1. Introduction. Suppose \( X_1, X_2, \ldots \) is a sequence of independent random variables taking values in an arbitrary measure space \((S, \mathcal{B})\) with common distribution \( P_X \). The empirical distribution of a sample \( s = (s_1, \ldots, s_n) \in S^n \) 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 \( n \)th Cartesian power of \( P_X \), the probability that the empirical distribution \( \hat{P}_n \) of \((X_1, \ldots, X_n)\) belongs to a set \( C \) of probability measure on \((S, \mathcal{B})\) is given by

\[
P\{\hat{P}_n \in C\} = P_X^n(A_n), \quad A_n = \{s: \hat{P}_n(s, \cdot) \in C\}.
\]

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

\[
\lim_{n \to \infty} \frac{1}{n} \log P\{\hat{P}_n \in C\} = -h(C, P_X),
\]

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

\[
h(Q, P_X) = \begin{cases} 
\int \log(\frac{dQ}{dP_X}) \, dQ, & \text{if } Q \ll P_X, \\
+\infty, & \text{otherwise.}
\end{cases}
\]

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

$$\bar{P}(\hat{P}_n \in C) = P_X^n(\bar{A}_n), \quad P(\hat{P}_n \in C) = P_X^n(A_n),$$

where $\bar{A}_n \supset A_n$ and $A_n \subset A_n$ respectively are sets in $\mathcal{B}^n$ 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

$$h(Q, P_X) = \sup_{\mathcal{B}} h_{\mathcal{B}}(Q, P_X), \quad h_{\mathcal{B}}(Q, P_X) = \sum_{i=1}^k Q(B_i) \log \frac{Q(B_i)}{P_X(B_i)},$$

where $\mathcal{B} = (B_1, \ldots, B_k)$ 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 $\Pi$ on $(S, \mathcal{B})$ is completely convex if for every probability space $(\Omega, \mathcal{F}, \mu)$ and $\mathcal{F}$-measurable mapping $\omega \to \nu(\omega, \cdot) \in \Pi$, the probability measure $\mu \nu$ defined by

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

also belongs to $\Pi$. A convex set of probability measures $\Pi$ is almost completely convex if there exist completely convex subsets $\Pi_1 \subset \Pi_2 \subset \cdots$ of $\Pi$ such that $\bigcup_{k=1}^\infty \Pi_k \supset \Pi \cap \Lambda_f, \Lambda_f$ the set of probability measures on $(S, \mathcal{B})$ 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 $X_1, \ldots, X_n$ 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 $X^n = (X_1, \ldots, X_n)$ 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

$$\lim_{n \to \infty} \frac{1}{n} h(P_X^n|\hat{P}_n \in C, (P^*)^n) = 0,$$

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

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

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

$$\lim_{n \to \infty} \inf \frac{1}{n} \log \frac{1}{n} \log P_x\left[\hat{P}_n(\omega, \cdot) \in G\right] \geq \inf_{\mu \in G} I(\mu),$$

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

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

where $\mathcal{U}_1$ 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 $C \subset \mathcal{M}(X)$,

$$\limsup_{n \to \infty} \frac{1}{n} \log \sup_{x \in X} P_x\left[\hat{P}_n(\omega, \cdot) \in C\right] \leq - \inf_{\mu \in C} I(\mu)$$

[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 $\mu$ so that if $C$ is a closed set in $\mathcal{M}(X)$,

$$I(C) = \inf_{\mu \in C} I(\mu) = I(\mu^*)$$
for some $\mu^* \in C$. Let $\Lambda_0$ 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 $\Lambda_0$ with marginals equal to $\mu^*$ which is naturally associated with $I(\mu^*)$. We define such a $P^*$ to be an $I$-projection of $\pi$ 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 $\Omega = \Pi_{i=-\infty}^{\infty} X_i$, $X_i = X$ for all $i$, and let $\mathcal{B}$ be the $\sigma$-field of Borel sets of $\Omega$. As before, $(\Omega, \mathcal{B}) = \Pi_{i=-\infty}^{\infty} (X_i, \mathcal{B}_i)$, where for each $i$, $\mathcal{B}_i$ is the $\sigma$-field of Borel sets on $X$. $\Omega$ 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 $\omega_n$ by

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

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

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

$R_{n, \omega}$ is the $n$th-order empirical distribution of $\omega_n$. 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 $(\Omega, \mathcal{B})$. For each $n$, $R_{n, \omega}(\cdot) : \Omega \to M_S(\Omega)$ is a continuous map of $(\omega_0, \ldots, \omega_{n-1})$. Let $C$ be a closed convex set with nonempty interior satisfying $I(C^0) < \infty$. Let $P^*$ be the $I$-projection of $\pi$ onto $C$. Then $P^*$ defines a stationary Markov process on $(\Omega, \mathcal{B})$ 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 $M_S(\Omega)$, $R_{n, \omega}(\cdot)$ converges to $P^*$ in conditional $P_\pi$-probability given $\hat{P}_n(\omega, \cdot) \in C$, uniformly for $\omega \in X$. Also for each $A$ a Borel set in $\mathcal{B}$ and each $n > 0$, $R_{n, \omega}(A) : \Omega \to \mathbb{R}$ is a measurable function of $\omega_0, \ldots, \omega_{n-1}$. Then it is possible to define stationary processes

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

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


Let \( u(x) \) be a probability density function with respect to \( \lambda(dx) \). Let \( P_n \) be the Markov process on \( \prod_{i=0}^{\infty}(X_i, \mathcal{B}_i) \), \( X_i = X \), \( \mathcal{B}_i \) the \( \sigma \)-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

\[
R_{n, u}^C(\cdot) = E^{P_n}[R_{n, u}(\cdot) | \hat{P}_n(\omega, \cdot) \in C]
\]

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

\[
E^{P_n}[R_{n, u}(B)|\hat{P}_n(\omega, \cdot) \in C] = P^{\lambda^n}[B|\hat{P}_n(\omega, \cdot) \in C],
\]

where \( \lambda^n \) is the \( n \)th 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 \( \omega_0, \omega_1, \ldots, \omega_m \) under the condition \( \hat{P}_n \in C \) converges to \( (P^*)^m \) as \( n \to \infty \), \( P^* \) the \( I \)-projection of \( \lambda \) 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 \( (X_0, \ldots, X_{n-1}) \), \( n = 1, 2, \ldots \) is said to be asymptotically quasi-Markov if there exists a stationary transition probability function \( Q(dy|x) \) such that

\[
\lim_{n \to \infty} \frac{1}{n} h(P_n, Q_n) = 0,
\]

where \( Q_n \) is the probability measure on \( (X_0, \ldots, X_{n-1}) \) 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 \( \prod_{i=0}^{\infty}(X_i, \mathcal{B}_i) \) give a sequence which is asymptotically quasi-Markov with respect to the transition probability function of \( P^* \) which is uniquely defined a.e. \( \lambda \). When the sequence of measures \( P_n \) in the definition comes from the restriction of stationary processes \( R_n \) on \( \Omega \) to \( \prod_{i=0}^{n-1}(X_i, \mathcal{B}_i) \), 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 \( P_n \) distribution of \( X_0, \ldots, X_{n-1} \) 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 \( R_{n,u}^C \) described above, so the implications of the asymptotic quasi-Markov property are less significant. However, one consequence is as follows. Let

\[(1.11) \quad P_n(\cdot) = P_u(\cdot) | \hat{P}_n(\omega, \cdot) \in C).\]

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 \( B_n \in \mathcal{F}_n-1, \)

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

if

\[
\overline{Q}_n(B_n) \leq \exp(-an), \quad n = 1, 2, \ldots,
\]

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

\[
P_n(B_n) \log \frac{P_n(B_n)}{Q_n(B_n)} + (1 - P_n(B_n)) \log \frac{1 - P_n(B_n)}{1 - Q_n(B_n)} \leq h(P_n, \overline{Q}_n).
\]

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

\[
h^1(Q|\pi) = h(Q, \bar{P}),
\]

where \( \bar{P} \) is the measure \( q(dx)\pi(dy|x), q(dx) \) the marginal of \( Q \). An I-projection \( P^* \) of \( \pi \) 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 \( \sigma \)-field given by the Borel sets, then \( h(Q, P) \) defined by (1.2) can alternatively be defined as

\[
(2.1) \quad h(Q, P) = \sup_{\mu \in \mathcal{F}_1} \left[ \int_X \log u(x) Q(dx) - \log \int_X u(x) P(dx) \right],
\]

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

\[
h^1(Q|\pi) = \sup_{u \in \mathcal{F}_1} \left[ \int \int_{X \times X} \log u(x, y) Q(dx, dy)
\right.
\]

\[
- \log \int \int_{X \times X} u(x, y) q(dx) \pi(dy|x) \right],
\]

\[
+ \int \int_{X \times X} \log u(x, y) Q(dx, dy)
\]

\[
- \log \int \int_{X \times X} u(x, y) q(dx) \pi(dy|x).
\]
where \( \mathcal{V}_1 \) is the set of continuous functions \( u(x, y) > 0 \) on \( X \times X \). For each \( u \in \mathcal{V}_1 \),

\[
\int \int_{X \times X} \log u(x, y)Q(dx, dy) - \log \int \int_{X \times X} u(x, y)q(dx)\pi(dy|x)
\]

is a continuous, convex function of \( Q \). The lemma follows. \( \square \)

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 \( \pi \) 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 \( (X, \mathcal{B}) \) be a Polish space with \( \mathcal{B} \) the \( \sigma \)-field of Borel sets. Let \( M_\mu \) be the set of probability measures on \( X \times X \) having both marginals \( \mu \). If \( \pi(dy|x) \) is the transition probability function of a discrete parameter Markov process with \( (X, \mathcal{B}) \) 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 infimum is actually achieved. If there exists a reference measure \( \lambda \) 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 \( 0 \leq a(x) < \infty \) a.e. \( \mu \) and \( 0 < b(y) \) a.e. \( \lambda \).

Now

\[
\inf_{Q \in M_C} h^1(Q|\pi) = \inf_{\mu \in C} \inf_{Q \in M_\mu} h^1(Q|\pi)
\]

\[= \inf_{\mu \in C} I(\mu)
\]

\[= I(C).
\]

For \( C \) a closed set, an \( I \)-projection \( P^* \) of \( \pi \) 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 \( \pi \) 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

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

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 $\lambda$.

**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 $\mathcal{M}(X)$ 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 $d\mu_n/d\lambda \geq 1/n \lambda$-a.e. The sequence $\mu_n$ converges in variation to $\mu$ so for sufficiently large $n$, $\mu_n \in C^0$. Let $\bar{\mu}$ be such a $\mu_n$ 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 $\eta$ and so large that $\int_X [m(x) \wedge n] \lambda(dx) \geq 1/2$. Then $\bar{\mu}_n(dx) = m_n(x)\lambda(dx)$. 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

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

**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 $Q \in M_C$,

$$h^1(Q|\pi) \geq h^1(Q|P^*(\cdot \cdot)) + h^1(P^*|\pi).$$
\textbf{Proof.} Suppose \( P \in M_C \) satisfies \( h^1(P|\pi) < \infty \). Consider \( h^1(\epsilon P + (1 - \epsilon)P^*|\pi), 0 \leq \epsilon \leq 1 \). By the convexity of \( h^1(Q|\pi) \) as a function of \( Q \), this is a convex function of \( \epsilon \). Since \( h^1(P^*|\pi) \leq h^1(\epsilon P + (1 - \epsilon)P^*|\pi), h^1(\epsilon P + (1 - \epsilon)P^*|\pi) \) is a nondecreasing function of \( \epsilon \). Then

\begin{equation}
\lim_{\epsilon \to 0} \frac{d}{d\epsilon} h^1(\epsilon P + (1 - \epsilon)P^*|\pi) \geq 0,
\end{equation}

provided the derivatives exist.

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

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

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

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

\begin{equation}
\begin{split}
h^1(\epsilon P + (1 - \epsilon)P^*|\pi) &= \int_\mathcal{X} \int_\mathcal{X} P_\epsilon(x, y) \log \frac{P_\epsilon(x, y)}{p_\epsilon(x) \pi(y|x)} \lambda(dx)\lambda(dy) \\
&= \int_\mathcal{X} \int_\mathcal{X} P_\epsilon(x, y) \log \frac{P_\epsilon(x, y)}{\pi(y|x)} \lambda(dx)\lambda(dy) \\
&\quad - \int_\mathcal{X} p_\epsilon(x) \log p_\epsilon(x) \lambda(dx).
\end{split}
\end{equation}

Further, for \( 0 \leq \epsilon \leq 1 \),

\begin{equation}
\begin{split}
(2.4)(i) & \quad \int_\mathcal{X} \int_\mathcal{X} P_\epsilon(x, y) \left| \log \frac{P_\epsilon(x, y)}{\pi(y|x)} \right| \lambda(dx)\lambda(dy) < \infty, \\
(2.4)(ii) & \quad \int_\mathcal{X} p_\epsilon(x) \log p_\epsilon(x) \lambda(dx) < \infty.
\end{split}
\end{equation}

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

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

Using the bounds

\begin{equation}
\begin{split}
\log^+ \frac{P_\epsilon(x, y)}{\pi(y|x)} &= \log^+ \left[ \frac{P(x, y)}{\pi(y|x)} + (1 - \epsilon) \frac{P^*(x, y)}{\pi(y|x)} \right] \\
&\leq \log^+ \left[ \frac{P(x, y)}{\pi(y|x)} + \frac{P^*(x, y)}{\pi(y|x)} \right] \\
&\leq \log 2 + \log^+ \left[ \frac{P_{1/2}(x, y)}{\pi(y|x)} \right]
\end{split}
\end{equation}

\( P_{1/2}(x, y) \) is a probability measure on \( \mathcal{X} \times \mathcal{X} \).
and
\[ \log^{-\frac{P_e(x,y)}{\pi(y|x)}} = \log^{-\left[ \frac{P(x,y)}{\pi(y|x)} + (1 - \varepsilon) \frac{P^*(x,y)}{\pi(y|x)} \right]} \]

(2.5)(ii)
\[ \leq \log^{-\left[ \min(\varepsilon, 1 - \varepsilon) \left( \frac{P(x,y)}{\pi(y|x)} + \frac{P^*(x,y)}{\pi(y|x)} \right) \right]} \]
\[ \leq \log^{-\left( 2 \min(\varepsilon, 1 - \varepsilon) \right) + \log^{-\left[ \frac{P_{1/2}(x,y)}{\pi(y|x)} \right]}} \]

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

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)
\[ = \int \int_{X \times X} (P(x,y) - P^*(x,y)) \log \frac{P(x,y)}{\pi(y|x)} \lambda(dx) \lambda(dy) \]
\[ - \int_X (p(x) - p^*(x)) \log p_e(x) \lambda(dx). \]

Now using the bound (2.5)(i) and the bound
\[ \log^{-\frac{P_e(x,y)}{\pi(y|x)}} \leq \log^{-\left[ (1 - \varepsilon) \frac{P^*(x,y)}{\pi(y|x)} \right]} \]
\[ \leq \log^{- (1 - \varepsilon) + \log^{-\left( \frac{P^*(x,y)}{\pi(y|x)} \right)}} \]

shows by dominated convergence that
\[ \lim_{\varepsilon \to 0} \int \int_{X \times X} P^*(x,y) \log \frac{P_e(x,y)}{\pi(y|x)} \lambda(dx) \lambda(dy) \]
\[ = \int \int_{X \times X} P^*(x,y) \log \frac{P^*(x,y)}{\pi(y|x)} \lambda(dx) \lambda(dy). \]

Similarly,
\[ \lim_{\varepsilon \to 0} \int_X p^*(x) \log p_e(x) \lambda(dx) = \int_X p^*(x) \log p^*(x) \lambda(dx). \]
It follows from (2.6) and (2.2) that
\[
\lim_{\epsilon \to 0} \int_{X \times X} P(x, y) \log \frac{P_\epsilon(x, y)}{p_\epsilon(x) \pi(y|x)} \lambda(dx) \lambda(dy)
\geq \int_{X \times X} P^*(x, y) \log \frac{P^*(x, y)}{p^*(x) \pi(y|x)} \lambda(dx) \lambda(dy).
\]

Rewriting the integral on the left-hand side as
\[
\int_{X \times X} P(x, y) \log \frac{P(x, y)}{p(x) \pi(y|x)} \lambda(dx) \lambda(dy)
- \int_{X \times X} P(x, y) \log \frac{P(x, y)}{p(x) P_\epsilon(y|x)} \lambda(dx) \lambda(dy)
\]
shows that
\[
\lim_{\epsilon \to 0} \int_{X \times X} P(x, y) \log \frac{P(x, y)}{p(x) P_\epsilon(y|x)} \lambda(dx) \lambda(dy)
\leq h^1(P|\pi) - h^1(P^*|\pi).
\]
We can write the integral on the left-hand side of (2.7) as
\[
\int_X p(x) \lambda(dx) h(P(dy|x), P_\epsilon(dy|x)),
\]
where \( P_\epsilon(dy|x) = P_\epsilon(y|x) \lambda(dy) \). Clearly \( h(P(dy|x), P_\epsilon(dy|x)) \) is defined for \( p(dx) \) a.e. \( x \).

By Fatou’s lemma,
\[
\int_X p(x) \lambda(dx) \liminf_{\epsilon \to 0} h(P(dy|x), P_\epsilon(dy|x))
\leq \lim_{\epsilon \to 0} \int_{X \times X} P(x, y) \log \frac{P(x, y)}{p(x) P_\epsilon(y|x)} \lambda(dx) \lambda(dy).
\]
Now on the set of \( p(dx) \) measure 1 where \( P_\epsilon(dy|x) \) is defined, we see that as \( \epsilon \to 0 \), \( P_\epsilon(dy|x) \) converges in variation to \( P^*(dy|x) \) when \( p^*(x) > 0 \) and \( P_\epsilon(dy|x) \) 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
\[
h \left( P(dy|x), \lim_{\epsilon \to 0} P_\epsilon(dy|x) \right)
\leq \liminf_{\epsilon \to 0} h(P(dy|x), P_\epsilon(dy|x)).
\]
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 
$f_y P^*(x)P^*(y|x)\lambda(dx) = p^*(y)$ for $\lambda$-a.e. $y$ so that $p^*(y) > 0$ for $\lambda$-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

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

or

$$(2.10) \quad h^1(P|\pi) \geq h^1(P|P^*(\cdot|\cdot)) + h^1(P^*|\pi)$$

for $P \in M_C$ satisfying $h^1(P|\pi) < \infty$. For the general case, $P^*(dy|x)$ is only defined $\lambda$-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 $\pi$ 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

$$h^1(P^*_1|\pi) \geq h^1(P^*_2|P^*_1(\cdot|\cdot)) + h^1(P^*_1|\pi).$$

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

$$h^1(P^*_2|P^*_1(\cdot|\cdot)) = \int_X P^*_2(dx)\int_P P^*_2(dy|x)\log \frac{P^*_2(y|x)}{P^*_1(y|x)},$$

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

$$(2.11) \quad P^*_2(dy|x) = P^*_1(dy|x), \quad \lambda \text{ a.e. } x.$$  

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,

$$I^*(P^*_1|P^*_1(\cdot|\cdot)) \leq h^1(P^*_1|P^*_1(\cdot|\cdot)) = 0,$$

so that $p^*(dx)$ 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 $(\Omega, \mathcal{F})$ with transition probability function $P^*_1(dy|x)$ and invariant measure $p^*(dx)$. 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. $\lambda$ 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 \( \pi \) onto \( C \) as above. Then \( P^*(dy|x) \) can be chosen to have the following property: There is a function \( b(y) \), \( 0 < b(y) < \infty \) a.e. \( y \) such that if \( h(y) \in L^1(1/b(y)\lambda(dy)) \), then

\[
\int_X P^*(dy|x) h(y)
\]

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. \( \lambda(dx) \),

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

By altering \( a(x) \) on a set of measure 0, it is possible to assume that this equation holds for all \( x \). Since \( a \leq \pi(y|x) \) for all \( x \) and \( \lambda \)-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 \( x \to L_r(\lambda) \) 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 \( C = \{ \mu \in \mathcal{M}(X) : \int_X f_i d\mu \geq \gamma_i, \ i = 1, \ldots, n \} \) for continuous, real-valued functions \( f_1, f_2, \ldots, f_n \) on \( X \). Suppose there is some \( \mu \in C \) satisfying

\[
(a) \quad \int_X f_i d\mu > \gamma_i \quad i = 1, \ldots, n,
\]

\[
(b) \quad I(\mu) < \infty.
\]

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

Let \( R^+_n \) denote \( \{ \xi \in \mathbb{R}^n, \xi_i \geq 0, \ i = 1, \ldots, n \} \). Let \( T_\xi \) be the mapping of the set of continuous functions on \( X, C(X) \), onto itself given by

\[
T_\xi g(x) = e^{\Sigma_{i=1}^n \xi_i f_i(x)} \int_X g(y) \pi(y|x) \lambda(dy).
\]

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

\[
\int_X V_\xi(x) \psi_\xi(x) \lambda(dx) = 1.
\]

Then

\[
I(C) = \max_{\xi \in \mathbb{R}^+_n} \left( \sum_{i=1}^n \xi_i \gamma_i - \log \rho_\xi \right)
\]

and

\[
P^*(x, y) = \left( V_\xi(y) \pi(y|x) e^{\Sigma_{i=1}^n \xi_i f_i(x)} \psi_\xi(x) \right) / \rho_\xi
\]

for \( \xi \) 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 \( T_\xi \) to be a compact operator [Edwards (1965), Proposition 9.5.17]. The lower bound (assumption 2) on \( \pi(y|x) \) and the continuity of \( f_i, \ i = 1, \ldots, n \), ensure that \( T_\xi \) is a strongly positive operator, so a theorem of Krein and Rutman (1948), Theorem 6.3, proves the existence of \( V_\xi, \psi_\xi \) and \( \rho_\xi \) as in the statement of the theorem.

The set \( C \) described in the theorem is weakly closed and convex. Any measure \( \mu \) 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 \( \pi \) onto \( C \) exists and is unique. To see \( I(C^0) = I(C) \), suppose \( \mu \) satisfies (2.13)(a) and (b) and let \( \nu_1 \) be any element of \( C \). Then \( \nu_\alpha = (1 - \alpha)\mu + \alpha \nu_1 \in C \) and by convexity,

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

The remainder of the theorem is proved in a sequence of four lemmas.
Lemma 2.10. Let $P^*$ be the unique $I$-projection of $\pi$ onto $C$. Then
\begin{equation}
I(C) = h^I(P^*|\pi) = \inf_{Q \in M_C} h(Q, p^*(dx)\pi(dy|x)),
\end{equation}
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
\begin{equation}
h(Q, p^*(dx)\pi(dy|x)) = h^I(Q|\pi) + h(q, p^*).
\end{equation}
From (2.10),
\begin{equation}
h(Q|\pi) \geq h^I(Q|P^*(\cdot|\cdot)) + h^I(P^*|\pi).
\end{equation}
Adding $h(q, p^*)$ to both sides and using $h(Q, P^*) = h(Q|P^*(\cdot|\cdot)) + h(q, P^*)$ shows
\begin{equation}
h(Q, p^*(dx)\pi(dy|x)) \geq h(Q, P^*) + h^I(P^*|\pi)
\geq h(Q, P^*) + \inf_{Q \in M_C} h(Q, p^*(dx)\pi(dy|x)).
\end{equation}
This last equation determines $P^*$ as the unique $I$-projection of $p^*(dx)\pi(dy|x)$ onto $M_C$ [Csiszár (1975), Theorem 2.2]. \qed

Lemma 2.11. Let $f_i(x)$, $i = 1, 2, \ldots, n$, be real-valued measurable functions on a measure space $(X, \mathcal{B}, \lambda)$. 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 $\lambda$-a.e. Suppose there exists a sequence $\sum_{i=1}^{n} \alpha_i^{m} f_i(x)$ converging pointwise $\lambda$-a.e. to a real-valued function $g(x)$. Let $c_{m_i}$ be the sequence
\begin{equation}
(\alpha_1 - \alpha_2, \alpha_1 - \alpha_3, \alpha_1 - \alpha_4, \ldots, \alpha_2 - \alpha_3, \alpha_2 - \alpha_4, \ldots, \alpha_n - \alpha_{n+1}, \alpha_n - \alpha_{n+2}, \ldots).
\end{equation}
Then $\lim_{m \to \infty} \sum_{i=1}^{n} c_{m_i} f_i(x) = 0$. Let
\begin{equation}
c_{m_i}' = c_{m_i}/\max(|c_{m_1}|, |c_{m_2}|, \ldots, |c_{m_n}|, 1).
\end{equation}
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 $(c_{m_i}')$ is 0 from the linear independence of the functions $f_i(x)$, $i = 1, \ldots, n$, $\lambda$-a.e. It follows that the sequence $(c_{m_i}')$ converges to 0. From the definition (2.17), it follows that the sequence $(c_{m_i})$ converges to 0. But then $(\alpha_i)$ 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 $\{f_{i1}\}_{i=1}^{n}$. 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 $(f_1, f_2, \ldots, f_n)$. 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 $V_\xi, \psi_\xi$ and $\rho_\xi$ be defined as in the statement of Theorem 2.9. Then the I-projection $P^*$ of $\pi$ onto $C$ has density

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

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

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

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

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

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

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 $\mathcal{F}$. Since $P^* \sim \lambda \times \lambda$, there exist functions

$$\sum_{i=1}^n \alpha_m f_i(x) - \gamma_i + \sum_{k=1}^M \beta_m g_k(x) - g_k(y),$$

$\alpha_m \geq 0$ and $\beta_m \in \mathbb{R}$ 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 $(m)$ and a sequence of constants $(a_m)$ so that

$$(2.18)(i) \lim_{m \to \infty} \left( \sum_{i=1}^n \alpha_m f_i(x) - \gamma_i + \sum_{k=1}^M \beta_m g_k(x) - a_m \right) = f(x)$$

exists for $\lambda$-a.e. $x$ and

$$(2.18)(ii) \lim_{m \to \infty} \left( - \sum_{k=1}^M \beta_m g_k(y) + a_m \right) = g(y)$$

exists for $\lambda$-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 $\lambda$-a.e. $x$,

$$\lim_{m \to \infty} \sum_{i=1}^n \alpha_m (f_i(x) - \gamma_i) = f(x) + g(x).$$
In view of Lemma 2.11, there exist constants $\xi_i$, $i = 1, \ldots, n$, $\xi_i \geq 0$ such that
\[
\sum_{i=1}^{n} \xi_i (f_i(x) - \gamma_i) = f(x) + g(x), \quad \lambda\text{-a.e.}
\]
Thus the conditional density satisfies
\[
P^*(y|x) = e^{I(C)\xi} e^{\sum_{i=1}^{n} f_i(x)(y) - \gamma_i} e^{-g(x)} e^{g(y)} \pi(y|x), \quad \lambda \times \lambda\text{-a.e.}
\]
Since $P^*(y|x)$ is a transition probability density function, there follows
\[
e^{\sum_{i=1}^{n} \xi_i f_i(x)} \frac{e^{g(y)} \pi(y|x) \lambda(dy)}{e^{I(C)\xi} e^{\sum_{i=1}^{n} \xi_i \gamma_i} e^{g(x)}}, \quad \lambda\text{-a.e.}
\]
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 $T_{\xi}^*$ with positive eigenvalue
\[
(2.19) \quad \rho_{\xi} = \exp \left( \sum_{i=1}^{n} \xi_i \gamma_i - I(C) \right).
\]
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 $\lambda$, there follows
\[
\int p^*(x) e^{-g(x)} e^{\sum_{i=1}^{n} f_i(x)} \pi(y|x) \lambda(dx)
= e^{-I(C)} e^{\sum_{i=1}^{n} \xi_i \gamma_i} p^*(y) e^{-g(y)}.
\]
Then $p^*(y)e^{-g(y)} \in L^1(\lambda)$ is the unique positive eigenvector for $T_{\xi}^*$ corresponding to the same eigenvalue. Since the product $V_{\xi}(x)\psi_{\xi}(x) = p^*(x)$, (2.14) holds. The conclusion of the lemma follows. \qed

**Lemma 2.13.** Under the assumption of Theorem 2.9,
\[
I(C) = \max_{\xi \in \mathbb{R}_+^n} \left( \sum_{i=1}^{n} \xi_i \gamma_i - \log \rho_{\xi} \right),
\]
where $\rho_{\xi}$ is the (positive) eigenvalue of greatest modulus for the operator $T_{\xi}$.

**Proof.** For any vector $\xi \in \mathbb{R}_+^n$, let
\[
P_{\xi}(x,y) = \left( V_{\xi}(y) \pi(y|x) e^{\sum_{i=1}^{n} \xi_i f_i(x)} \psi_{\xi}(x) \right) / \rho_{\xi},
\]
where $V_., \psi_\xi$ and $\rho_\xi$ are as defined in Theorem 2.9. From Lemma 2.12, the $I$-projection $P^*$ of $\pi$ onto $C$ has $\lambda \times \lambda$ density $P^*(x,y) = P_{\xi^*}(x,y)$ for some $\xi^* \in \mathbb{R}_+^n$. It follows from (2.19) that
\[
(2.20) \quad I(C) = \sum_{i=1}^{n} \xi^*_i \gamma_i - \log \rho_{\xi^*}.
\]
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

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

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 $P' \in \Lambda$,

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

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

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

one obtains

$$I(C) \geq \inf_{Q \in \tilde{M}_C} \int \int_{X \times X} Q(x, y) \log \frac{V_\zeta(y) e^{2\gamma - 2 \psi_\zeta(x)}}{V_\zeta(x) \rho_\zeta}$$

$$= \inf_{Q \in \tilde{M}_C} \left( \int \int_{X \times X} Q(x, y) \sum_{i=1}^n \zeta_i f_i(x) - \log \rho_\zeta \right)$$

$$\geq \left( \sum_{i=1}^n \zeta_i \gamma_i - \log \rho_\zeta \right) \text{ for } \zeta \in \mathbb{R}_n^+, \quad (2.21)$$

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 $(\Omega, \mathcal{B})$ be as in Section 1. Let $h_{\mathcal{F}_m}(\cdot , \cdot)$ denote the entropy when the supremum in (2.1) is taken over positive functions $u(x) \in C(\Omega)$, the continuous functions on $\Omega$, which depend only on the coordinates $\omega_i$, $n \leq i \leq m$.

Let $P$ be a measure on $\mathcal{F}_m^s$ with $s \leq t$. Suppose $P(\omega; \omega(t) = \bar{\omega}(t)) = 1$. Define a measure $\delta_\omega \otimes_t P$ on $\Omega$ by

$$\left( \delta_\omega \otimes_t P \right) \{ \omega(t_1) \in A_1, \omega(t_2) \in A_2, \ldots, \omega(t_n) \in A_n \}$$

$$= \chi_{A_1}(\bar{\omega}(t_1)) \chi_{A_2}(\bar{\omega}(t_2)) \cdots \chi_{A_n}(\bar{\omega}(t_k))$$

$$\times P(\omega(t_{k+1}) \in A_{k+1}, \ldots, \omega(t_n) \in A_n).$$

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

$$\tilde{Q} = \int_{\Omega} P_\omega Q(d\omega).$$

(3.1)
For $Q \in M_S(\Omega)$ define the entropy of $Q$ with respect to $\pi$ by

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

By (2.1),

$$H(Q|\pi) = \sup_{u \in \mathcal{F}^-} \left[ \int \log u(\omega)Q(\omega) - \log \int u(\omega)\hat{Q}(\omega) \right],$$

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

$$\int \int u(\omega)\hat{Q}(\omega) = \int E^{P}(u)Q(\omega).$$

Under assumption 3 on the probability transition function $\pi(dy|x)$, $\omega \rightarrow E^{P}(u)$ is a continuous function for $u \in \mathcal{F}$. 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 $\mathcal{M}(X)$ satisfying $I(C) = I(C^0) < \infty$. Let $P^*$ be the $I$-projection of $\pi$ onto $C$ considered as a stationary process on $(\Omega, \mathcal{B})$. 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

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

uniformly for $x \in X$. Let $\Pi_C$ 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 $(\Omega, \mathcal{B})$ is a Polish space, the weak topology on the set of probability measures on $\Omega$ is metrizable. Let $\Delta(\cdot, \cdot)$ denote this metric. Define

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

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

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

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 \( H(\Pi_C^*|\pi) = \inf_{Q \in \Pi_C} H(Q|\pi) \) 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 \( \epsilon > 0 \), \( H(\Pi_C^*|\pi) > H(\Pi_C|\pi) \). Fix \( \epsilon \) and pick \( \epsilon^1 \) such that \( 2\epsilon^1 < H(\Pi_C^*|\pi) - H(\Pi_C|\pi) \).

It follows from (3.3) and (3.4) that \( \exists N \) such that for \( n \geq N \) and every \( x \in X \),

\[
P_x \{ \Delta(R_n, \omega, P^*) \geq \epsilon | \hat{P}_n(\omega, \cdot) \in C \} \leq e^{-n(H(\Pi_C^*|\pi) - \epsilon^1)} \leq e^{-n(H(\Pi_C^*|\pi) - \epsilon^1) - 2\epsilon^1},
\]

so that

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

which establishes the theorem. \( \Box \)

**Theorem 3.2.** The stationary processes defined by

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

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

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

\[
\lim_{n \to \infty} P_x \left\{ \left| \int f dR_n - \int f dP^* \right| \geq \epsilon | P_n(\omega, \cdot) \in C \right\} = 0,
\]

uniformly for \( x \in X \).

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

\[
\int_{\Omega} f dR_{n, x}^C = E^{P_x} \left\{ \int f dR_n | \hat{P}_n(\omega, \cdot) \in C \right\}.
\]
Then for \( f \in C(\Omega) \),
\[
\left| \int_\Omega f \, dR^c_{n,x} - \int_\Omega f \, dP^* \right|
\leq E^{P_x}\left( \left| \int_\Omega f \, dR_{n,\omega} - \int_\Omega f \, dP^* \right| \mid \hat{P}_n(\omega, \cdot) \in C \right)
\leq 2|f|P_x\left( \left| \int_\Omega f \, dR_{n,\omega} - \int_\Omega f \, dP^* \right| \geq \epsilon \mid \hat{P}_n(\omega, \cdot) \in C \right)
+ \epsilon \left( P_x\left( \left| \int_\Omega f \, dR_{n,\omega} - \int_\Omega f \, dP^* \right| < \epsilon \mid \hat{P}_n(\omega, \cdot) \in C \right) \right),
\]
so by (3.5),
\[
\lim_{n \to \infty} \left| \int_\Omega f \, dR^c_{n,x} - \int_\Omega f \, dP^* \right| \leq \epsilon.
\]
Since \( \epsilon \) is arbitrary, the weak convergence of \( R^c_{n,x} \) to \( P^* \) is established.

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

**Lemma 3.3.** Let \( P^* \), \( \Pi_C \) 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 \( \mathcal{F}_1^0 \). Let \( \hat{Q}_{1,\omega}^0 \) be the regular conditional probability distribution of \( \hat{Q} \) given \( \mathcal{F}_1^0 \). Then \( E^{\hat{Q}_{1,\omega}}[dQ/d\hat{Q}] \) is a version of \( dQ_1^0/d\hat{Q}_1^0 \). It follows that
\[
I(C) \leq h^1(Q_1^0|\pi)
= \int E^{\hat{Q}_{1,\omega}} \frac{dQ}{d\hat{Q}} \log E^{\hat{Q}_{1,\omega}} \left[ \frac{dQ}{d\hat{Q}} \right] d\hat{Q}_1^0(\omega)
\leq \int E^{\hat{Q}_{1,\omega}} \left[ \frac{dQ}{d\hat{Q}} \log \frac{dQ}{d\hat{Q}} \right] d\hat{Q}_1^0(\omega)
= \int dQ \log \frac{dQ}{d\hat{Q}} = I(C),
\]
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 \( Q_{1,0} - \text{a.e.} \ \omega \). Since \( x \log x \) is strictly convex, this implies for \( Q - \text{a.e.} \ \omega , \)

\[
\frac{dQ}{dQ} (\omega) = E^{Q_{1,0}} \left[ \frac{dQ}{dQ} \right].
\]

Let \( Q_{\omega(0)} \) denote the regular conditional probability distribution of \( Q_{1}^{0} \) given \( F_{0}^{0} \) and note that \( P_{\omega(0)} \) is the regular conditional probability distribution of \( Q_{1}^{0} \) given \( F_{0}^{0} \). The measures \( Q_{1}^{0} \) and \( Q_{1}^{0} \) have the same marginal distribution on \( F_{0}^{0} \), which we denote by \( q \). Then \( dQ_{1}/dQ_{1}^{0} \) is the Radon-Nikodym derivative of \( Q_{\omega(0)} \) with respect to \( P_{\omega(0)} \) which exists for \( q - \text{a.e.} \ \omega(0) \). It now follows from (3.6) that

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

This shows \( Q \) is a stationary Markov process as follows: Let \( B \in \mathcal{F}_{0}^{-j} \), \( A \in \mathcal{F}_{1}^{1} \):

\[
Q[A \cap B] = \int_{A \cap B} \frac{dQ}{dQ} Q(d\omega) = \int_B E^{P_{\omega}} \left[ \frac{dQ}{dQ} \chi_A \right] Q(d\omega) = \int_B E^{P_{\omega}} \left[ \frac{dQ_{\omega(0)}}{dP_{\omega(0)}} \chi_A \right] Q(d\omega) \quad \text{[by (3.1)]}
\]

\[
= \int_B E^{P_{\omega(0)}} \left[ \frac{dQ_{\omega(0)}}{dP_{\omega(0)}} \chi_A \right] Q(d\omega) = \int_B Q_{\omega(0)}(A) Q(d\omega).
\]

Since \( Q_{\omega(0)}(A) = E^{Q}[A|F_{0}^{0}] \), this shows that \( E^{Q}[A|F_{0}^{j-1}] = E^{Q}[A|F_{0}^{0}] \) for any \( A \in \mathcal{F}_{1}^{1} \), \( j > 0 \). It follows that \( Q \) is a stationary Markov process with transition probability function \( Q(A|x) = Q_{x}(A) \) for \( A \in \mathcal{F}_{1}^{1} \). 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 \( \lambda \), let

\[
R_{n,u}^{C}(\cdot) = E^{P_{u}} \left\{ R_{n,u}(\cdot) | \hat{P}_{n}(\omega, \cdot) \in C \right\}
\]

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

We begin by establishing a fundamental lemma. Let \( Q \) be a stationary process on \((\Omega, \mathcal{B})\) with marginal \( q \). Let \( Q_{1}^{0} \) be the restriction of \( Q \) to \( F_{1}^{0} \) and let \( Q_{\omega(0)} \) denote the regular conditional probability distribution of \( Q_{1}^{0} \) w.r.t.
\( S^0 \). Then, as before, for \( A \in S^1 \), \( Q(A|x) = Q^0(A) \) 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 \( \prod_{i=0}^\infty (X_i, \mathcal{B}_i), X_i = X, \mathcal{B}_i \) the Borel \( \sigma \)-field on \( X \), \( 0 \leq i < \infty \), with a probability transition function \( P(dy|x) \) and initial distribution \( q(dx) \). Let \( Q, \tilde{Q} \) be as above. Then for any \( n \),

\[
h_{S^0_n}(Q, \bar{P}) = h_{S^0_n}(Q, \tilde{Q}) + nh^1(Q_1^0|P(\cdot|\cdot)) .
\]

**Proof.** Let \( \bar{P}^0 \) denote the restriction of \( \bar{P} \) to \( S^0 \). Then we can assume \( Q \ll \bar{Q} \) on \( S^0 \) and \( Q^0 \ll \bar{P}^0 \); otherwise both sides are \( \infty \). To establish this, suppose \( h_{S^0}(Q, \bar{P}) < \infty \). Then \( Q \ll \bar{P} \) on \( S^0 \), in particular, \( Q^0 \ll \bar{P}^0 \). Since \( Q^0 \) and \( \bar{P}^0 \) both have marginal \( q \) on \( S^0 \), \( dQ^0/d\bar{P}^0 \) 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 \( M \in S^0, Q(M) = 0 \). However,

\[
\tilde{Q}(M) = \int_M \frac{dQ^0}{d\bar{P}^0_1}(\omega_0, \omega_1) \cdots \frac{dQ^0}{d\bar{P}^0_n}(\omega_{n-1}, \omega_n) \frac{d\bar{P}(\omega_0, \ldots, \omega_n)}{d\bar{P}^0}(\omega_0, \ldots, \omega_n),
\]

so that \( \tilde{Q}(M) = 0 \) implies that \( \bar{P} \)-a.e. on \( S^0 \), \( dQ^0/d\bar{P}^0(\omega_0, \omega_1) \cdots dQ^0/d\bar{P}^0_n(\omega_{n-1}, \omega_n) = 0 \). Let \( N \) be the \( S^0 \) set of \( \bar{P} \)-measure 0, where this product is positive. Let \( T_i = (\omega_{i-1}, \omega_i) : dQ^0_1/d\bar{P}^0_1(\omega_{i-1}, \omega_i) = 0 \). 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 \( Q \ll \bar{Q} \) on \( S^0 \) and \( Q^0 \ll \bar{P}^0 \), which we have seen implies \( \tilde{Q} \ll \bar{P} \) on \( S^0 \), we have

\[
\frac{dQ}{d\bar{P}} = \frac{dQ}{d\tilde{Q}} \frac{d\tilde{Q}}{d\bar{P}}
\]

on \( S^0 \). Taking log of both sides, integrating over \( Q \) and using (4.2) gives

\[
h_{S^0}(Q, \bar{P}) = h_{S^0}(Q, \tilde{Q}) + \int dQ(\omega_0, \ldots, \omega_n) \log \frac{dQ^0}{d\bar{P}^0_1}(\omega_0, \omega_1) \cdots \frac{dQ^0}{d\bar{P}^0_n}(\omega_{n-1}, \omega_n).
\]

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

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

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

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 $\mathcal{F}_{n-1}^0$-measurable set of $\omega$ 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 $(\omega_0, \ldots, \omega_{n-1})$. It follows that

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

converges as $j \to \infty$ to

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

$$= \int \omega \ dR_{n,u}^C.$$

Thus the measure

$$\sum_{k=0}^{k-k} a_{jk} R_{n,\omega_{e_{jk}}}$$

converges weakly as $j \to \infty$ to $R_{n,u}^C$. 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 $\Pi_C$, each of the measures in (4.3) is in $\Pi_C$. Thus $R_{n,\omega}$ is a limit point of $\Pi_C$, which, being closed, implies $R_{n,u}^C \in \Pi_C$. □

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

$$R_{n,u}^C(\omega_0, \ldots, \omega_{n-1}) = \frac{1}{P_u\{E_n\}} \chi_{E_n}(\omega_0, \ldots, \omega_{n-1}) \frac{1}{n} \sum_{i=0}^{n-1} \pi_i(\omega_0, \ldots, \omega_{n-1}),$$
where
\[
\pi_0(\omega_0, \ldots, \omega_{n-1}) = u(\omega_0) \pi(\omega_1|\omega_0) \cdots \pi(\omega_{n-1}|\omega_{n-2})
\]
and
\[
\pi_i(\omega_0, \ldots, \omega_{n-1}) = \pi(\omega_0|\omega_{n-1}) \pi(\omega_1|\omega_0) \cdots \pi(\omega_{n-i-1}|\omega_{n-i-2})
\times u(\omega_{n-i}) \pi(\omega_{n-i+1}|\omega_{n-i}) \cdots \pi(\omega_{n-1}|\omega_{n-2}).
\]
(4.4)

**Proof.** Observe that the map \( \Omega \to \Omega \) defined by \( \omega \mapsto \omega_n \) is continuous, hence \( \mathcal{F}_n^0 \)-measurable, and the maps \( \theta_i, \theta_i^{-1}; \Omega \to \Omega \) are continuous, hence measurable. Let \( i > 0 \) and let \( A \) be a measurable set in \( \mathcal{F}_n^{-1} \). Then
\[
E_u[\chi_A(\theta_i \omega_n)]
\]
(4.5)
\[
= \int_X \cdots \int_X \chi_A(\theta_i \omega_n) \pi_0(\omega_0, \ldots, \omega_{n-1}) \, d\lambda^n
\]
\[
= \int_X \cdots \int_X \chi_A(\omega_n) \pi_0(\theta_i^{-1} \omega_n) \, d\lambda^n
\]
by Fubini's theorem. It is easy to see that for \( i > 0 \),
\[
\pi_0(\theta_i^{-1} \omega_n) = \pi(\omega_0|\omega_{n-1}) \pi(\omega_1|\omega_0) \cdots \pi(\omega_{n-i-1}|\omega_{n-i-2})
\times u(\omega_{n-i}) \pi(\omega_{n-i+1}|\omega_{n-i}) \cdots \pi(\omega_{n-1}|\omega_{n-2}).
\]
(4.6)

To obtain the density of
\[
R^{C}_{n,u}(\cdot) = E_u[R_{n,u}(\cdot)|\hat{P}_n(\omega, \cdot) \in C],
\]
observe that \( \omega \in E_n \) if and only if \( \theta_i \omega_n \in E_n \) for any \( 0 \leq i \leq n - 1 \). Then
\[
R^{C}_{n,u}(\cdot) = E_u[R_{n,u}(\cdot)|\omega \in E_n]
\]
\[
= \frac{1}{n} \sum_{i=0}^{n-1} E_u[\chi_{\cdot}(\theta_i \omega_n)|\theta_i \omega_n \in E_n].
\]
It now follows from (4.5) and (4.6) that for any set \( A \in \mathcal{F}_n^0 \),
\[
R^{C}_{n,u}(\cdot) = \int \cdots \int \frac{1}{A P_u(E_n)} \chi_{E_n}(\omega_0, \ldots, \omega_{n-1}) \frac{1}{n} \sum_{i=0}^{n-1} \pi_i(\omega_0, \ldots, \omega_{n-1}) \, d\lambda^n,
\]
which proves the lemma. \( \square \)

Let \( \bar{P}_{n,u}^* \) be the measure on \( \mathcal{F}_n^0 \) defined by the transition probability function \( P^*(dy|x) \), \( P^* \) the \( I \)-projection of \( \pi \) onto \( C \) and initial distribution given by the marginal of \( R^{C}_{n,u} \). 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
\[ \frac{1}{n} \log P_u(\hat{P}_n(\omega, \cdot) \in C) \]
\[ \leq - \frac{1}{n} h_{\mathcal{F}_{n-1}}(R_{n,u}^C, \overline{P}_{n,u}^*) - \frac{(n - 1)}{n} I(C) + \frac{1}{n} \log \left( \frac{\sup uA}{a} \right), \]
where \( a \) and \( A \) are the bounds on \( \pi(y|x) \) given by assumption 3.

**Proof.** Let
\[ \hat{R}_{n,u}^C(\cdot) = E_u(R_{n,u}(\cdot)), \]
\[ \frac{1}{n} \sum_{i=0}^{n-1} E_u(\chi_i(\theta_i \omega_n)). \]
Then
\[ - \log P_u(\hat{P}_n(\omega, \cdot) \in C) = h_{\mathcal{F}_{n-1}}(R_{n,u}^C, \hat{R}_{n,u}^C) \]
(4.7)
\[ = \int_X \cdots \int_X dR_{n,u}^C \log \frac{R_{n,u}^C(\omega_0, \ldots, \omega_{n-1})}{\hat{R}_{n,u}^C(\omega_0, \ldots, \omega_{n-1})}, \]
where \( R_{n,u}^C(\omega_0, \ldots, \omega_{n-1}) \) and \( \hat{R}_{n,u}^C(\omega_0, \ldots, \omega_{n-1}) \) are the densities of \( R_{n,u}^C \) and \( \hat{R}_{n,u}^C \) respectively, with respect to \( \lambda^n \). Let \( \pi^l(\omega_0, \omega_1, \ldots, \omega_{n-1}) = \lambda(\omega_l|\omega_0)\pi(\omega_0|\omega_1)\cdots \omega(\omega_{n-1}|\omega_{n-2}) \). Let \( r_{n,u}(\omega_0) \) denote the density of the marginal of \( R_{n,u}^C \) with respect to \( \lambda \) and let \( \pi_{n,u} \) be the measure on \( \mathcal{F}_n^0 \) with density \( r_{n,u}(\omega_0)\pi^l(\omega_0, \ldots, \omega_{n-1}) \) with respect to \( \lambda^n \). Now for \( R_{n,u}^C \)-a.e. \( (\omega_0, \ldots, \omega_{n-1}), \)
\[ \frac{R_{n,u}^C(\omega_0, \ldots, \omega_{n-1})}{\hat{R}_{n,u}^C(\omega_0, \ldots, \omega_{n-1})} = \frac{R_{n,u}^C(\omega_0, \ldots, \omega_{n-1})}{\pi_{n,u}(\omega_0, \ldots, \omega_{n-1})} \]
\[ \times r_{n,u}(\omega_0) \frac{\pi^l(\omega_0, \ldots, \omega_{n-1})}{\hat{R}_{n,u}^C(\omega_0, \ldots, \omega_{n-1})}. \]

It is possible to take the log of both sides and integrate over \( R_{n,u}^C \) to obtain
\[ \int_X \cdots \int_X dR_{n,u}^C \log \frac{R_{n,u}^C}{\hat{R}_{n,u}^C} \]
(4.8)
\[ = \int_X \cdots \int_X dR_{n,u}^C \log \frac{R_{n,u}^C}{\pi_{n,u}} \]
\[ + \int_X \cdots \int_X dR_{n,u}^C \log r_{n,u}(\omega_0) + \int_X \cdots \int_X dR_{n,u}^C \log \frac{\pi^l}{\hat{R}_{n,u}^C}, \]
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 \( \pi \) and \( u \) that

\[
\pi_{i}(\omega_{0}, \ldots, \omega_{n-1}) \leq \frac{\sup uA}{a} \pi^{1}(\omega_{0}, \ldots, \omega_{n-1}),
\]

so that

\[
\int_{X} \cdots \int_{X} dR_{n,u}^{C} \log \frac{\pi^{1}}{\hat{R}_{n,u}^{C}} \geq -\log \left( \frac{\sup uA}{a} \right).
\]

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

\[
-\log P_{u}[\hat{P}_{n}(\omega, \cdot) \in C] \geq h_{\varphi_{n-1}}^{0}(R_{n,u}^{C}, \pi_{n,u}) - \log \left( \frac{\sup uA}{a} \right).
\]

Applying Lemma 4.1 and letting \( \hat{R}_{n,u}^{C} \) denote the stationary Markov process defined by \( R_{n,u}^{C} \), we have

\[
h_{\varphi_{n-1}}^{0}(R_{n,u}^{C}, \pi_{n,u}) = h_{\varphi_{n-1}}^{0}(R_{n,u}^{C}, \hat{R}_{n,u}^{C})
\]

\[
+ (n - 1) h^{1}(R_{n,u,1}^{C}1|\pi).
\]

Since \( R_{n,u}^{C} \in \Pi_{C} \) by Lemma 4.2, \( R_{n,u,1}^{C} \in M_{C} \) and it follows from Theorem 2.3 that

\[
h^{1}(R_{n,u,1}^{C}1|\pi) \geq h^{1}(R_{n,u,1}^{C0}P^{*}(\cdot|\cdot)) + I(C).
\]

From (4.10) we have

\[
-\log P_{u}[\hat{P}_{n}(\omega, \cdot) \in C] \geq h_{\varphi_{n-1}}^{0}(R_{n,u}^{C}, \hat{R}_{n,u}^{C})
\]

\[
+ (n - 1) h^{1}(R_{n,u,1}^{C0}P^{*}(\cdot|\cdot))
\]

\[
+ (n - 1) I(C) - \log \left( \frac{\sup uA}{a} \right).
\]

Applying Lemma 4.1 again gives

\[
h_{\varphi_{n-1}}^{0}(R_{n,u}^{C}, \bar{P}^{*}) = h_{\varphi_{n-1}}^{0}(R_{n,u}^{C}, \hat{R}_{n,u}^{C}) + (n - 1) h^{1}(R_{n,u,1}^{C0}P^{*}(\cdot|\cdot)).
\]

Substituting this into (4.11) yields

\[
-\log P_{u}[\hat{P}_{n}(\omega, \cdot) \in C] \geq h_{\varphi_{n-1}}^{0}(R_{n,u}^{C}, \bar{P}^{*}) + (n - 1) I(C) - \log \left( \frac{\sup uA}{a} \right).
\]

The lemma follows. \( \square \)

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

\[
\lim_{n \to \infty} \frac{1}{n} h_{\varphi_{n-1}}^{0}(R_{n,u}^{C}, \bar{P}^{*}) = 0,
\]

so that the measures \( R_{n,u}^{C} \) are asymptotically quasi-independent with respect to \( P^{*}(dy|x) \), the transition probability function of the I-projection of \( \pi \) onto \( C \).
PROOF. Using the uniformity of the estimate (1.5), it follows that
\[ \liminf_{n \to \infty} \frac{1}{n} \log P\{\hat{P}_n(\omega, \cdot) \in C^0\} \geq -I(C^0) = -I(C). \]
It follows from Lemma 4.4 that
\[
\limsup_{n \to \infty} \left\{ \frac{1}{n} \log P\{\hat{P}_n(\omega, \cdot) \in C\} + \frac{1}{n} h_{\mathcal{F}^{\mathcal{G}}_{n-\mathcal{G}}}(R^C_{n,u}, \bar{P}^*_n, u) \right\} \\
\leq -I(C).
\] (4.12)
In particular,
\[ \lim_{n \to \infty} \frac{1}{n} \log P\{\hat{P}_n(\omega, \cdot) \in C\} = -I(C). \]
It follows from (4.12) that
\[ \limsup_{n \to \infty} \frac{1}{n} h_{\mathcal{F}^{\mathcal{G}}_{n-\mathcal{G}}}(R^C_{n,u}, \bar{P}^*_n, u) \leq 0, \]
which establishes the theorem. \( \square \)

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 \( \pi \) onto \( C \). From Theorem 2.3, this function is defined for \( \lambda \)-a.e. \( x \). Extend it as described in Corollary 2.8. Let \( \hat{Q} \) be the measure on \( \Omega \) defined by
\[
\int_{\Omega} \delta_\omega \otimes_0 P^*_0 Q(d\omega).
\]
For \( Q \in M_S(\Omega) \), define \( h^j(Q|P^*(\cdot|\cdot)) = h_{\mathcal{F}^1}(Q, \hat{Q}) \).

COROLLARY 5.1. For \( h^j(\cdot|\cdot) \) defined as above,
\[ \lim_{n \to \infty} h^j(R^C_{n,u}|P^*(\cdot|\cdot)) = 0. \]

PROOF. From Theorem 4.5, we have that
\[ \lim_{n \to \infty} \frac{1}{n} h_{\mathcal{F}^{\mathcal{G}}_{n-\mathcal{G}}}(R^C_{n,u}, \bar{P}^*_n, u) = 0, \]
where \( \bar{P}^*_n \) is the measure on \( \mathcal{F}^{\mathcal{G}}_{n-1} \) defined by the transition probability function \( \bar{P}^*(dy|x) \) and initial distribution given by the marginal of \( R^C_{n,u} \). We can assume without loss of generality that for \( n > 1 \), \( h_{\mathcal{F}^{\mathcal{G}}_{n-\mathcal{G}}}(R^C_{n,u}, \bar{P}^*_n, u) < \infty \). It follows from the proof of Lemma A.4 in the Appendix [(A.5)] that
\[
\begin{equation}
\tag{5.2}
h_{\mathcal{F}^{\mathcal{G}}_{n-1}}(R^C_{n,u}, \bar{P}^*_n, u) = \sum_{i=1}^{n-1} h^j(R^C_{n,u}|P^*(\cdot|\cdot)).
\end{equation}
\]
From their definition, \( h^k(R_{n,u}^C | P^*(\cdot | \cdot)) \leq h^l(Q | P^*(\cdot | \cdot)) \) if \( k < l \). Then for \( j \leq n - 1 \),

\[
\frac{n - j}{n} h^j(R_{n,u}^C | P^*(\cdot | \cdot)) \leq \frac{1}{n} \sum_{i=1}^{n-1} h^i(R_{n,u}^C | P^*(\cdot | \cdot)) = \frac{1}{n} h_{\mathcal{F}_{n-1}^\delta}(R_{n,u}^C, \overline{P}_{n,u}^*),
\]

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

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 \( \{R_{n,u}^C\} \) contains a subsequence which converges weakly to the stationary Markov process \( P^* \), so that \( R_{n,u}^C \) 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 \( \mathcal{F}_{1,i}^\delta \), and suppose that for \( |t| \) sufficiently small, \( e^{i f(\omega_{i_1}, \ldots, \omega_{i_j})} \) is integrable with respect to \( (1/b(\omega_{i_1}) \cdots 1/b(\omega_{i_j}))^{\lambda^i} \), where the functions \( b(\cdot) \) are as in Corollary 2.8. Then \( \int f(\omega) dR_{n,u}^C \to \int f(\omega) dP^* \) as \( n \to \infty \).

**Proof.** The proof is similar to Csiszár (1975), Lemma 3.1. Using the stationarity of \( R_{n,u}^C \), we can assume without loss of generality that \( f(\omega) \) is measurable with respect to \( \mathcal{F}_1^\delta \). Using (5.2), it follows from Corollary 5.1 that

\[
\lim_{n \to \infty} h_{\mathcal{F}_{1}^\delta}(R_{n,u}^C, \overline{P}_{n,u}^*) = 0.
\]

Let \( f_{0,j,n} \) be the Radon-Nikodym derivative of \( R_{n,u}^C \) with respect to \( \overline{P}_{n,u}^* \) on \( \mathcal{F}_{j,0}^\delta \). Let \( Y = \bigcap_{i=0}^j X_i, \ X_i = X, \ i = 1, \ldots, j \). The Csiszár–Kemperman–Kullback inequality is that for two probability measures \( P \) and \( Q \) on a measure space \( (X, \mathcal{X}) \):

\[
|P - Q| \leq \sqrt{2h(P, Q)}
\]

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

\[
\lim_{n \to \infty} \left| R_{n,u}^C - \overline{P}_{n,u}^* \right|_{\mathcal{F}_{j,0}^\delta} = \lim_{n \to \infty} \int \left| f_{0,j,n} - 1 \right| d\overline{P}_{n,u}^* = 0,
\]

where \( \left| \cdot \right|_{\mathcal{F}_{j,0}^\delta} \) denotes the variation norm for measures on \( \mathcal{F}_{j,0}^\delta \). Let \( A_K = \{ \omega: f(\omega) \leq K \} \). Then

\[
\lim_{n \to \infty} \left| \int_{A_K} f(\omega) dR_{n,u}^C - \int_{A_K} f(\omega) d\overline{P}_{n,u}^* \right| = 0.
\]

However, for any \( g(\omega) \) measurable with respect to \( \mathcal{F}_{j,1}^\delta \) which is integrable
with respect to \(1/b(\omega_1) \cdots 1/b(\omega_j)\lambda^j(\omega_1, \ldots, \omega_j)\), it follows from Corollary 2.8 that
\[
\int_{\prod_{i=1}^j x_i} g(\omega_1, \ldots, \omega_j) P^*(\omega_1|\omega_0) P^*(\omega_2|\omega_1) \cdots P^*(\omega_j|\omega_{j-1}) \, d\lambda^j(\omega_1, \ldots, \omega_j)
\]
is a continuous function of \(\omega_0\). Since \(R_{n,u}^C\) converges weakly to \(P^*\), the marginals of \(R_{n,u}^C\) converge weakly to the marginals of \(P^*\) and it follows that \(\int g(\omega) \, dP_{n,u}^* \to \int g(\omega) \, dP^*\) as \(n \to \infty\). Then (5.3) implies
\[
\lim_{n \to \infty} \int_{A_K} f(\omega) \, dR_{n,u}^C = \int_{A_K} f(\omega) \, dP^*.
\]
To complete the proof, it suffices to show that for any \(\varepsilon > 0\), there exists \(K\) such that
\[
(5.4)(i) \quad \limsup_{n \to \infty} \int_{Y/A_K} |f| \, dR_{n,u}^C = \limsup_{n \to \infty} \int_{Y/A_k} |f| f_{0,j,n} \, dP_{n,u}^* < \varepsilon
\]
and
\[
(5.4)(ii) \quad \int_{Y/A_K} |f| \, dP^* < \varepsilon.
\]
To obtain this, we prove that
\[
(5.5) \quad \lim_{n \to \infty} h_{\mathcal{F}_j}(R_{n,u}^C, \overline{P}_{n,u}^*) = \lim_{n \to \infty} \int_Y f_{0,j,n} \log f_{0,j,n} \, d\overline{P}_{n,u}^* = 0
\]
implies that for any \(A \in \mathcal{F}_j^{-1}\),
\[
(5.6) \quad \lim_{n \to \infty} \int_A f_{0,j,n} \log f_{0,j,n} \, d\overline{P}_{n,u}^* = 0.
\]
Now
\[
\int_A f_{0,j,n} \log f_{0,j,n} \, d\overline{P}_{n,u}^* \geq R_{n,u}^C(A) \log \frac{R_{n,u}^C(A)}{\overline{P}_{n,u}^*(A)},
\]
so that
\[
(5.7) \quad \liminf_{n \to \infty} \int_A f_{0,j,n} \log f_{0,j,n} \, d\overline{P}_{n,u}^* \geq P^*(A) \log \frac{P^*(A)}{P^*(A)} = 0.
\]
Similarly,
\[
(5.8) \quad \liminf_{n \to \infty} \int_{Y/A} f_{0,j,n} \log f_{0,j,n} \, d\overline{P}_{n,u}^* \geq 0.
\]
In view of (5.5), (5.7) and (5.8), (5.6) follows.

Now pick \(t > 0\) and \(K\) so that \(\int_{Y/A_K} |f| \, dP^* < \varepsilon t\) 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
\[ \int_{Y/A_K} t |f| f_{0,j,n} d\tilde{P}_n^* \leq \int_{Y/A_K} f_{0,j,n} \log f_{0,j,n} d\tilde{P}_n^* + \int_{Y/A_K} e^{tf} d\tilde{P}_n^*. \]

It follows that
\[ \limsup_{n \to \infty} \int_{Y/A_K} |f| f_{0,j,n} d\tilde{P}_n^* \leq \frac{1}{t} \lim_{n \to \infty} \int_{Y/A_K} e^{tf} d\tilde{P}_n^* \]
\[ = \frac{1}{t} \int_{Y/A_K} e^{tf} dP^* \]
\[ < \varepsilon, \]

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 \( P_n \)-distribution of \( X_0, \ldots, X_{n-1} \) 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 \( \mathcal{F}^0_{n-1} \) measurable set \{ \omega: \hat{P}_n(\omega, \cdot) \in C \}. \) Then the conditional \( P_n \)-distribution of \( X_0, \ldots, X_{n-1} \) under the condition \( \hat{P}_n(\omega, \cdot) \in C \) has the density
\[ P_{n,u}(\omega_0, \ldots, \omega_n-1) \]
\[ = \frac{1}{P_u(E_n)} \chi_{E_n}(\omega_0, \ldots, \omega_n-1)u(\omega_0)\pi(\omega_1|\omega_0) \cdots \pi(\omega_{n-1}|\omega_{n-2}). \]

Let \( R_{n,u}^C(\omega_0, \ldots, \omega_n-1) \) be the density of \( R_{n,u}^C \) on \( \mathcal{F}^0_{n-1} \). Then from (4.4), we have
\[ (5.9) \begin{align*}
(5.9)(i) \quad & P_{n,u}(\omega_0, \ldots, \omega_n-1) \geq \inf uA \frac{R_{n,u}^C(\omega_0, \ldots, \omega_n-1).}
(5.9)(ii) \quad & P_{n,u}(\omega_0, \ldots, \omega_n-1) \leq \sup uA \frac{R_{n,u}^C(\omega_0, \ldots, \omega_n-1)\lambda^n.}
\end{align*} \]

If \( p_{n,u}(\omega_0) \) is the density with respect to \( \lambda \) of the first marginal of \( P_1(\cdot| \hat{P}_n(\omega, \cdot) \in C) \) on \( \mathcal{F}^0_{n-1} \), then the same bounds must hold with respect to the density \( r_{n,u}(\omega_0) \) of the marginals of \( R_{n,u}^C \) with respect to \( \lambda \). Let \( \hat{P}_n \) be the probability measure on \( \mathcal{F}^0_{n-1} \) 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 \( \lambda^n \). Let \( \hat{P}_{n,u}(\omega_0, \ldots, \omega_n-1) \) denote this density. Similarly, let \( \hat{P}_{n,u}(\omega_0, \ldots, \omega_n-1) \) denote the density of \( \hat{P}_{n,u} \) with respect to \( \lambda^n \).
Then from (5.9)(i) and (ii),
\[
\int P_{n,u} \log \frac{P_{n,u}}{\bar{P}_{n,u}} \, d\lambda^n \geq \left( \inf \frac{ua}{\sup uA} \right) \int R_{n,u}^C \frac{R_{n,u}}{\bar{P}_{n,u}} \, d\lambda^n \\
+ 2 \left( \inf \frac{ua}{\sup uA} \right) \log \left( \frac{\inf \frac{ua}{\sup uA}}{\sup uA} \right)
\]
and
\[
\int P_{n,u} \log \frac{P_{n,u}}{\bar{P}_{n,u}} \, d\lambda^n \leq \left( \sup \frac{uA}{\inf uA} \right) \int R_{n,u}^C \frac{R_{n,u}}{\bar{P}_{n,u}} \, d\lambda^n \\
+ 2 \left( \frac{\sup uA}{\inf uA} \right) \log \left( \frac{\sup uA}{\inf uA} \right).
\]
It follows from Theorem 4.5 that
\[
\lim_{n \to \infty} \frac{1}{n} \int P_{n,u} \log \frac{P_{n,u}}{\bar{P}_{n,u}} \, d\lambda^n = 0,
\]
which establishes that the sequence of measures $P_u(\cdot | \bar{P}_n(\omega, \cdot) \in C)$ on $\mathcal{F}_{n-1}^0$ 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 $A \subset M_2(\Omega)$,

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

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 $(X, \Sigma)$ be a Polish space and $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \Sigma$ be sub-$\sigma$-fields. Let $\mu$ and $\lambda$ be two measures on $(X, \Sigma)$ and suppose $\mu \ll \lambda$ on the $\sigma$-field $\mathcal{F}_1$. Let $\mu_\omega' = \int X \lambda_\omega \mu(\omega) d\omega$, where $\lambda_\omega$ is the conditional probability distribution of $\lambda$ given $\mathcal{F}_1$. Then

\[
h_{\mathcal{F}_2}(\mu, \lambda) = h_{\mathcal{F}_1}(\mu, \lambda) + h_{\mathcal{F}_2}(\mu, \mu').
\]
PROOF. For $E \in \Sigma$, 

$$
\mu'(E) = \int_X \lambda_\omega(E) \mu(d\omega) \\
= \int_X \lambda_\omega(E) \frac{d\mu}{d\lambda} \bigg| \mathcal{F}_1 \lambda(d\omega) \\
= \int_X E^{\lambda_\omega} \left( \lambda_\omega \frac{d\mu}{d\lambda} \bigg| \mathcal{F}_1 \right) \lambda(d\omega) \\
= \int \frac{d\mu}{d\lambda} \bigg| \mathcal{F}_1 \lambda(d\omega),
$$

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

$$
\mu(E) = \int \frac{d\mu}{d\lambda} \bigg| \mathcal{F}_2 \lambda(d\omega) \\
= \int \left( \frac{d\mu}{d\lambda} \bigg| \mathcal{F}_2 \right) \left( \frac{d\mu}{d\lambda} \bigg| \mathcal{F}_1 \right) \frac{d\mu}{d\lambda} \bigg| \mathcal{F}_1 \lambda(d\omega) \\
= \int \left( \frac{d\mu}{d\lambda} \bigg| \mathcal{F}_2 \right) \left( \frac{d\mu}{d\lambda} \bigg| \mathcal{F}_1 \right) \mu'(d\omega).
$$

It follows that 

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

Taking the logarithm of both sides and integrating with respect to $\mu$ completes the argument. \(\square\)

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

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

$$
h_{\mathcal{F}_n}(Q, \hat{Q}) = nH(Q|\pi).
$$

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

$$
h_{\mathcal{F}_{n+1}}(Q, \hat{Q}) = h_{\mathcal{F}_n}(Q, \hat{Q}) + h_{\mathcal{F}_n}(Q, Q'),
$$

where $Q' = \int \hat{Q}_\omega Q(d\omega)$, $\hat{Q}_\omega$ the conditional probability distribution of $\hat{Q}$ given $\mathcal{F}_{n-\infty}$. But $\hat{Q}_\omega = \delta_\omega \otimes P_{\omega(j)}$, so that, using the stationary of $Q$, $h_{\mathcal{F}_{n+1}}(Q, Q') = H(Q|\pi)$. \(\square\)
Lema A.4.

\begin{equation}
\lim_{n \to \infty} \frac{1}{n} h_{\mathcal{F}_{n-1}^0}(Q, \hat{Q}) = H(Q|\pi).
\end{equation}

**Proof.** By Lemma A.3, either $h_{\mathcal{F}_{n}^{-}}(Q, \hat{Q}) = +\infty$ for all $n$ or

\begin{equation}
\frac{1}{n} h_{\mathcal{F}_{n-1}^0}(Q, \hat{Q}) \leq \frac{1}{n} h_{\mathcal{F}_{n-1}^{-}}(Q, \hat{Q}) = \frac{n-1}{n} H(Q|\pi).
\end{equation}

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

\begin{equation}
h_{\mathcal{F}_{j+1}^0}(Q, \hat{Q}) - h_{\mathcal{F}_{j}^0}(Q, \hat{Q}) = h_{\mathcal{F}_{j+1}^0}(Q, Q'),
\end{equation}

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

\begin{equation}
h_{\mathcal{F}_{j+1}^0}(Q, Q') = h_{\mathcal{F}_{j+1}^c}(Q, \hat{Q}) = h^j(Q|\pi).
\end{equation}

From (A.4), it follows that

\begin{equation}
\frac{1}{n} h_{\mathcal{F}_{n-1}^0}(Q, \hat{Q}) = \frac{1}{n} \sum_{j=1}^{n-1} h^j(Q|\pi).
\end{equation}

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 $h^j(Q|\pi) \leq M$. It follows from Moy (1961), Lemma 3, that

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

concluding the proof of the lemma. $\square$

**Lemma A.5.** Let $\Lambda_j$ denote the set of continuous functions $\phi$ on $\Omega$ depending only on the coordinates $\omega_i$, $0 \leq i \leq j$, which satisfy $E^{P_\omega}(e^\phi) \leq 1$ for all $x \in X$. Assume the transition probability function $\pi(dy|x)$ satisfies assumption 3. Then

\begin{equation}
h_{\mathcal{F}_{j}^0}(Q, \hat{Q}) = \sup_{\phi \in \Lambda_j} E^{Q}(\phi).
\end{equation}

**Proof.** By (2.1),

\begin{equation}
h_{\mathcal{F}_{j}^0}(Q, \hat{Q}) = \sup_{u \in \mathcal{U}} \left[ \int \log u(\omega) Q(d\omega) - \log \int_{\Omega} E^{P_\omega}(u) Q(d\omega) \right],
\end{equation}
where \( \mathcal{W} \) consists of the positive, continuous functions depending only on the coordinates \( \omega_i, 0 \leq i \leq j \). Writing \( \log u(\omega) = \phi(\omega) \) for \( u(\omega) \in \mathcal{W} \) shows

\[
h_{\mathcal{F}^0_j}(Q, \hat{Q}) \geq \sup_{\phi \in \Lambda_j} E^Q(\phi).
\]

Let \( \Phi \) denote the set of continuous function depending on the coordinates \( \omega_i, 0 \leq i \leq j \). 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

\[
E^{P_x}(e^\phi) = e^{P_x}(e^{\phi(\omega) - \bar{\psi}(\omega(0))})
\]

\[
= e^{-\bar{\psi}(x)} E^{P_x}(e^\phi) = 1,
\]

so \( \phi \in \Lambda_j \). Then

\[
h_{\mathcal{F}^0_j}(Q, \hat{Q}) = \sup_{\phi \in \Phi} \left[ \int_\Omega \psi(\omega) Q(d\omega) - \log \int_\Omega E^{P_x}(e^\phi) Q(d\omega) \right]
\]

\[
\leq \sup_{\phi \in \Phi} \left[ \int_\Omega \psi(\omega) Q(d\omega) - \int_\Omega \log E^{P_x}(e^\phi) Q(d\omega) \right]
\]

by Jensen's inequality. The right-hand side

\[
= \sup_{\phi \in \Phi} \left[ \int_\Omega (\psi(\omega) - \bar{\psi}(\omega(0))) Q(d\omega) \right]
\]

\[
\leq \sup_{\phi \in \Lambda_j} E^Q(\phi). \quad \square
\]

**Lemma A.6.** Suppose \( \phi \) is \( \mathcal{F}^0_{N-1} \) measurable and \( E^{P_x}(e^\phi) \leq 1 \) for all \( x \in X \). Then

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

for all \( n \).

**Proof.** For \( j = 0, 1, \ldots, N - 1 \), define

\[
\psi_j(\omega) = \sum_{k: k \geq 0 \atop j + kN \leq n} \phi(\theta_{j+kN} \omega).
\]

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

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

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

Under the hypothesis on \( \phi \), \( E^P_x(\exp \psi(\omega)) \leq 1 \).

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 \( \phi \) is a bounded \( \mathcal{F}^0_{N-1} \)-measurable function satisfying \( E^P_x(e^{\phi}) \leq 1 \) for all \( x \). Then

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

**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 \( \omega_n \) is defined as in Section 1. Then

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

which, in view of Lemma A.6, establishes the corollary. \( \square \)

Let

\[ J(A) = \lim_{n \to \infty} \sup \frac{1}{n} \log \sup_{x \in X} \Gamma_{n,x}(A). \]

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

\[ J(A) \leq - \sup_{l: A_1, A_2, \ldots, A_l} \inf_{1 \leq j \leq l} \sup_{N > 0} \inf_{\phi \in \Lambda_{N-1}} \frac{1}{N} \int_{\Omega} \phi(\omega)Q(d\omega). \]
**Proof.** From Corollary A.7, for any \( A \in M_S(\Omega) \) and any \( \phi \in \Lambda_{N-1} \),

\[
\Gamma_{n,x}(A) \leq \exp \left( 2 \sup_{\omega \in \Omega} \phi(\omega) \right) \exp \left( -\frac{n}{N} \inf_{Q \in A} \int_{\Omega} \phi(\omega) Q(d\omega) \right).
\]

Then

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

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

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

\[
\sup_{A \subseteq \bigcup_{j=1}^{l} A_j} \inf_{1 \leq j \leq l} \inf_{N>0} \sup_{\phi \in \Lambda_{N-1}} \inf_{Q \in A_j} \frac{1}{N} \int_{\Omega} \phi(\omega) Q(d\omega) \geq \inf_{Q \in A} \mathcal{H}(Q|\pi).
\]

**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

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

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

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

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

\[
\sup_{1 \leq j \leq l} \sup_{N>0} \sup_{\phi \in \Lambda_{N-1}} \inf_{Q \in G_j} \frac{1}{N} \int_{\Omega} \phi(\omega) Q(d\omega) \geq \inf_{Q \in A} \mathcal{H}(Q|\pi) - \varepsilon.
\]

The statement of the lemma follows. □

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

**REFERENCES**


Mathematics Department
University of Massachusetts
Lowell, Massachusetts 01854