional probability. Let $(\Omega, \mathscr{B})$ be as in Section 1. Let $h_{\mathscr{F}_m^n}(\cdot,\cdot)$ denote the entropy when the supremum in (2.1) is taken over positive functions $u(x) \in C(\Omega)$, the continuous functions on Ω, which depend only on the coordinates ω_i, $n \leq i \leq m$.

Let P be a measure on $\mathscr{F}_\infty^s$ with $s \leq t$. Suppose $P\{\omega: \omega(t) = \overline{\omega}(t)\} = 1$. Define a measure $\delta_{\overline{\omega}} \otimes_t P$ on Ω by

$$\begin{aligned} (\delta_{\overline{\omega}} \otimes_t P)&\{\omega(t_1) \in A_1, \omega(t_2) \in A_2, \dots, \omega(t_n) \in A_n\} \\ &= \chi_{A_1}(\overline{\omega}(t_1))\chi_{A_2}(\overline{\omega}(t_2)) \cdots \chi_{A_k}(\overline{\omega}(t_k)) \\ &\quad \times P\{\omega(t_{k+1}) \in A_{k+1}, \dots, \omega(t_n) \in A_n\}, \end{aligned}$$

where $t_1 < t_2 < \cdots < t_k \leq t \leq t_{k+1} < \cdots < t_n$. Suppose $\pi(dy|x)$ is a transition probability function giving rise to a Markov process P_x on Ω_x. For $\omega \in \Omega$, let $P_\omega = \delta_\omega \otimes_0 P_{\omega(0)}$ and define a measure $\hat{Q}$ on Ω by

$$\hat{Q} = \int_\Omega P_\omega Q(d\omega). \qquad (3.1)$$

For $Q \in M_S(\Omega)$ define the entropy of Q with respect to π by

$$H(Q|\pi) = h_{\mathscr{F}_1^{-\infty}}(Q, \hat{Q}).$$

By (2.1),

$$(3.2) \qquad H(Q|\pi) = \sup_{u \in \mathscr{U}} \left[\int_\Omega \log u(\omega) Q(d\omega) - \log \int_\Omega u(\omega) \hat{Q}(d\omega) \right],$$

where $\mathscr{U}$ is the set of positive continuous functions which only depend on ω_i, $i \leq 1$. By (3.1),

$$\int_\Omega u(\omega) \hat{Q}(d\omega) = \int_\Omega E^{P_\omega}(u) Q(d\omega).$$

Under assumption 3 on the probability transition function $\pi(dy|x)$, $\omega \to E^{P_\omega}(u)$ is a continuous function for $u \in \mathscr{U}$. Then the expression in brackets in (3.2) is a continuous function of Q. It follows that $H(Q|\pi)$ is lower semicontinuous.

These definitions are required for the proof of the following lemma.

Lemma 3.1. *Let C be a closed convex set in $\mathscr{M}(X)$ satisfying $I(C) = I(C^0) < \infty$. Let P^* be the I-projection of π onto C considered as a stationary process on $(\Omega, \mathscr{B})$. Then in terms of the metric for the weak topology on $M_S(\Omega)$, $R_{n,\omega}$ converges to P^* in conditional P_x-probability given $\hat{P}_n(\omega, \cdot) \in C$, uniformly for $x \in X$.*

Proof. It follows from (1.5), (1.7) and the assumption that $I(C) = I(C^0) < \infty$ that

$$(3.3) \qquad \lim_{n \to \infty} \frac{1}{n} \log P_x\{\hat{P}_n(\omega, \cdot) \in C\} = -I(C),$$

uniformly for $x \in X$. Let Π_C be the set of $Q \in M_S(\Omega)$ with marginals in C. Then $\hat{P}_n(\omega, \cdot) \in C$ is equivalent to $R_{n,\omega} \in \Pi_C$.

Since $(\Omega, \mathscr{B})$ is a Polish space, the weak topology on the set of probability measures on Ω is metrizable. Let $\Delta(\cdot, \cdot)$ denote this metric. Define

$$\Pi_C^\varepsilon = \{Q \in \Pi_C : \Delta(Q, P^*) \geq \varepsilon\}.$$

Π_C and Π_C^ε are closed sets of $M_S(\Omega)$, which is compact, so both Π_C and Π_C^ε are compact. Under assumption 3 the methods of Donsker and Varadhan (1983) show that

$$(3.4) \qquad \limsup_{n \to \infty} \frac{1}{n} \log \sup_{x \in X} P_x\{R_{n,\omega} \in \Pi_C^\varepsilon\} \leq - \inf_{Q \in \Pi_C^\varepsilon} H(Q|\pi).$$

A proof of (3.4) is given in the Appendix, Theorem A.1.

$H(Q|\pi)$ is lower-semicontinuous in Q so that both $H(\Pi_C|\pi) = \inf_{Q \in \Pi_C} H(Q|\pi)$ and $H(\Pi_C^\varepsilon|\pi) = \inf_{Q \in \Pi_C^\varepsilon} H(Q|\pi)$ are achieved. Using the contraction principle of Donsker and Varadhan (1983), Theorem 6.1,

$$H(\Pi_C|\pi) = \inf_{\mu \in C} \inf_{\{Q:\, q(Q) = \mu\}} H(Q|\pi)$$

$$= \inf_{\mu \in C} I(\mu) = I(C),$$

where $q(Q)$ denotes the marginal of Q. Thus $H(\Pi_C|\pi)$ is achieved by P^*. We will show in Lemma 3.3 that P^* is the unique minimum. Then for any $\varepsilon > 0$, $H(\Pi_C^\varepsilon|\pi) > H(\Pi_C|\pi)$. Fix ε and pick ε^1 such that $2\varepsilon^1 < H(\Pi_C^\varepsilon|\pi) - H(\Pi_C|\pi)$. It follows from (3.3) and (3.4) that $\exists\, N$ such that for $n \geq N$ and every $x \in X$,

$$P_x\{\Delta(R_{n,\omega}, P^*) \geq \varepsilon | \hat{P}_n(\omega, \cdot) \in C\}$$

$$= \frac{P_x\{R_{n,\omega} \in \Pi_C^\varepsilon\}}{P_x\{\hat{P}_n(\omega, \cdot) \in C\}}$$

$$\leq \frac{e^{-n(H(\Pi_C^\varepsilon|\pi) - \varepsilon^1)}}{e^{-n(H(\Pi_C|\pi) + \varepsilon^1)}}$$

$$= e^{-n(H(\Pi_C^\varepsilon|\pi) - H(\Pi_C|\pi) - 2\varepsilon^1)},$$

so that

$$\lim_{n \to \infty} P_x\{\Delta(R_{n,\omega}, P^*) \geq \varepsilon | \hat{P}_n(\omega, \cdot) \in C\} = 0, \quad \text{uniformly in } x,$$

which establishes the theorem. □

THEOREM 3.2. *The stationary processes defined by*

$$R_{n,x}^C(\cdot) = E^{P_x}\{R_{n,\omega}(\cdot) | \hat{P}_n(\omega, \cdot) \in C\}$$

converge weakly to P^* *for all* $x \in X$.

PROOF. For any $f \in C(\Omega)$, it follows as in Theorem 3.1 that

$$\lim_{n \to \infty} P_x\left\{\left|\int f\, dR_{n,\omega} - \int f\, dP^*\right| \geq \varepsilon | P_n(\omega, \cdot) \in C\right\} = 0, \tag{3.5}$$

uniformly for $x \in X$.

Now for any $f \in L^1(R_{n,x}^C)$,

$$\int_\Omega f\, dR_{n,x}^C = E^{P_x}\left\{\int f\, dR_{n,\omega} | \hat{P}_n(\omega, \cdot) \in C\right\}.$$

Then for $f \in C(\Omega)$,

$$\begin{aligned}
&\left|\int_{\Omega} f\,dR_{n,x}^{C} - \int_{\Omega} f\,dP^*\right| \\
&\quad \le E^{P_x}\left\{\left|\int_{\Omega} f\,dR_{n,\omega} - \int_{\Omega} f\,dP^*\right| \,\middle|\, \hat{P}_n(\omega,\cdot) \in C\right\} \\
&\quad < 2|f| P_x\left\{\left|\int_{\Omega} f\,dR_{n,\omega} - \int_{\Omega} f\,dP^*\right| \ge \varepsilon \,\middle|\, \hat{P}_n(\omega,\cdot) \in C\right\} \\
&\qquad + \varepsilon\left(P_x\left\{\left|\int_{\Omega} f\,dR_{n,\omega} - \int_{\Omega} f\,dP^*\right| < \varepsilon \,\middle|\, \hat{P}_n(\omega,\cdot) \in C\right\}\right),
\end{aligned}$$

so by (3.5),

$$\lim_{n\to\infty}\left|\int_{\Omega} f\,dR_{n,x}^{c} - \int_{\Omega} f\,dP^*\right| \le \varepsilon.$$

Since ε is arbitrary, the weak convergence of $R_{n,x}^{C}$ to P^* is established.

To complete the proof of Lemma 3.1, we establish the following lemma.

Lemma 3.3. *Let P^*, Π_C and $H(Q|\pi)$ be as in the proof of Theorem 3.1. Then*

$$H(\Pi_C|\pi) = \inf_{Q\in\Pi_C} H(Q|\pi)$$

is attained uniquely by P^*.

Proof. Suppose that $Q \in \Pi_C$ achieves the above infimum, which, by assumption, is finite. Let $\hat{Q}$ be as defined by (3.1). Then $Q \ll \hat{Q}$. Denote by $\hat{Q}_1^0$ the restriction of $\hat{Q}$ to $\mathscr{F}_1^{\,0}$. Let $\hat{Q}_{1,\omega}^0$ be the regular conditional probability distribution of $\hat{Q}$ given $\mathscr{F}_1^{\,0}$. Then $E^{\hat{Q}_{1,\omega}^0}[dQ/d\hat{Q}]$ is a version of $dQ_1^0/d\hat{Q}_1^0$. It follows that

$$\begin{aligned}
I(C) &\le h^1(Q_1^0|\pi) \\
&= \int E^{\hat{Q}_{1,\omega}^0}\left[\frac{dQ}{d\hat{Q}}\right]\log E^{\hat{Q}_{1,\omega}^0}\left[\frac{dQ}{d\hat{Q}}\right] d\hat{Q}_1^0(\omega) \\
&\le \int E^{\hat{Q}_{1,\omega}^0}\left[\frac{dQ}{d\hat{Q}}\log\frac{dQ}{d\hat{Q}}\right] d\hat{Q}_1^0(\omega) \\
&= \int dQ \log\frac{dQ}{d\hat{Q}} = I(C),
\end{aligned}$$

where Jensen's inequality for the measure $\hat{Q}_{1,\omega}^0$ has been used. However, in this case, we must have equality holding in the Jensen estimate for $\hat{Q}_1^0$-a.e. ω. Since $x \log x$ is strictly convex, this implies for $\hat{Q}$-a.e. ω,

$$\frac{dQ}{d\hat{Q}}(\omega) = E^{\hat{Q}_{1,\omega}^0}\left[\frac{dQ}{d\hat{Q}}\right]. \tag{3.6}$$

Let $Q_{\omega(0)}$ denote the regular conditional probability distribution of Q_1^0 given $\mathscr{F}_0^{\,0}$ and note that $P_{\omega(0)}$ is the regular conditional probability distribution of $\hat{Q}_1^0$ given $\mathscr{F}_0^{\,0}$. The measures Q_1^0 and $\hat{Q}_1^0$ have the same marginal distribution on $\mathscr{F}_0^{\,0}$, which we denote by q. Then $dQ_1^0/d\hat{Q}_1^0$ is the Radon-Nikodym derivative of $Q_{\omega(0)}$ with respect to $P_{\omega(0)}$ which exists for q-a.e. $\omega(0)$. It now follows from (3.6) that

$$\frac{dQ}{d\hat{Q}} = \frac{dQ_{\omega(0)}}{dP_{\omega(0)}} \quad \text{for } \hat{Q}\text{-a.e. } \omega.$$

This shows Q is a stationary Markov process as follows: Let $B \in \mathscr{F}_0^{-\infty}$, $A \in \mathscr{F}_1^{\,1}$:

$$\begin{aligned} Q[A \cap B] &= \int_{A\cap B} \frac{dQ}{d\hat{Q}} \hat{Q}(d\omega) \\ &= \int_B E^{P_\omega}\left[\frac{dQ}{d\hat{Q}}\chi_A\right] Q(d\omega) \\ &= \int_B E^{P_\omega}\left[\frac{dQ_{\omega(0)}}{dP_{\omega(0)}}\chi_A\right] Q(d\omega) \quad [\text{by } (3.1)] \\ &= \int_B E^{P_{\omega(0)}}\left[\frac{dQ_{\omega(0)}}{dP_{\omega(0)}}\chi_A\right] Q(d\omega) \\ &= \int_B Q_{\omega(0)}(A) Q(d\omega). \end{aligned}$$

Since $Q_{\omega(0)}(A) = E^Q[A|\mathscr{F}_0^{\,0}]$, this shows that $E^Q[A|\mathscr{F}_0^{-j}] = E^Q[A|\mathscr{F}_0^{\,0}]$ for any $A \in \mathscr{F}_1^{\,1}$, $j > 0$. It follows that Q is a stationary Markov process with transition probability function $Q(A|x) = Q_x(A)$ for $A \in \mathscr{F}_1^{\,1}$. Since $I(C) = h^1(Q_1^0|\pi)$, it follows from Corollary 2.7 that Q is the stationary Markov process P^*. □

4. A large deviations estimate.

For $u(x)$ a probability density function with respect to λ, let

$$R_{n,u}^C(\cdot) = E^{P_u}\{R_{n,\omega}(\cdot)|\hat{P}_n(\omega,\cdot) \in C\} \tag{4.1}$$

be defined as in (1.9). In this section we show the sequence of measures $R_{n,u}^C(\cdot)$, $n = 1, 2, \dots$, is asymptotically quasi-Markov.

We begin by establishing a fundamental lemma. Let Q be a stationary process on $(\Omega, \mathscr{B})$ with marginal q. Let Q_1^0 be the restriction of Q to $\mathscr{F}_1^{\,0}$ and let $Q_{\omega(0)}$ denote the regular conditional probability distribution of Q_1^0 w.r.t.

$\mathscr{F}_0^0$. Then, as before, for $A \in \mathscr{F}_1^1$, $Q(A|x) = Q_x(A)$ defines a transition probability function a.e. q. Let $\tilde{Q}$ be the stationary Markov process with transition probability $Q(A|x)$ and invariant measure q. We say that $\tilde{Q}$ *is the stationary Markov process defined by* Q. Then we have the following lemma.

LEMMA 4.1. *Let $\bar{P}$ be the Markov process on $\prod_{i=0}^{\infty}(X_i, \mathscr{B}_i)$, $X_i = X$, $\mathscr{B}_i$ the Boreal σ-field on X, $0 \le i < \infty$, with a probability transition function $P(dy|x)$ and initial distribution $q(dx)$. Let $Q, \tilde{Q}$ be as above. Then for any n,*

$$h_{\mathscr{F}_n^0}(Q, \bar{P}) = h_{\mathscr{F}_n^0}(Q, \tilde{Q}) + nh^1(Q_1^0|P(\cdot|\cdot)).$$

PROOF. Let $\bar{P}_1^0$ denote the restriction of $\bar{P}$ to $\mathscr{F}_1^0$. Then we can assume $Q \ll \tilde{Q}$ on $\mathscr{F}_n^0$ and $Q_1^0 \ll \bar{P}_1^0$; otherwise both sides are ∞. To establish this, suppose $h_{\mathscr{F}_n^0}(Q, \bar{P}) < \infty$. Then $Q \ll \bar{P}$ on $\mathscr{F}_n^0$; in particular, $Q_1^0 \ll \bar{P}_1^0$. Since Q_1^0 and $\bar{P}_1^0$ both have marginal q on $\mathscr{F}_0^0$, $dQ_1^0/d\bar{P}_1^0$ is the Radon-Nikodym derivative of $Q(dy|x)$ with respect to $P(dy|x)$, which exists for q-a.e. x. Now suppose for $M \in \mathscr{F}_n^0$, $\tilde{Q}(M) = 0$. However,

$$\tilde{Q}(M) = \int_M \frac{dQ_1^0}{d\bar{P}_1^0}(\omega_0, \omega_1) \cdots \frac{dQ_1^0}{d\bar{P}_1^0}(\omega_{n-1}, \omega_n)\, d\bar{P}(\omega_0, \dots, \omega_n), \tag{4.2}$$

so that $\hat{Q}(M) = 0$ implies that $\bar{P}$-a.e. on $\mathscr{F}_n^0$, $dQ_1^0/d\bar{P}_1^0(\omega_0, \omega_1) \cdots dQ_1^0/d\bar{P}_1^0(\omega_{n-1}, \omega_n) = 0$. Let N be the $\mathscr{F}_n^0$ set of $\bar{P}$-measure 0, where this product is positive. Let $T_i = \{(\omega_{i-1}, \omega_i): dQ_1^0/d\bar{P}_1^0(\omega_{i-1}, \omega_i) = 0\}$. Using the stationarity of Q, each T_i has Q-measure 0. Then $M \subseteq N \cup \bigcup_{i=1}^n T_i$ so M has Q-measure 0 and $Q \ll \tilde{Q}$.

Assuming that $Q \ll \hat{Q}$ on $\mathscr{F}_n^0$ and $Q_1^0 \ll \bar{P}_1^0$, which we have seen implies $\tilde{Q} \ll \bar{P}$ on $\mathscr{F}_n^0$, we have

$$\frac{dQ}{d\bar{P}} = \frac{dQ}{d\tilde{Q}} \frac{d\tilde{Q}}{d\bar{P}}$$

on $\mathscr{F}_n^0$. Taking log of both sides, integrating over Q and using (4.2) gives

$$h_{\mathscr{F}_n^0}(Q, \bar{P}) = h_{\mathscr{F}_n^0}(Q, \tilde{Q}) + \int dQ(\omega_0, \dots, \omega_n) \log \frac{dQ_1^0}{d\bar{P}_1^0}(\omega_0, \omega_1) \cdots \frac{dQ_1^0}{d\bar{P}_1^0}(\omega_{n-1}, \omega_n).$$

Using the stationarity of Q, the last integral on the right is $nh^1(Q_1^0|P(\cdot|\cdot))$. □

The following lemma establishes the analog in this situation of the almost completely convex condition required by Csiszár (1984) on the convex set C described in Section 1.

LEMMA 4.2. *Let $R^C_{n,u}(\cdot)$ be as defined in* (4.1) *and let Π_C be the set of $M_S(\Omega)$ whose marginals are in C, a weakly closed convex set. Then $R^C_{n,u}(\cdot) \in \Pi_C$.*

PROOF. Consider $P_u\{\cdot|\hat{P}_n(\omega, \cdot) \in C\}$ as a measure of $\mathscr{F}^0_{n-1}$. Now $\Pi^\infty_{i=0}(X_i, \mathscr{B}_i)$ is $(\Pi^\infty_{i=0}X_i, \mathscr{B})$, where $\mathscr{B}$ is the Borel σ-field on $\Pi^\infty_{i=0}X_i$, which is in particular a separable metric space. It is standard that there are probability measures μ_j on $\mathscr{F}^0_{n-1}$,

$$\mu_j = \sum_{k=0}^{k=k_j} a_{jk}\delta_{e_{jk}},$$

whose supports are finite sets which converge weakly to $P_u\{\cdot|\hat{P}_n(\omega, \cdot) \in C\}$ [Parthasarathy (1967), Theorem (6.3)]. Let E_n be the $\mathscr{F}^0_{n-1}$-measurable set of ω satisfying $\hat{P}_n(\omega, \cdot) \in C$. Without loss of generality, it may be assumed that the finite set $\{e_{jk}\}$ lies in E_n for each j and k.

Now let $f \in C(\Omega)$. Then for each $\omega \in \Omega$,

$$\int f\,dR_{n,\omega} = \frac{1}{n}\sum_{i=0}^{n-1} f(\theta_i\omega_n)$$

is a continuous function of $(\omega_0, \dots, \omega_{n-1})$. It follows that

$$\int\left(\int f\,dR_{n,\omega}\right)\mu_j = \sum_{k=0}^{k=k_j} a_{jk}\left(\int f\,dR_{n,e_{jk}}\right)$$

converges as $j \to \infty$ to

$$E_u\left\{\int f\,dR_{n,\omega}|\hat{P}_n(\omega, \cdot) \in C\right\}$$

$$= \int_\Omega f(\omega)\,dR^C_{n,u}.$$

Thus the measure

$$\sum_{k=0}^{k=k_j} a_{jk}R_{n,e_{jk}} \tag{4.3}$$

converges weakly as $j \to \infty$ to $R^C_{n,u}$. Since for each j and k, $e_{jk} \in E_n$, it follows that $\hat{P}_n(e_{jk}, \cdot) \in C$ or equivalently that $R_{n,e_{jk}} \in \Pi_C$. By the convexity of Π_C, each of the measures in (4.3) is in Π_C. Thus $R^C_{n,\omega}$ is a limit point of Π_C, which, being closed, implies $R^C_{n,u} \in \Pi_C$. □

LEMMA 4.3. *Let E_n be the $\mathscr{F}^0_{n-1}$-measurable set $\{\omega: \hat{P}_n(\omega, \cdot) \in C\}$. The measure $R^C_{n,u}$ defined by* (4.1) *has a density for sets A in $\mathscr{F}^0_{n-1}$ with respect to λ^n given by*

$$R^C_{n,u}(\omega_0, \dots, \omega_{n-1}) = \frac{1}{P_u\{E_n\}}\chi_{E_n}(\omega_0, \dots, \omega_{n-1})\frac{1}{n}\sum_{i=0}^{n-1}\pi_i(\omega_0, \dots, \omega_{n-1}),$$

where

$$\pi_0(\omega_0, \dots, \omega_{n-1}) = u(\omega_0)\pi(\omega_1|\omega_0) \cdots \pi(\omega_{n-1}|\omega_{n-2})$$

and

$$\pi_i(\omega_0, \dots, \omega_{n-1}) = \pi(\omega_0|\omega_{n-1})\pi(\omega_1|\omega_0) \cdots \pi(\omega_{n-i-1}|\omega_{n-i-2}) \times u(\omega_{n-i})\pi(\omega_{n-i+1}|\omega_{n-i}) \cdots \pi(\omega_{n-1}|\omega_{n-2}). \tag{4.4}$$

PROOF. Observe that the map $\Omega \to \Omega$ defined by $\omega \mapsto \omega_n$ is continuous, hence $\mathscr{F}_n^0$-measurable, and the maps θ_i, θ_i^{-1}: $\Omega \to \Omega$ are continuous, hence measurable. Let $i > 0$ and let A be a measurable set in $\mathscr{F}_{n-1}^0$. Then

$$\begin{aligned} &E_u\{\chi_A(\theta_i\omega_n)\} \\ &\quad = \int_X \cdots \int_X \chi_A(\theta_i\omega_n)\pi_0(\omega_0, \dots, \omega_{n-1})\, d\lambda^n \\ &\quad = \int_X \cdots \int_X \chi_A(\omega_n)\pi_0(\theta_i^{-1}\omega_n)\, d\lambda^n \end{aligned} \tag{4.5}$$

by Fubini's theorem. It is easy to see that for $i > 0$,

$$\pi_0(\theta_i^{-1}\omega_n) = \pi(\omega_0|\omega_{n-1})\pi(\omega_1|\omega_0) \cdots \pi(\omega_{n-i-1}|\omega_{n-i-2}) \times u(\omega_{n-i})\pi(\omega_{n-i+1}|\omega_{n-i}) \cdots \pi(\omega_{n-1}|\omega_{n-2}). \tag{4.6}$$

To obtain the density of

$$R_{n,u}^C(\cdot) = E_u\{R_{n,\omega}(\cdot)|\hat{P}_n(\omega, \cdot) \in C\},$$

observe that $\omega \in E_n$ if and only if $\theta_i\omega_n \in E_n$ for any $0 \le i \le n-1$. Then

$$\begin{aligned} R_{n,u}^C(\cdot) &= E_u\{R_{n,\omega}(\cdot)|\omega \in E_n\} \\ &= \frac{1}{n}\sum_{i=0}^{n-1} E_u\{\chi_{(\cdot)}(\theta_i\omega_n)|\theta_i\omega_n \in E_n\}. \end{aligned}$$

It now follows from (4.5) and (4.6) that for any set $A \in \mathscr{F}_{n-1}^0$,

$$R_{n,u}^C(\cdot) = \int \cdots \int_A \frac{1}{P_u\{E_n\}}\chi_{E_N}(\omega_0, \dots, \omega_{n-1})\frac{1}{n}\sum_{i=0}^{n-1}\pi_i(\omega_0, \dots, \omega_{n-1})\, d\lambda^n,$$

which proves the lemma. □

Let $\bar{P}_{n,u}^*$ be the measure on $\mathscr{F}_{n-1}^0$ defined by the transition probability function $P^*(dy|x)$, P^* the I-projection of π onto C and initial distribution given by the marginal of $R_{n,u}^C$. We now establish the following lemma.

LEMMA 4.4. *Let C be a closed convex set with nonempty interior C^0 satisfying $I(C^0) < \infty$. Suppose the probability density function in* (4.1) *is* *bounded from above. Then*

$$\frac{1}{n}\log P_u\{\hat{P}_n(\omega,\cdot)\in C\}$$
$$\le -\frac{1}{n}h_{\mathscr{F}^0_{n-1}}\big(R^C_{n,u},\bar{P}^*_{n,u}\big)-\frac{(n-1)}{n}I(C)+\frac{1}{n}\log\left(\frac{\sup uA}{a}\right),$$

where a and A are the bounds on $\pi(y|x)$ given by assumption 3.

PROOF. Let

$$\hat{R}^C_{n,u}(\cdot)=E_u\{R_{n,\omega}(\cdot)\}$$
$$=\frac{1}{n}\sum_{i=0}^{n-1}E_u\{\chi_{(\cdot)}(\theta_i\omega_n)\}.$$

Then

$$-\log P_u\{\hat{P}_n(\omega,\cdot)\in C\}=h_{\mathscr{F}^0_{n-1}}\big(R^C_{n,u},\hat{R}^C_{n,u}\big)$$
$$=\int_X\cdots\int_X dR^C_{n,u}\log\frac{R^C_{n,u}(\omega_0,\dots,\omega_{n-1})}{\hat{R}^C_{n,u}(\omega_0,\dots,\omega_{n-1})}, \tag{4.7}$$

where $R^C_{n,u}(\omega_0,\dots,\omega_{n-1})$ and $\hat{R}^C_{n,u}(\omega_0,\dots,\omega_{n-1})$ are the densities of $R^C_{n,u}$ and $\hat{R}^C_{n,u}$, respectively, with respect to λ^n. Let $\pi^1(\omega_0,\omega_1,\dots,\omega_{n-1})=\pi(\omega_1|\omega_0)\pi(\omega_2|\omega_1)\cdots\ \omega(\omega_{n-1}|\omega_{n-2})$. Let $r_{n,u}(\omega_0)$ denote the density of the marginal of $R^C_{n,u}$ with respect to λ and let $\pi_{n,u}$ be the measure on $\mathscr{F}_n^0$ with density $r_{n,u(\omega_0)}\pi^1(\omega_0,\dots,\omega_{n-1})$ with respect to λ^n. Now for $R^C_{n,u}$-a.e. $(\omega_0,\dots,\omega_{n-1})$,

$$\frac{R^C_{n,u}(\omega_0,\dots,\omega_{n-1})}{\hat{R}^C_{n,u}(\omega_0,\dots,\omega_{n-1})}=\frac{R^C_{n,u}(\omega_0,\dots,\omega_{n-1})}{\pi_{n,u}(\omega_0,\dots,\omega_{n-1})}$$
$$\times\, r_{n,u}(\omega_0)\frac{\pi^1(\omega_0,\dots,\omega_{n-1})}{\hat{R}^C_{n,u}(\omega_0,\dots,\omega_{n-1})}.$$

It is possible to take the log of both sides and integrate over $R^C_{n,u}$ to obtain

$$\int_X\cdots\int_X dR^C_{n,u}\log\frac{R^C_{n,u}}{\hat{R}^C_{n,u}}$$
$$=\int_X\cdots\int_X dR^C_{n,u}\log\frac{R^C_{n,u}}{\pi_{n,u}} \tag{4.8}$$
$$+\int_X\cdots\int_X dR^C_{n,u}\log r_{n,u}(\omega_0)+\int_X\cdots\int_X dR^C_{n,u}\log\frac{\pi^1}{\hat{R}^C_{n,u}},$$

provided the right-hand side is well defined. However, the first integral on the right is evidently positive as is the second, which is just $h(r_{n,u},\lambda)$. For the third, it follows from (4.4) and the bounds on π and u that

$$\pi_i(\omega_0,\dots,\omega_{n-1}) \le \frac{\sup uA}{a}\pi^1(\omega_0,\dots,\omega_{n-1}), \tag{4.9}$$

so that

$$\int_X\cdots\int_X dR^C_{n,u}\log\frac{\pi^1}{\hat{R}^C_{n,u}} \ge -\log\left(\frac{\sup uA}{a}\right).$$

For (4.7), (4.8), and (4.9) it now follows that

$$-\log P_u\{\hat{P}_n(\omega,\cdot)\in C\} \ge h_{\mathscr{F}^0_{n-1}}(R^C_{n,u},\pi_{n,u}) - \log\left(\frac{\sup uA}{a}\right). \tag{4.10}$$

Applying Lemma 4.1 and letting $\tilde{R}^C_{n,u}$ denote the stationary Markov process defined by $R^C_{n,u}$, we have

$$h_{\mathscr{F}^0_{n-1}}(R^C_{n,u},\pi_{n,u}) = h_{\mathscr{F}^0_{n-1}}(R^C_{n,u},\tilde{R}^C_{n,u}) + (n-1)h^1(R^C_{n,u}{}^0_1|\pi).$$

Since $R^C_{n,u}\in\Pi_C$ by Lemma 4.2, $R^C_{n,u}{}^0_1\in M_C$ and it follows from Theorem 2.3 that

$$h^1(R^C_{n,u}{}^0_1|\pi) \ge h^1(R^C_{n,u}{}^0_1|P^*(\cdot\mid\cdot)) + I(C).$$

From (4.10) we have

$$\begin{aligned}-\log P_u\{\hat{P}_n(\omega,\cdot)\in C\} \ge{}& h_{\mathscr{F}^0_{n-1}}(R^C_{n,u}\tilde{R}^C_{n,u})\\ &+(n-1)h^1(R^C_{n,u}{}^0_1|P^*(\cdot\mid\cdot))\\ &+(n-1)I(C) - \log\left(\frac{\sup uA}{a}\right).\end{aligned} \tag{4.11}$$

Applying Lemma 4.1 again gives

$$h_{\mathscr{F}^0_{n-1}}(R^C_{n,u},\bar{P}^*_{n,u}) = h_{\mathscr{F}^0_{n-1}}(R^C_{n,u},\tilde{R}^C_{n,u}) + (n-1)h^1(R^C_{n,u}{}^0_1|P^*(\cdot\mid\cdot)).$$

Substituting this into (4.11) yields

$$-\log P_u\{\hat{P}_n(\omega,\cdot)\in C\} \ge h_{\mathscr{F}^0_{n-1}}(R^C_{n,u},\bar{P}^*_{n,u}) + (n-1)I(C) - \log\left(\frac{\sup uA}{a}\right).$$

The lemma follows. □

THEOREM 4.5. *Suppose that in addition to the hypothesis of Lemma* 4.4 *we have* $I(C)=I(C^0)<\infty$. *Then*

$$\lim_{n\to\infty}\frac{1}{n}h_{\mathscr{F}^0_{n-1}}(R^C_{n,u},\bar{P}^*_{n,u}) = 0,$$

so that the measures $R^C_{n,u}$ *are asymptotically quasi-independent with respect to* $P^*(dy|x)$, *the transition probability function of the I-projection of* π *onto* C.

PROOF. Using the uniformity of the estimate (1.5), it follows that

$$\liminf_{n\to\infty} \frac{1}{n} \log P_u\{\hat{P}_n(\omega, \cdot) \in C^0\} \geq -I(C^0) = -I(C).$$

It follows from Lemma 4.4 that

$$\limsup_{n\to\infty} \left(\frac{1}{n} \log P_u\{\hat{P}_n(\omega, \cdot) \in C\} + \frac{1}{n} h_{\mathscr{F}^0_{n-1}}\left(R^C_{n,u}, \bar{P}^*_{n,u}\right)\right) \tag{4.12}$$
$$\leq -I(C).$$

In particular,

$$\lim_{n\to\infty} \frac{1}{n} \log P_u\{\hat{P}_n(\omega, \cdot) \in C\} = -I(C).$$

It follows from (4.12) that

$$\limsup_{n\to\infty} \frac{1}{n} h_{\mathscr{F}^0_{n-1}}\left(R^C_{n,u}, \bar{P}^*_{n,u}\right) \leq 0,$$

which establishes the theorem. □

5. Corollaries. Let C be a closed convex set with nonempty interior satisfying $I(C^0) = I(C) < \infty$, so that the measures $R^C_{n,u}$ defined by (1.9) for $u(x)$ bounded from above are asymptotically quasi-independent with respect to $P^*(dy|x)$, the probability transition function of the I-projection of π onto C. From Theorem 2.3, this function is defined for λ-a.e. x. Extend it as described in Corollary 2.8. Let $\hat{Q}$ be the measure on Ω defined by

$$\int_\Omega \delta_\omega \otimes_0 P^*_{\omega(0)} Q(d\omega). \tag{5.1}$$

For $Q \in M_S(\Omega)$, define $h^j(Q|P^*(\cdot | \cdot)) = h_{\mathscr{F}_1^{-j}}(Q, \hat{Q})$.

COROLLARY 5.1. *For $h^j(\cdot | \cdot)$ defined as above,*

$$\lim_{n\to\infty} h^j\left(R^C_{n,u}|P^*(\cdot | \cdot)\right) = 0.$$

PROOF. From Theorem 4.5, we have that

$$\lim_{n\to\infty} \frac{1}{n} h_{\mathscr{F}^0_{n-1}}\left(R^C_{n,u}, \bar{P}^*_{n,u}\right) = 0,$$

where $\bar{P}^*_{n,u}$ is the measure on $\mathscr{F}^0_{n-1}$ defined by the transition probability function $P^*(dy|x)$ and initial distribution given by the marginal of $R^C_{n,u}$. We can assume without loss of generality that for $n > 1$, $h_{\mathscr{F}^0_{n-1}}(R^C_{n\,u}, \bar{P}^*_{n,u}) < \infty$. It follows from the proof of Lemma A.4 in the Appendix [(A.5)] that

$$h_{\mathscr{F}^0_{n-1}}\left(R^C_{n,u}, \bar{P}^*_{n,u}\right) = \sum_{i=1}^{n-1} h^i\left(R^C_{n,u}|P^*(\cdot | \cdot)\right). \tag{5.2}$$

From their definition, $h^k(R^C_{n,u}|P^*(\cdot|\cdot)) \le h^l(Q|P^*(\cdot|\cdot))$ if $k < l$. Then for $j \le n-1$,

$$\frac{n-j}{n} h^j\big(R^C_{n,u}|P^*(\cdot|\cdot)\big) \le \frac{1}{n}\sum_{i=1}^{n-1} h^i\big(R^C_{n,u}|P^*(\cdot|\cdot)\big)$$
$$= \frac{1}{n} h_{\mathscr{F}^0_{n-1}}\big(R^C_{n,u}, \bar{P}^*_{n,u}\big),$$

so that $\lim_{n\to\infty} h^j(R^C_{n,u}|P^*(\cdot|\cdot)) = 0$. □

Using Corollary 2.8, it follows that $h^j(Q|P^*(\cdot|\cdot))$ is a lower semicontinuous function of Q. Since $M_S(\Omega)$ is compact, it follows from Corollary 5.1 that any subsequence of $\{R^C_{n,u}\}$ contains a subsequence which converges weakly to the stationary Markov process P^*, so that $R^C_{n,u}$ converges weakly to P^*. Of course, this follows immediately from Lemma 3.1 using the uniformity of convergence for $x \in X$. However, more can be concluded from Corollary 5.1.

COROLLARY 5.2. *Let $f(\omega)$ be measurable with respect to $\mathscr{F}^{i_1}_{i_j}$ and suppose that for $|t|$ sufficiently small, $e^{tf(\omega_{i_1},\dots,\omega_{i_j})}$ is integrable with respect to $(1/b(\omega_{i_1})\cdots 1/b(\omega_{i_j})\lambda^j)$, where the functions $b(\cdot)$ are as in Corollary 2.8. Then $\int f(\omega)\,dR^C_{n,u} \to \int f(\omega)\,dP^*$ as $n\to\infty$.*

PROOF. The proof is similar to Csiszár (1975), Lemma 3.1. Using the stationarity of $R^C_{n,u}$, we can assume without loss of generality that $f(\omega)$ is measurable with respect to $\mathscr{F}^1_j$. Using (5.2), it follows from Corollary 5.1 that

$$\lim_{n\to\infty} h_{\mathscr{F}^0_j}\big(R^C_{n,u}, \bar{P}^*_{n,u}\big) = 0.$$

Let $f_{0,j,n}$ be the Radon-Nikodym derivative of $R^C_{n,u}$ with respect to $\bar{P}^*_{n,u}$ on $\mathscr{F}^0_j$. Let $Y = \prod_{i=0}^j X_i$, $X_i = X$, $i = 1,\dots,j$. The Csiszár–Kemperman–Kullback inequality is that for two probability measures P and Q on a measure space $(X,\mathscr{X})$:

$$|P - Q| \le \sqrt{2h(P,Q)}$$

[Csiszár (1967), Theorem 4.1, Kemperman (1969), Theorem 6.11, and Kullback (1967)]. It follows that

$$\lim_{n\to\infty} \big|R^C_{n,u} - \bar{P}^*_{n,u}\big|_{\mathscr{F}^0_j} = \lim_{n\to\infty}\int \big|f_{0,j,n} - 1\big|\,d\bar{P}^*_{n,u} = 0,$$

where $|\cdot - \cdot|_{\mathscr{F}^0_j}$ denotes the variation norm for measures on $\mathscr{F}^0_j$. Let $A_K = \{\omega\colon f(\omega) \le K\}$. Then

$$\lim_{n\to\infty}\left|\int_{A_K} f(\omega)\,dR^C_{n,u} - \int_{A_K} f(\omega)\,d\bar{P}^*_{n,u}\right| = 0. \tag{5.3}$$

However, for any $g(\omega)$ measurable with respect to $\mathscr{F}^1_j$ which is integrable with respect to $(1/b(\omega_1)\cdots 1/b(\omega_j)\lambda^j(\omega_1,\dots,\omega_j))$, it follows from Corollary 2.8 that

$$\int_{\prod_{i=1}^{j} X_i} g(\omega_1,\dots,\omega_j) P^*(\omega_1|\omega_0) P^*(\omega_2|\omega_1)\cdots P^*(\omega_j|\omega_{j-1})\, d\lambda^j(\omega_1,\dots,\omega_j)$$

is a continuous function of ω_0. Since $R^C_{n,u}$ converges weakly to P^*, the marginals of $R^C_{n,u}$ converge weakly to the marginals of P^* and it follows that $\int g(\omega)\, d\bar{P}^*_{n,u} \to \int g(\omega)\, dP^*$ as $n \to \infty$. Then (5.3) implies

$$\lim_{n\to\infty} \int_{A_K} f(\omega)\, dR^C_{n,u} = \int_{A_K} f(\omega)\, dP^*.$$

To complete the proof, it suffices to show that for any $\varepsilon > 0$, there exists K such that

$$\text{(5.4)(i)} \qquad \limsup_{n\to\infty} \int_{Y/A_K} |f|\, dR^C_{n,u} = \limsup_{n\to\infty} \int_{Y/A_k} |f| f_{0,j,n}\, d\bar{P}^*_{n,u} < \varepsilon$$

and

$$\text{(5.4)(ii)} \qquad \int_{Y/A_K} |f|\, dP^* < \varepsilon.$$

To obtain this, we prove that

$$\text{(5.5)} \qquad \lim_{n\to\infty} h_{\mathscr{F}_j^0}\big(R^C_{n,u}, \bar{P}^*_{n,u}\big) = \lim_{n\to\infty} \int_Y f_{0,j,n} \log f_{0,j,n}\, d\bar{P}^*_{n,u} = 0$$

implies that for any $A \in \mathscr{F}_j^1$,

$$\text{(5.6)} \qquad \lim_{n\to\infty} \int_A f_{0,j,n} \log f_{0,j,n}\, d\bar{P}^*_{n,u} = 0.$$

Now

$$\int_A f_{0,j,n} \log f_{0,j,n}\, d\bar{P}^*_{n,u} \ge R^C_{n,u}(A) \log \frac{R^C_{n,u}(A)}{\bar{P}^*_{n,u}(A)},$$

so that

$$\text{(5.7)} \qquad \liminf_{n\to\infty} \int_A f_{0,j,n} \log f_{0,j,n}\, d\bar{P}^*_{n,u} \ge P^*(A) \log \frac{P^*(A)}{P^*(A)} = 0.$$

Similarly,

$$\text{(5.8)} \qquad \liminf_{n\to\infty} \int_{Y/A} f_{0,j,n} \log f_{0,j,n}\, d\bar{P}^*_{n,u} \ge 0.$$

In view of (5.5), (5.7) and (5.8), (5.6) follows.

Now pick $t > 0$ and K so that $\int_{Y/A_K} e^{t|f|}\, dP^* < \varepsilon t$ so that (ii) of (5.4) is satisfied. Using the inequality $ab < a \log a + e^b$, where $a = f_{0,j,n}$ and $b = t|f|$

yields

$$\int_{Y/A_K} t|f|f_{0,j,n}\,d\bar{P}^*_{n,u} \le \int_{Y/A_K} f_{0,j,n}\log f_{0,j,n}\,d\bar{P}^*_{n,u} + \int_{Y/A_K} e^{t|f|}\,d\bar{P}^*_{n,u}.$$

It follows that

$$\begin{aligned}\limsup_{n\to\infty}\int_{Y/A_K}|f|f_{0,j,n}\,d\bar{P}^*_{n,u} &\le \frac{1}{t}\lim_{n\to\infty}\int_{Y/A_K} e^{t|f|}\,d\bar{P}^*_{n,u}\\ &= \frac{1}{t}\int_{Y/A_K} e^{t|f|}\,dP^*\\ &< \varepsilon,\end{aligned}$$

which completes the proof. □

Finally, we have the following corollary.

COROLLARY 5.3. *Suppose the hypotheses of Theorem* 4.5 *are satisfied and that additionally the probability density* $u(x)$ *is bounded away from* 0. *Then the conditional* P_u*-distribution of* $X_0,\dots,X_{n-1}$ *under the condition* $\hat{P}_n(\omega,\cdot)\in C$ *is asymptotically quasi-Markov with respect to the probability transition function* $P^*(dy|x)$.

PROOF. Let E_n be the $\mathscr{F}^0_{n-1}$ measurable set $\{\omega\colon \hat{P}_n(\omega,\cdot)\in C\}$. Then the conditional P_u-distribution of $X_0,\dots,X_{n-1}$ under the condition $\hat{P}_n(\omega,\cdot)\in C$ has the density

$$\begin{aligned}&P_{n,u}(\omega_0,\dots,\omega_{n-1})\\ &\quad=\frac{1}{P_u\{E_n\}}\chi_{E_n}(\omega_0,\dots,\omega_{n-1})u(\omega_0)\pi(\omega_1|\omega_0)\cdots\pi(\omega_{n-1}|\omega_{n-2}).\end{aligned}$$

Let $R^C_{n,u}(\omega_0,\dots,\omega_{n-1})$ be the density of $R^C_{n,u}$ on $\mathscr{F}^0_{n-1}$. Then from (4.4), we have

$$\text{(5.9)(i)}\qquad P_{n,u}(\omega_0,\dots,\omega_{n-1}) \ge \frac{\inf ua}{\sup uA}R^C_{n,u}(\omega_0,\dots,\omega_{n-1}).$$

Similarly

$$\text{(5.9)(ii)}\qquad P_{n,u}(\omega_0,\dots,\omega_{n-1}) \le \frac{\sup uA}{\inf ua}R^C_{n,u}(\omega_0,\dots,\omega_{n-1}).$$

If $p_{n,u}(\omega_0)$ is the density with respect to λ of the first marginal of $P_u\{\cdot|\hat{P}_n(\omega,\cdot)\in C\}$ on $\mathscr{F}^0_{n-1}$, then the same bounds must hold with respect to the density $r_{n,u}(\omega_0)$ of the marginals of $R^C_{n,u}$ with respect to λ. Let $\tilde{P}^*_{n,u}$ be the probability measure on $\mathscr{F}^0_{n-1}$ with initial distribution $p_{n,u}(\omega_0)\,d\lambda(\omega_0)$ and transition probability function $P^*(dy|x)$. Theorem 2.3 insures that this has a density with respect to λ^n. Let $\tilde{P}^*_{n,u}(\omega_0,\dots,\omega_{n-1})$ denote this density. Similarly, let $\bar{P}^*_{n,u}(\omega_0,\dots,\omega_{n-1})$ denote the density of $\bar{P}^*_{n,u}$ with respect to λ^n.

Then from (5.9)(i) and (ii),

$$\int P_{n,u} \log \frac{P_{n,u}}{\tilde{P}^*_{n,u}}\, d\lambda^n \geq \left(\frac{\inf ua}{\sup uA}\right) \int R^C_{n,u} \log \frac{R^C_{n,u}}{\overline{P}^*_{n,u}}\, d\lambda^n + 2\left(\frac{\inf ua}{\sup uA}\right) \log\left(\frac{\inf ua}{\sup uA}\right)$$

and

$$\int P_{n,u} \log \frac{P_{n,u}}{\tilde{P}^*_{n,u}}\, d\lambda^n \leq \left(\frac{\sup uA}{\inf ua}\right) \int R^C_{n,u} \log \frac{R^C_{n,u}}{\overline{P}^*_{n,u}}\, d\lambda^n + 2\left(\frac{\sup uA}{\inf ua}\right) \log\left(\frac{\sup uA}{\inf ua}\right).$$

It follows from Theorem 4.5 that

$$\lim_{n\to\infty} \frac{1}{n} \int P_{n,u} \log \frac{P_{n,u}}{\tilde{P}^*_{n,u}}\, d\lambda^n = 0,$$

which establishes that the sequence of measures $P_u\{\cdot\,|\hat{P}_n(\omega,\cdot) \in C\}$ on $\mathscr{F}^0_{n-1}$ is asymptotically quasi-independent with respect to $P^*(dy|x)$. □

APPENDIX

In this appendix, the following theorem is established.

THEOREM A.1. *Suppose that the probability transition function* $\pi(dy|x)$ *satisfies assumption* 3 *of Section* 1. *Then for any closed set* $A \subset M_S(\Omega)$,

$$\limsup_{n\to\infty} \frac{1}{n} \log \sup_{x\in X} P_x\{R_{n,\omega} \in A\} \leq -\inf_{Q\in A} H(Q|\pi).$$

The results of this section are, unless otherwise noted, direct translations of results of Donsker and Varadhan (1983) (cf. Sections 2, 3 and 4) into the language of discrete parameter processes. They are provided here for the convenience of the reader.

LEMMA A.2. *Let* (X, Σ) *be a Polish space and* $\mathscr{F}_1 \subset \mathscr{F}_2 \subset \Sigma$ *be sub-σ-fields. Let* μ *and* λ *be two measures on* (X, Σ) *and suppose* $\mu \ll \lambda$ *on the σ-field* $\mathscr{F}_1$. *Let* $\mu' = \int_X \lambda_\omega \mu(d\omega)$, *where* λ_ω *is the conditional probability distribution of* λ *given* $\mathscr{F}_1$. *Then*

$$h_{\mathscr{F}_2}(\mu,\lambda) = h_{\mathscr{F}_1}(\mu,\lambda) + h_{\mathscr{F}_2}(\mu,\mu'). \tag{A.1}$$

PROOF. For $E \in \Sigma$,

$$\begin{aligned}\mu'(E) &= \int_X \lambda_\omega(E)\mu(d\omega)\\ &= \int_X \lambda_\omega(E)\frac{d\mu}{d\lambda}\bigg|_{\mathscr{F}_1}\lambda(d\omega)\\ &= \int_X E^{\lambda_\omega}\left(\chi_E \frac{d\mu}{d\lambda}\bigg|_{\mathscr{F}_1}\right)\lambda(d\omega)\\ &= \int_E \frac{d\mu}{d\lambda}\bigg|_{\mathscr{F}_1}\lambda(d\omega),\end{aligned}$$

so that $d\mu'/d\lambda = d\mu/d\lambda|_{\mathscr{F}_1}$. In particular, $d\mu/d\lambda|_{\mathscr{F}_2}$ exists or both sides of (A.1) are equal to $+\infty$. Then for $E \in \Sigma$,

$$\begin{aligned}\mu(E) &= \int_E \frac{d\mu}{d\lambda}\bigg|_{\mathscr{F}_2}\lambda(d\omega)\\ &= \int_E \left(\frac{d\mu}{d\lambda}\bigg|_{\mathscr{F}_2}\bigg/\frac{d\mu}{d\lambda}\bigg|_{\mathscr{F}_1}\right)\frac{d\mu}{d\lambda}\bigg|_{\mathscr{F}_1}\lambda(d\omega)\\ &= \int_E \left(\frac{d\mu}{d\lambda}\bigg|_{\mathscr{F}_2}\bigg/\frac{d\mu}{d\lambda}\bigg|_{\mathscr{F}_1}\right)\mu'(d\omega).\end{aligned}$$

It follows that

$$\frac{d\mu}{d\lambda}\bigg|_{\mathscr{F}_2} = \frac{d\mu}{d\lambda}\bigg|_{\mathscr{F}_1}\frac{d\mu}{d\mu'} \quad \text{a.e. } \mu'.$$

Taking the logarithm of both sides and integrating with respect to μ completes the argument. □

Suppose that $\hat{Q}$ is defined as in (3.1).

LEMMA A.3. *Either* $h_{\mathscr{F}_n^{-\infty}}(Q, \hat{Q}) = +\infty$ *for all* $n > 0$ *or*

$$h_{\mathscr{F}_n^{-\infty}}(Q, \hat{Q}) = nH(Q|\pi).$$

PROOF. If $H(Q|\pi) = +\infty$, then $h_{\mathscr{F}_n^{-\infty}}(Q, \hat{Q}) = +\infty$ for all $n > 0$. It may then be assumed that $H(Q|\pi) < \infty$. To argue by induction, assume $h_{\mathscr{F}_j^{-\infty}}(Q, \hat{Q}) < \infty$. Then by Lemma A.2,

$$h_{\mathscr{F}_{j+1}^{-\infty}}(Q, \hat{Q}) = h_{\mathscr{F}_j^{-\infty}}(Q, \hat{Q}) + h_{\mathscr{F}_{j+1}^{-\infty}}(Q, Q'),$$

where $Q' = \int \hat{Q}_\omega Q(d\omega)$, $\hat{Q}_\omega$ the conditional probability distribution of $\hat{Q}$ given $\mathscr{F}_j^{-\infty}$. But $\hat{Q}_\omega = \delta_\omega \otimes_j P_{\omega(j)}$, so that, using the stationary of Q, $h_{\mathscr{F}_{j+1}^{-\infty}}(Q, Q') = H(Q|\pi)$. □

LEMMA A.4.

$$\lim_{n\to\infty} \frac{1}{n} h_{\mathscr{F}_{n-1}^0}(Q, \hat{Q}) = H(Q|\pi). \tag{A.2}$$

PROOF. By Lemma A.3, either $h_{\mathscr{F}_n^{-\infty}}(Q, \hat{Q}) = +\infty$ for all n or

$$\frac{1}{n} h_{\mathscr{F}_{n-1}^0}(Q, \hat{Q}) \le \frac{1}{n} h_{\mathscr{F}_{n-1}^{-\infty}}(Q, \hat{Q}) = \frac{n-1}{n} H(Q|\pi). \tag{A.3}$$

Then, if for some $k > 0$, $h_{\mathscr{F}_k^0}(Q, \hat{Q}) = +\infty$, both sides of (A.2) are equal to $+\infty$. It may then be assumed that for all $k > 0$, $h_{\mathscr{F}_k^0}(Q, \hat{Q}) < \infty$. Applying Lemma A.2 gives

$$h_{\mathscr{F}_{j+1}^0}(Q, \hat{Q}) - h_{\mathscr{F}_j^0}(Q, \hat{Q}) = h_{\mathscr{F}_{j+1}^0}(Q, Q'), \tag{A.4}$$

where $Q' = \int \hat{Q}_\omega Q(d\omega)$, $\hat{Q}_\omega$ the conditional distribution of $\hat{Q}$ given $\mathscr{F}_j^0$. Here $\hat{Q}_\omega = \delta_\omega \otimes_j P_{\omega(j)}$ considered as a measure on $\mathscr{F}_{j+1}^0$. Recalling the definition of $h^j(Q|\pi)$ in Section 5 and using the stationarity of Q,

$$h_{\mathscr{F}_{j+1}^0}(Q, Q') = h_{\mathscr{F}_1^{-j}}(Q, \hat{Q}) = h^j(Q|\pi).$$

From (A.4), it follows that

$$\frac{1}{n} h_{\mathscr{F}_{n-1}^0}(Q, \hat{Q}) = \frac{1}{n} \sum_{j=1}^{n-1} h^j(Q|\pi). \tag{A.5}$$

The sequence $\{h^j(Q|\pi)\}$ is increasing. If it increases without bound, it follows from (A.5) and (A.3) that both sides of (A.2) are equal to $+\infty$. Otherwise, there is some M so that $h^j(Q|\pi) \le M$. It follows from Moy (1961), Lemma 3, that

$$\lim_{j\to\infty} h^j(Q|\pi) = H(Q|\pi),$$

concluding the proof of the lemma. □

LEMMA A.5. *Let Λ_j denote the set of continuous functions ϕ on Ω depending only on the coordinates ω_i, $0 \le i \le j$, which satisfy $E^{P_x}\{e^\phi\} \le 1$ for all $x \in X$. Assume the transition probability function $\pi(dy|x)$ satisfies assumption 3. Then*

$$h_{\mathscr{F}_j^0}(Q, \hat{Q}) = \sup_{\phi\in\Lambda_j} E^Q\{\phi\}.$$

PROOF. By (2.1),

$$h_{\mathscr{F}_j^0}(Q, \hat{Q}) = \sup_{u\in\mathscr{U}'} \left[\int_\Omega \log u(\omega) Q(d\omega) - \log \int_\Omega E^{P_\omega}(u) Q(d\omega)\right],$$

where $\mathscr{U}'$ consists of the positive, continuous functions depending only on the coordinates ω_i, $0 \le i \le j$. Writing $\log u(\omega) = \phi(\omega)$ for $u(\omega) \in \mathscr{U}'$ shows

$$h_{\mathscr{F}_j^0}(Q, \hat{Q}) \ge \sup_{\phi \in \Lambda_j} E^Q\{\phi\}.$$

Let Φ denote the set of continuous function depending on the coordinates ω_i, $0 \le i \le j$. For $\psi \in \Phi$, define

$$\bar{\psi}(x) = \log E^{P_x}\{e^{\psi}\}.$$

Under assumption 3 on $\pi(dy|x)$, $\bar{\psi}(x)$ is a continuous function of x. Let $\phi(\omega) = \psi(\omega) - \bar{\psi}(\omega(0))$. Then

$$\begin{aligned} E^{P_x}\{e^{\phi}\} &= e^{P_x}\{e^{\psi(\omega) - \bar{\psi}\omega(0)}\} \\ &= e^{-\bar{\psi}(x)} E^{P_x}\{e^{\psi}\} = 1, \end{aligned}$$

so $\phi \in \Lambda_j$. Then

$$\begin{aligned} h_{\mathscr{F}_j^0}(Q, \hat{Q}) &= \sup_{\psi \in \Phi} \left[\int_{\Omega} \psi(\omega) Q(d\omega) - \log \int_{\Omega} E^{P_\omega}(e^{\psi}) Q(d\omega) \right] \\ &\le \sup_{\psi \in \Phi} \left[\int_{\Omega} \psi(\omega) Q(d\omega) - \int_{\Omega} \log E^{P_\omega}(e^{\psi}) Q(d\omega) \right] \end{aligned}$$

by Jensen's inequality. The right-hand side

$$\begin{aligned} &= \sup_{\psi \in \Phi} \left[\int_{\Omega} (\psi(\omega) - \bar{\psi}(\omega(0))) Q(d\omega) \right] \\ &\le \sup_{\phi \in \Lambda_j} E^Q\{\phi\}. \end{aligned}$$

□

LEMMA A.6. *Suppose ϕ is $\mathscr{F}_{N-1}^0$ measurable and $E^{P_x}\{e^{\phi}\} \le 1$ for all $x \in X$. Then*

$$E^{P_x}\left\{ \exp\left(\frac{1}{N} \sum_{i=0}^{n-1} \phi(\theta_i \omega) \right) \right\} \le 1 \tag{A.6}$$

for all n.

PROOF. For $j = 0, 1, \dots, N-1$, define

$$\psi_j(\omega) = \sum_{\substack{k:\, k \ge 0 \\ j + kN \le n}} \phi(\theta_{j+kN}\omega).$$

The left-hand side of (A.6) is

$$E^{P_x}\left\{ \exp\left(\frac{1}{N} \sum_{j=0}^{N-1} \psi_j(\omega) \right) \right\}.$$

Jensen's inequality implies

$$E^{P_x}\left\{\exp\left(\frac{1}{N}\sum_{j=0}^{N-1}\psi_j(\omega)\right)\right\} \le E^{P_x}\left\{\frac{1}{N}\sum_{j=0}^{N-1}\exp\psi_j(\omega)\right\}$$
$$= \frac{1}{N}\sum_{j=0}^{N-1}E^{P_x}\{\exp\psi_j(\omega)\}.$$

Under the hypothesis on ϕ, $E^{P_x}\{\exp\psi_j(\omega)\} \le 1$.

Define a measure on $M_S(\Omega)$ by

$$\Gamma_{n,x}(A) = P_x\{\omega \in \Omega,\, R_{n,\omega} \in A\}.$$ □

Corollary A.7. *Suppose that ϕ is a bounded $\mathscr{F}_{N-1}^0$-measurable function satisfying $E^{P_x}\{e^{\phi}\} \le 1$ for all x. Then*

$$E^{\Gamma_{n,x}}\left\{\exp\left(\frac{n}{N}\int_\Omega\phi(\omega)Q(d\omega)\right)\right\}$$
$$\le \exp\left\{2\sup_{\omega\in\Omega}\phi(\omega)\right\}.$$

Proof.

$$E^{\Gamma_{n,x}}\left\{\exp\left(\frac{n}{N}\int_\Omega\phi(\omega)Q(d\omega)\right)\right\}$$
$$= E^{P_x}\left\{\exp\left(\frac{n}{N}\int_\Omega\phi(\omega')R_{n,\omega}(d\omega')\right)\right\}.$$

Now

$$\int_\Omega\phi(\omega')R_{n,\omega}(d\omega') = \frac{1}{n}\sum_{i=0}^{n-1}\phi(\theta_i\omega_n),$$

where ω_n is defined as in Section 1. Then

$$\left|\sum_{i=0}^{n-1}\phi(\theta_i\omega) - n\int_\Omega\phi(\omega')R_{n,\omega}(d\omega')\right| \le 2(N-1)\sup_{\omega\in\Omega}|\phi(\omega)|$$

which, in view of Lemma A.6, establishes the corollary. □

Let

$$J(A) = \limsup_{n\to\infty}\frac{1}{n}\log\sup_{x\in X}\Gamma_{n,x}(A).$$

Lemma A.8. *Let Λ_{N-1} be as defined in the statement of Lemma A.5. For any set $A \in M_S(\Omega)$,*

$$J(A) \le -\sup_{\substack{l:\,A_1,A_2,\dots,A_l\\ A\subset\bigcup_{j=1}^{l}A_j}}\ \inf_{1\le j\le l}\ \sup_{N>0}\ \sup_{\phi\in\Lambda_{N-1}}\ \inf_{Q\in A_j}\frac{1}{N}\int_\Omega\phi(\omega)Q(d\omega).$$

PROOF. From Corollary A.7, for any $A \in M_S(\Omega)$ and any $\phi \in \Lambda_{N-1}$,

$$\Gamma_{n,x}(A) \le \exp\Big(2 \sup_{\omega\in\Omega} \phi(\omega)\Big)\exp\Big(-\frac{n}{N} \inf_{Q\in A} \int_\Omega \phi(\omega) Q(d\omega)\Big).$$

Then

$$J(A) \le -\sup_{N>0} \sup_{\phi\in\Lambda_{N-1}} \inf_{Q\in A} \frac{1}{N}\int_\Omega \phi(\omega)Q(d\omega).$$

The proof is concluded upon the observation that $J(A \cup B) \le \max(J(A), J(B))$. □

LEMMA A.9. *Let A be a closed, thus compact set in $M_S(\Omega)$. Then*

$$\sup_{\substack{l:\, A_1, A_2,\dots,A_l \\ A\subset \cup_{j=1}^l A_j}} \inf_{1\le j\le l} \sup_{N>0} \sup_{\phi\in\Lambda_{N-1}} \inf_{Q\in A_j} \frac{1}{N}\int_\Omega \phi(\omega)Q(d\omega) \ge \inf_{Q\in A} H(Q|\pi).$$

PROOF. From Lemmas A.4 and A.5, it follows that for any $\overline{Q} \in A$ and $\varepsilon > 0$, there is an $N_{\overline{Q}}$ and a $\phi_{\overline{Q}}$ such that

$$\frac{1}{N_{\overline{Q}}}\int_\Omega \phi_{\overline{Q}}(\omega)\overline{Q}(d\omega) \ge \inf_{Q\in A} H(Q|\pi) - \varepsilon/2.$$

Since $\phi_{\overline{Q}}$ is a continuous function on Ω, there is a neighborhood $G_{\overline{Q}}$ of $\overline{Q}$ in $M_S(\Omega)$ such that for $Q \in G_{\overline{Q}}$,

$$\frac{1}{N_{\overline{Q}}}\int_\Omega \phi_{\overline{Q}}(\omega)Q(d\omega) \ge \inf_{Q\in A} H(Q|\pi) - \varepsilon.$$

The neighborhoods $G_{\overline{Q}}$ form an open cover of the compact set A. Let $G_1, G_2, \dots, G_l$ be a finite subcover. Then

$$\inf_{1\le j\le l} \sup_{N>0} \sup_{\phi\in\Lambda_{N-1}} \inf_{Q\in G_j} \frac{1}{N}\int_\Omega \phi(\omega)Q(d\omega) \ge \inf_{Q\in A} H(Q|\pi) - \varepsilon.$$

The statement of the lemma follows. □

Acknowledgment. The author would like to thank I. Csiszár for suggesting this problem to her.

REFERENCES

CSISZÁR, I. (1967). Information-type measures of difference of probability distributions and indirect observations. *Studia Sci. Math. Hungar.* **2** 299–318.

CSISZÁR, I. (1975). *I*-divergence geometry of probability distributions and minimization problems. *Ann. Probab.* **3** 146–158.

CSISZÁR, I. (1984). Sanov property, generalized *I*-projection and a conditional limit theorem. *Ann. Probab.* **12** 768–793.

CSISZÁR, I., COVER, T. and CHO, B. S. (1987). Conditional limit theorems under Markov conditioning. *IEEE Trans. Inform. Theory* **IT-33** 788–801.

DONSKER, M. D. and VARADHAN, S. R. S. (1975). Asymptotic evaluation of certain Markov process expectations for large time. I. *Comm. Pure Appl. Math.* **28** 1–47.

DONSKER, M. D. and VARADHAN, S. R. S. (1976). Asymptotic evaluation of certain Markov process expectations for large time. III. *Comm. Pure Appl. Math.* **29** 389–461.

DONSKER, M. D. and VARADHAN, S. R. S. (1983). Asymptotic evaluation of certain Markov process expectations for large time. IV. *Comm. Pure Appl. Math.* **36** 183–212.

EDWARDS, R. E. (1965). *Functional Analysis and Applications.* Holt, Rinehart and Winston, New York.

HARRIS, T. E. (1956). The existence of stationary measures for certain Markov processes. *Proc. Third Berkeley Symp. Math. Statist. Probab.* **2** 113–124. Univ. California Press, Berkeley.

JUSTESEN, J. and HOHOLDT, T. (1984). Maxentropic Marov chains. *IEEE Trans. Inform. Theory* **IT-30** 665–667.

KEMPERMAN, J. H. B. (1969). On the optimum rate of transmitting information. *Probability and Information Theory. Lecture Notes in Math.* **89** 126–169. Springer, New York.

KREIN, M. and RUTMAN, M. (1948). Linear operators leaving invariant a cone in a Banach space. *Uspekhi. Math. Nauk* **3** 3–95. [*Amer. Math. Soc. Trans.* (1950) **10** 199–235.]

KULLBACK, S. (1967). A lower bound for discrimination information in terms of variation. *IEEE Trans. Inform. Theory* **IT-13** 126–127.

MOY, S. C. (1961). Generalizations of Shannon–McMillan Theorem. *Pacific J. Math.* **11** 705–714.

PARTHASARATHY, K. R. (1967). *Probability Measures on Metric Spaces.* Academic, San Diego,

PINSKER, M. S. (1964). *Information and Information Stability of Random Variables and Processes.* Holden-Day, San Francisco.

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

SANOV, I. N. (1957). On the probability of large deviations of random variables. *Mat. Sb.* **42** 11–14. [*Sel. Trans. Math. Statist. Probab.* (1961) **1** 213–244.]

SPITZER, F. (1972). A variational characterization of finite Markov chains. *Ann. Math. Statist.* **43** 303–307.

MATHEMATICS DEPARTMENT
UNIVERSITY OF MASSACHUSETTS
LOWELL, MASSACHUSETTS 01854