QUASISTATIONARITY AND THE RELATIVE ENTROPY OF PROBABILITY MEASURES

L.A. BREYER

Abstract. Consider a Markov process X with finite lifetime. In this paper, we derive sufficient conditions for the convergence of the law of X, conditioned on longer and longer lifetimes, using information theoretic techniques.

1. Introduction

Consider a Markov process X_t with state space E, and suppose that with probability 1, this process exits E in a finite time. Denote the first exit time from E by ζ, and assume that we have a.s. $\zeta > 0$. In this paper, we shall derive sufficient conditions for the sequence of processes $X^n = (X_s : s \leq t \,|\, \zeta > n)$ to converge to some limiting process $Z = (Z_s : s \leq t)$ on E, by using entirely information theoretic arguments.

The convergence problem we discuss enters into the study of *quasistationarity* (see for example Pollett, 1995), where it is known or assumed that ζ is large. The process Z_t then gives a better approximation to the observed system described by X_t.

There is currently much interest in rare event simulation for engineering or insurance systems. The straightforward Monte Carlo method consists in simulating the sample paths of the process of interest, and analysing the (very small) subset of those paths which obey the rare event constraint. Often, this can be phrased in terms of a Markov process X_t and a real valued function f, and one is interested in simulating the paths of X_t with the constraint $f(X_t) \geq 0$, or equivalently, $X_t \in E := f^{-1}(0)$. A superior approach uses Monte Carlo importance sampling, wherein one simulates first the paths of the process X_t, *conditioned on* $X_t \in E$,

Date: April, 1999.

1991 *Mathematics Subject Classification.* Primary 60J, Secondary 60F.

Key words and phrases. Markov chain , Quasistationarity, Kullback-Leibler divergence.

and then weighs the results according to the appropriate likelyhood ratio. We refer the interested reader to the papers collected in RESIM'99.

Finally, note that we could frame the problem treated here in terms of the empirical occupation measure of X_t. More specifically, if we set $L_t(X, dy) = \frac{1}{t}\int_0^t \delta_{X_s}(dy)ds$, then

$$\{\zeta > t\} = \{L_t(X, E) = 1\} = \{L_t(X, \cdot) \in C_E\},$$

where C_E is the set of probability measures on E. In this way, the conditioning problem we consider is related to one considered in Statistical Physics and Large Deviation Theory (see the Gibbs Conditioning Principle, Dembo and Zeitouni, 1998, or the excellent review by Bolthausen, 1993). However, for these results to apply here, our Markov process X_t must be identified as the restriction to E of another process Y_t whose state space includes E as a proper subset. Moreover, this artificial process Y_t is required to be uniformly ergodic, which is a very strong assumption. By contrast, we make no such assumptions on X_t here, and our sufficient conditions are simply integrability assumptions for the limiting Radon-Nikodym densities.

2. Convergence in Information

In the following sections, we shall prove convergence results in information. Here we collect useful facts about this type of convergence for the convenience of the reader. For proofs of the assertions, we refer the reader to (Csiszar, 1975).

Recall that if $\mathbb{Q}$ and $\mathbb{P}$ are two probability measures on a measurable space $(\Omega, \mathcal{F})$, the Kullback-Leibler or information divergence $D(\mathbb{Q} \parallel \mathbb{P})$ is defined by

$$D(\mathbb{Q} \parallel \mathbb{P}) = \begin{cases} \int \log(d\mathbb{Q}/d\mathbb{P})d\mathbb{Q} & \text{if } \mathbb{Q} \ll \mathbb{P}, \\ +\infty & \text{otherwise.} \end{cases} \tag{1}$$

In line with this definition, we say that a sequence of probability measures $(\mathbb{Q}_n)$ converges in information to a probability $\mathbb{Q}$ if $D(\mathbb{Q} \parallel \mathbb{P}) \to 0$. In particular, this requires $D(\mathbb{Q} \parallel \mathbb{P}) < \infty$ for all n sufficiently large.

Convergence in information is stronger than total variation convergence. If $\mathbb{Q}_n$ converges to $\mathbb{P}$ in information, then

$$\int F d\mathbb{Q}_n \to \int F d\mathbb{P} \text{ for all } F \text{ such that } \int e^{\epsilon |F|} d\mathbb{P} < \infty, \text{ for some } \epsilon > 0.$$

A converse is true as well. If $(\mathbb{Q}_n)$ are probability measures such that

$$\sup_n D\big(\mathbb{Q}_n \parallel \mathbb{P}\big) < +\infty, \tag{2}$$

then since $D\big(\mathbb{Q}_n \parallel \mathbb{P}\big) = \int F\big(d\mathbb{Q}_n/d\mathbb{P}\big)d\mathbb{P}$ where $F(x) = x\log x$ is convex and such that $F(x)/x \to \infty$ as $x \to \infty$, the Radon-Nilkdym derivatives $d\mathbb{Q}_n/d\mathbb{P}$ are uniformly integrable. Consequently, some subsequence $\mathbb{Q}_{n(k)}$ converges in $L^1(d\mathbb{P})$, and this implies that $\mathbb{Q}_{n(k)}$ converges in total variation to some probability measure $\mathbb{P}^* \ll \mathbb{P}$. If in fact $D\big(\mathbb{Q}_{n(k)} \parallel \mathbb{Q}\big) \to 0$ for some $\mathbb{Q}$, then $\mathbb{P}_* \equiv \mathbb{Q}$.

It is important to realize that the above results depend crucially on the chosen σ-algebra $\mathcal{F}$. Since $\mathbb{Q} \ll \mathbb{P}$ if and only if, for all $A \in \mathcal{F}$, $\mathbb{P}(A) = 0$ implies $\mathbb{Q}(A) = 0$, we can have $D\big(\mathbb{Q} \parallel \mathbb{P}\big) = \infty$ on a probability space $(\Omega, \mathcal{F})$, and have $D\big(\mathbb{Q} \parallel \mathbb{P}\big) < \infty$ on $(\Omega, \mathcal{G})$ where $\mathcal{G}$ is a sub-σ-algebra of $\mathcal{F}$.

For $\mathbb{Q}$ and $\mathbb{P}$ probability measures on a space $(\Omega, \mathcal{F})$, we shall write $D_{\mathcal{G}}\big(\mathbb{Q} \parallel \mathbb{P}\big)$ to mean $D\big(\mathbb{Q}|_{\mathcal{G}} \parallel \mathbb{P}|_{\mathcal{G}}\big)$, i.e. computed in $(\Omega, \mathcal{G})$. We then have the simple rule

$$D_{\mathcal{G}_1}\big(\mathbb{Q} \parallel \mathbb{P}\big) \le D_{\mathcal{G}_2}\big(\mathbb{Q} \parallel \mathbb{P}\big) \quad \text{if } \mathcal{G}_1 \subseteq \mathcal{G}_2 \subseteq \mathcal{F}. \tag{3}$$

This follows immediately from the familiar alternative definition to (1):

$$D_{\mathcal{G}}\big(\mathbb{Q} \parallel \mathbb{P}\big) = \sup\Big\{\sum_{i=1}^{k} \mathbb{Q}(B_i)\log\Big(\mathbb{Q}(B_i)/\mathbb{P}(B_i)\Big) : \\ (B_1, \dots, B_k) \text{ a } \mathcal{G} \text{ measurable partition of } \Omega\Big\}.$$

In the next section, we shall work with a stochastic process $(X_t : t \ge 0)$ with law $\mathbb{P}$ which is defined on a filtered probability space $(\Omega, \mathcal{F}, \mathcal{F}_t : t \ge 0)$. Here $\mathcal{F}_s \subset \mathcal{F}_t$ if $s < t$. We shall define probabilities $\mathbb{Q}_n$ and show that $\sup_n D_{\mathcal{F}_t}\big(\mathbb{Q}_n \parallel \mathbb{P}\big) < \infty$ for all finite t, and thereby that the laws converge on $(\Omega, \mathcal{F}_t)$.

We define one further quantity. Denote by $\mathcal{P}(\Omega, \mathcal{F})$ the class of probability measures on $(\Omega, \mathcal{F})$. For a given probability $\mathbb{P} \in \mathcal{P}(\Omega, \mathcal{F})$, the I-projection of $\mathbb{P}$ on a convex subset $\mathcal{C} \subseteq \mathcal{P}(\Omega, \mathcal{F})$ is that probability $\mathbb{P}^* \in \mathcal{P}(\Omega, \mathcal{F})$ which satisfies

$$D\big(\mathbb{P}^* \parallel \mathbb{P}\big) = \inf\big\{D\big(\mathbb{Q} \parallel \mathbb{P}\big) : \mathbb{Q} \in \mathcal{C}\big\} < \infty,$$

where by convention $\inf \emptyset = \infty$ (thus if $\mathcal{C} \cap \{D(\cdot \parallel \mathbb{P}) < \infty\} = \emptyset$, the I-projection does not exist). If the set $\mathcal{C}$ is closed in the topology of total variation, then in fact $\mathbb{Q}^* \in \mathcal{C}$.

A fundamental example of the above occurs when $\mathcal{C}$ is defined by linear constraints, i.e. there exist functions f_α and constants c_α such that

$$\mathcal{C} = \Big\{\mathbb{Q} \in \mathcal{P}(\Omega, \mathcal{F}) : \int f_\alpha d\mathbb{Q} = c_\alpha \text{ for all } \alpha\Big\}.$$

If the I-projection of $\mathbb{P}$ on $\mathcal{C}$ exists, it must be of the form (Csiszar, 1975, Theorem 3.1)

$$d\mathbb{P}^*(\omega) = Z^{-1} 1_{N^c}(\omega) \exp U(\omega) d\mathbb{P}(\omega), \tag{4}$$

where Z is a constant, the function U belongs to the closed linear subspace of $L^1(d\mathbb{P})$ spanned by the f_α, and N is a set such that $\mathbb{Q}(N) = 0$ whenever $\mathbb{Q} \in \mathcal{C}$ and $D\big(\mathbb{Q} \parallel \mathbb{P}\big) < \infty$.

For a more concrete example, let $A \subseteq \Omega$ be $\mathcal{F}$-measurable and satisfy $\mathbb{P}(A) > 0$. The set

$$\mathcal{C}(A) = \big\{\mathbb{Q} \in \mathcal{P}(\Omega, \mathcal{F}) : \mathbb{Q}(A) = 1\big\}$$

is convex, total variation closed and contains at least one measure, namely $\mathbb{P}(\cdot \,|\, A) = \mathbb{P}(\cdot \cap A)/\mathbb{P}(A)$.

Since $D\Big(\mathbb{P}(\cdot \,|\, A) \parallel \mathbb{P}\Big) = -\log \mathbb{P}(A) < \infty$, the I-projection exists and from (4) is of the form $d\mathbb{P}^* = (a 1_A + b 1_{A^c}) d\mathbb{P}$. The constraint $\mathbb{P}^* \in \mathcal{C}(A)$ gives $a = 1/\mathbb{P}(A)$, while requiring $\mathbb{P}^*(\Omega) = 1$ implies $b = 0$. Thus $\mathbb{P}^* = \mathbb{P}(\cdot \,|\, A)$, that is, the conditional probability minimizes the Kullback-Leibler divergence among all probabilities $\mathbb{Q}$ such that $\mathbb{Q}(A) = 1$.

If we have $\mathbb{P}(A) = 0$, the probability $\mathbb{P}(\cdot \,|\, A)$ can no longer be defined, but we can attempt a limiting argument.

Lemma 1. *Let $A \subseteq \Omega$ be $\mathcal{F}$ measurable and such that $\mathbb{P}(A) = 0$. Suppose there exist $\mathcal{F}$-measurable sets $A_n \subseteq \Omega$ decreasing monotonically to A, and also satisfying $\mathbb{P}(A_n) > 0$ for each n. If $\mathcal{G} \subseteq \mathcal{F}$ is a σ-algebra, define*

$$\mathcal{C}(A, \mathcal{G}) = \{\mathbb{Q} \in \mathcal{P}(\Omega, \mathcal{F}) : \mathbb{Q} \ll \mathbb{P} \text{ on } \mathcal{G} \text{ and } \mathbb{Q}(A) = 1\},$$

and suppose that $D_{\mathcal{G}}\big(\mathbb{Q} \parallel \mathbb{P}\big) < \infty$ for some $\mathbb{Q} \in \mathcal{C}(A, \mathcal{G})$. Then there exists a probability $\mathbb{P}^* \in \mathcal{P}(\Omega, \mathcal{F})$ such that $\mathbb{P}(\cdot \,|\, A_n)$ converges to $\mathbb{P}^*$ in information on the space $(\Omega, \mathcal{G})$, and in fact on $(\Omega, \mathcal{G}')$ for any sub-σ-algebra $\mathcal{G}'$ of $\mathcal{G}$. .

Proof. In $\mathcal{P}(\Omega,\mathcal{G})$, define the set $\mathcal{C}(A_n,\mathcal{G})$ analogously to $\mathcal{C}(A,\mathcal{G})$ above, and note that this is convex and variation closed, both as a subset of $\mathcal{P}(\Omega,\mathcal{F})$ and of $\mathcal{P}(\Omega,\mathcal{G})$. When viewed as a subset of $\mathcal{P}(\Omega,\mathcal{G})$, it can also be written

$$\mathcal{C}(A_n,\mathcal{G}) = \Big\{\mathbb{Q}\in\mathcal{P}(\Omega,\mathcal{G}) : \int \mathbb{P}(A_n\,|\,\mathcal{G})d\mathbb{Q} = 1\Big\}.$$

Indeed, this is because $\mathbb{P}(A_n\,|\,\mathcal{G}) = \mathbb{Q}(A_n\,|\,\mathcal{G})$ a.s. $\mathbb{Q}$ for all $\mathbb{Q}\in\mathcal{C}(A_n,\mathcal{G})$ (original definition). To see this, take G bounded and $\mathcal{G}$ measurable but arbitrary, and observe

$$\begin{aligned}\int G\mathbb{P}(A\,|\,\mathcal{G})d\mathbb{Q} &= \int G\Big(d\mathbb{Q}|_{\mathcal{G}}/d\mathbb{P}|_{\mathcal{G}}\Big)\mathbb{P}(A\,|\,\mathcal{G})d\mathbb{P}\\ &= \int \mathbb{E}^{\mathbb{P}}\Big(1_{A_n}G\Big(d\mathbb{Q}|_{\mathcal{G}}/d\mathbb{P}|_{\mathcal{G}}\Big)\,|\,\mathcal{G}\Big)d\mathbb{P}\\ &= \int 1_{A_n}G\Big(d\mathbb{Q}|_{\mathcal{G}}/d\mathbb{P}|_{\mathcal{G}}\Big)d\mathbb{P}\\ &= \int G1_{A_n}d\mathbb{Q}\\ &= \int G\mathbb{Q}(A_n\,|\,\mathcal{G})d\mathbb{Q},\end{aligned}$$

Next, we wish to compute in $\mathcal{P}(\Omega,\mathcal{G})$ the I-projection $\mathbb{P}^*$ of $\mathbb{P}$ on $\mathcal{C}(A_n,\mathcal{G})$. From the characterization (4), we have a set $N\in\mathcal{G}$ and constants a and Z such that

$$\begin{aligned}d\mathbb{P}^* &= Z^{-1}\exp\big\{a\mathbb{P}(A_n\,|\,\mathcal{G})\big\}1_{N^c}d\mathbb{P}\\ &= Z^{-1}\Big\{\sum_{k=0}^{\infty}\frac{a^k}{k!}\mathbb{P}(A_n\,|\,\mathcal{G})^k\Big\}1_{N^c}d\mathbb{P}\\ &= Z^{-1}\mathbb{P}(A_n\,|\,\mathcal{G})\Big\{\sum_{k=0}^{\infty}\frac{a^k}{k!}\Big\}1_{N^c}d\mathbb{P},\end{aligned}$$

where we used the relation

$$\begin{aligned}\mathbb{P}(A_n|\mathcal{G})^2 &= \mathbb{E}\Big(1_{A_n}\mathbb{E}\big(1_{A_n}|\mathcal{G}\big)|\mathcal{G}\Big)\\ &= \mathbb{E}\Big(\mathbb{E}\big(1_{A_n}|\mathcal{G}\big)|\mathcal{G}\Big)\\ &= \mathbb{P}(A_n|\mathcal{G}).\end{aligned}$$

The set N must satisfy $\mathbb{Q}(N)=0$ for all $\mathbb{Q}\in\mathcal{C}(A_n,\mathcal{G})$ which have, in $\mathbb{P}(\Omega,\mathcal{G})$, a finite Kullback Leibler distance to $\mathbb{P}$. In particular, this must hold for the measure $d\mathbb{Q} = \frac{\mathbb{P}(A_n|\mathcal{G})}{\mathbb{P}(A_n)}d\mathbb{P}$, restricted to $(\Omega,\mathcal{G})$. In that case, $\mathbb{Q}(N)=0$ implies that

$\mathbb{P}(A_n|\mathcal{G}) = 0$ a.s. on N. Hence we may simplify the formula for $\mathbb{P}^*$ above, arriving at $d\mathbb{P}^*_{\mathcal{G}} = \widetilde{Z}^{-1}\mathbb{P}(A_n \,|\, \mathcal{G})d\mathbb{P}$, and the requirement $\mathbb{P}^*(\Omega) = 1$ then gives $\widetilde{Z} = \mathbb{P}(A_n)$.

Summarizing, the I projection of $\mathbb{P}$ on $\mathcal{C}(A_n, \mathcal{G})$ computed in $\mathcal{P}(\Omega, \mathcal{G})$ is simply the I-projection $\mathbb{P}^*_n = \mathbb{P}(\cdot \,|\, A_n)$ of $\mathbb{P}$ on $\mathcal{C}(A_n)$ computed in $\mathcal{P}(\Omega, \mathcal{F})$,and conditioned to $\mathcal{G}$. Consequently, we have

$$\begin{aligned} D_{\mathcal{G}}\big(\mathbb{P}^*_n \parallel \mathbb{P}\big) &= \inf\{D_{\mathcal{G}}\big(\mathbb{Q} \parallel \mathbb{P}\big) : \mathbb{Q} \in \mathcal{C}(A_n, \mathcal{G})\} \\ &\leq \inf\{D_{\mathcal{G}}\big(\mathbb{Q} \parallel \mathbb{P}\big) : \mathbb{Q} \in \mathcal{C}(A_{n+1}, \mathcal{G})\} \\ &= D_{\mathcal{G}}\big(\mathbb{P}^*_{n+1} \parallel \mathbb{P}\big) \\ &\leq \cdots \leq \inf_{\mathbb{Q}\in\mathcal{C}(A,\mathcal{G})} D_{\mathcal{G}}\big(\mathbb{Q} \parallel \mathbb{P}\big) < \infty, \end{aligned}$$

where the inequalities are due to the fact that

$$\mathcal{C}(A_n, \mathcal{G}) \supseteq \mathcal{C}(A_{n+1}, \mathcal{G}) \supseteq \cdots \supseteq \mathcal{C}(A, \mathcal{G}) = \cap_n \mathcal{C}(A_n, \mathcal{G}).$$

Having bounded $D_{\mathcal{G}}\big(\mathbb{P}^*_n \parallel \mathbb{P}\big)$ uniformly in n, it follows that the Radon-Nikodym derivatives φ^*_n of $\mathbb{P}^*_n$ with respect to $\mathbb{P}$, computed in $(\Omega, \mathcal{G})$, are uniformly integrable, hence weakly sequentially compact in $L^1(\Omega, \mathcal{G})$. Let φ be any weak limit point. Taking a subsequence if necessary, we may assume that $\varphi_n \to \varphi$ a.s. $\mathbb{P}$ and we shall show that all these φ are equivalent. Indeed, defining $\mathbb{P}^* \ll \mathbb{P}$ on $\mathcal{G}$ by $d\mathbb{P}^*/d\mathbb{P} = \varphi$, it is clear that

$$\mathbb{P}^*(A_n) = \lim_{k\to\infty} \mathbb{P}^*_k(A_n) = 1, \quad n = 1, 2, 3, \ldots$$

and thus also $\mathbb{P}^* \in \mathcal{C}(A, \mathcal{G})$. By Fatou's lemma, it holds that

$$\begin{aligned} D_{\mathcal{G}}\big(\mathbb{P}^* \parallel \mathbb{P}\big) &\leq \liminf_{n\to\infty} D_{\mathcal{G}}\big(\mathbb{P}^*_n \parallel \mathbb{P}\big) \\ &= \lim_{n\to\infty} \inf\big\{D_{\mathcal{G}}\big(\mathbb{Q} \parallel \mathbb{P}\big) : \mathbb{Q} \in \mathcal{C}(A_n, \mathcal{G})\big\} \\ &\leq \inf\big\{D_{\mathcal{G}}\big(\mathbb{Q} \parallel \mathbb{P}\big) : \mathbb{Q} \in \mathcal{C}(A)\big\} < \infty, \end{aligned}$$

whence $\mathbb{P}^*$, when restricted to $(\Omega, \mathcal{F})$, is the necessarily unique I-projection of $\mathbb{P}$ on $\mathcal{C}(A, \mathcal{G})$. Finally, if $\mathcal{G}' \subset \mathcal{G}$, then (3) shows that the $\mathbb{P}^*_n$ also converge to $\mathbb{P}^*$ in information on $(\Omega, \mathcal{G}')$. □

3. Convergence of conditioned processes

Let $(X_t : t \geq 0)$ be a time homogeneous Strong Markov process with state space E, right continuous and left limited paths, and defined on a canonical filtered probability space $(\Omega, \mathcal{F}, \mathbb{P}, \mathcal{F}_t : t \geq 0)$. The probability $\mathbb{P}$ is the law of the process $(X_t : t \geq 0)$ on path space, and its initial distribution is $\nu = \mathbb{P} \circ X_0^{-1}$. As usual, $\sigma(X_s : s \leq t) \subseteq \mathcal{F}_t$, and we introduce the lifetime ζ of the process,

$$\zeta = \inf\{t > 0 : X_t \notin E \text{ or } X_{t-} \notin E\},$$

assuming that $\mathbb{P}(\zeta > t) > 0$ for all t. To keep technical aspects at bay, we will always assume all functions $f : E \to \mathbb{R}$ to be measurable, and moreover that the process $t \mapsto f(X_t)$ is always locally bounded.

In this section, we shall find conditions ensuring the convergence of the conditioned processes X^n defined by

$$X^n = (X_s : s \geq 0 \,|\, \zeta > n).$$

Note that if $\mathbb{P}(\zeta = \infty) > 0$, this problem is trivial since then

$$\lim_{n\to\infty} \mathbb{P}(B \,|\, \zeta > n) = \mathbb{P}(B \cap \{\zeta = \infty\})/\mathbb{P}(\zeta = \infty), \quad B \in \mathcal{F}, \tag{5}$$

and under the measure on the right, the paths of $(X_t : t \geq 0)$ a.s. equal those of a Markov process with transition semigroup $P_t(x, dy)c(y)/c(x)$, where $P_t(x, dy)$ is the semigroup of $(X_t : t \geq 0)$ under the original path measure $\mathbb{P}$ and $c(x) = \mathbb{P}(\zeta = \infty \,|\, X_0 = x)$. By the discussion in the previous section, the limiting measure in (5) is the I-projection of $\mathbb{P}$ on the set $\mathcal{C} = \{\mathbb{Q} \in \mathcal{P}(\Omega, \mathcal{F}) : \mathbb{Q}(\zeta = \infty) = 1\}$, and the convergence in (5) actually occurs in information on $(\Omega, \mathcal{F})$.

In case $\mathbb{P}(\zeta = \infty) = 0$, we proceed using Lemma 4, with $A_n = \{\zeta > n\}$, $A = \{\zeta = \infty\}$. However, it is never possible to take $\mathcal{G} = \mathcal{F}$, since

$$D_{\mathcal{G}}\big(\mathbb{P}(\cdot \,|\, \zeta > n) \,\|\, \mathbb{P}\big) = -\log \mathbb{P}(\zeta > n),$$

and this converges to infinity as $n \to \infty$. Instead, we may hope at best that $\mathcal{G} = \mathcal{F}_t$ or $\mathcal{G} = \mathcal{F}_T$ works, where T is a stopping time.

We shall need a convenient class of probability measures on which to test the condition $\mathbb{Q} \in \mathcal{C}(A, \mathcal{G})$. To this end, make the definition

Definition 2. *A function $\psi \geq 0$ on E is called λ-invariant for the Markov process $(X_t : t \geq 0)$ if $e^{\lambda t}\psi(X_t)$ is a nonzero martingale on $(\Omega, \mathcal{F}_t, \mathbb{P})$.*

The functions which are λ-invariant for $(X_t : t \geq 0)$ typically arise as positive eigenfunctions of the local martingale generator (but the zero function is excluded). Recall that this is the (multivalued) operator L such that, for any (not necessarily bounded) function $f : E \to \mathbb{R}$ in its domain, Lf is a function such that

$$M_t^f = f(X_t) - f(X_0) - \int_0^t Lf(X_s)ds$$

is a right continuous, left limited local martingale on $(\Omega, \mathcal{F}_t, \mathbb{P})$. This means that there exist stopping times $T_n \to \infty$ such that $M^f_{t\wedge T_n}$ is a uniformly integrable martingale for each n. When f is bounded and M_t^f is a true martingale, then Lf reduces to the (Hille-Yosida) strong generator.

Recall that a stopping time $T : \Omega \to \mathbb{R}_+$ is a random variable such that $\{T \leq t\} \in \mathcal{F}_t$ for each t. We also define $\mathcal{F}_T = \{A \in \mathcal{F} : A \cap \{T \leq t\} \in \mathcal{F}_t\}$, which is again a σ-algebra, and then $\mathcal{F}_S \subseteq \mathcal{F}_T$ whenever $S \leq T$ are stopping times (Revuz and Yor, 1994).

Generally, if ψ is λ-invariant, then $L\psi = -\lambda\psi$. For the converse, if $L\psi = -\lambda\psi$, then by a straightforward calculation using Ito's formula,

$$e^{\lambda t}\psi(X_t) = \psi(X_0) + \int_0^t e^{\lambda s} dM_s^\psi. \tag{6}$$

Thus $e^{\lambda t}\psi(X_t)$ will be a martingale if M_t^ψ is and $\lambda \leq 0$, for example, but in general $e^{\lambda t}\psi(X_t)$ is only a local martingale.

Every λ-invariant function ψ gives rise to a probability measure $d\mathbb{Q}$ on $(\Omega, \vee_t \mathcal{F}_t)$, which is defined uniquely by

$$dQ|_{\mathcal{F}_t} = e^{\lambda t}\psi(X_t)d\mathbb{P}|_{\mathcal{F}_t}, \quad t \geq 0, \qquad \int \psi(X_0)d\mathbb{P} = 1. \tag{7}$$

This measure belongs to $\mathcal{C}(A, \mathcal{F}_t)$ for every t, since $\mathbb{Q}(\zeta > t) = \int e^{\lambda t}\psi(X_t)d\mathbb{P} = 1$.

The following result is true in discrete or continuous time.

Theorem 3. *Suppose there exists a λ-invariant function ψ satisfying*

$$\int \psi(X_t)\log\psi(X_t)d\mathbb{P} < \infty \quad \textit{for all } t, \tag{8}$$

then the measures $\mathbb{P}(\cdot \,|\, \zeta > n)$ *converge in information on each space* $(\Omega, \mathcal{F}_t)$ *to (the restriction of) a probability* $\mathbb{P}^* \in \mathcal{P}(\Omega, \vee_t \mathcal{F}_t)$. *This latter probability satisfies*

$$d\mathbb{P}^*|_{\mathcal{F}_t} = e^{\lambda_* t}\varphi_*(X_t) d\mathbb{P}|_{\mathcal{F}_t}, \quad \int \varphi_*(X_0) d\mathbb{P} = 1, \tag{9}$$

where φ_* *is some* λ_**-invariant function, and we have for every* t

$$\mathbb{P}^* = \textit{arg min}\big\{D_{\mathcal{F}_t}\big(\mathbb{Q} \parallel \mathbb{P}\big) : \mathbb{Q} \in \mathcal{C}(\zeta = \infty, \mathcal{F}_t)\big\}. \tag{10}$$

Proof. Define $\mathbb{Q} \in \mathcal{C}(\{\zeta = \infty\}, \mathcal{F}_t)$ as in (7), and compute

$$\begin{aligned} D_{\mathcal{F}_t}\big(\mathbb{Q} \parallel \mathbb{P}\big) &= \int \log\Big(e^{\lambda t}\psi(X_t)\Big) d\mathbb{Q} \\ &= \lambda t + \int e^{\lambda t}\psi(X_t) \log \psi(X_t) d\mathbb{P} < \infty, \end{aligned}$$

by (8). Applying Lemma 4 gives the convergence $D_{\mathcal{F}_t}\Big(\mathbb{P}(\cdot \,|\, \zeta > n) \parallel \mathbb{P}^*_t\Big) \to 0$, where $\mathbb{P}^*_t$ is the I-projection of $\mathbb{P}$, restricted to $(\Omega, \mathcal{F}_t)$, onto $\mathcal{C}(\{\zeta = \infty\}, \mathcal{F}_t)$. Moreover, we have $\mathbb{P}^*_t|_{\mathcal{F}_s} = \mathbb{P}^*_s$ for $t > s$. To exhibit the form of the limit, use the Markov property to get

$$\frac{\mathbb{P}(\zeta > n \,|\, \mathcal{F}_t)}{\mathbb{P}(\zeta > n)} = \frac{\mathbb{P}(\zeta > n - t \,|\, X_t)}{\mathbb{P}(\zeta > n)} = \varphi^*_{n,t}(X_t).$$

Since these functions are, for fixed t, uniformly integrable, a standard argument (Jacka and Roberts, 1996), implies that the limit $\varphi^* = \lim_{n\to\infty} \varphi^*_n$ is λ_*-invariant for some λ_*, that is, $\mathbb{E}\big(\varphi^*(X_t) \,|\, X_s\big) = e^{-\lambda_*(t-s)}\varphi^*(X_s)$. Finally, (10) holds because $\mathbb{P}^*_t = \mathbb{P}^*|_{\mathcal{F}_t}$ is the I-projection of $\mathbb{P}$ on $\mathcal{C}(\zeta = \infty, \mathcal{F}_t)$. □

We note that when viewed under the law $\mathbb{P}^*$, the coordinate process X_t is necessarily a Markov process. This follows from the fact that $\mathbb{P}^* \ll \mathbb{P}$ on every σ-algebra $\mathcal{F}_t$, and is a by-product of the convergence in information.

As remarked before the theorem, the λ-invariant functions ψ arise typically as eigenfunctions of the local martingale generator (e.g. the second order differential operator when X is a diffusion process, or the q-matrix when X is a Markov chain on a countable state space, see Rogers and Williams, 1987).

We now propose to give a sufficient condition which guarantees that a positive eigenfunction of L (with eigenvalue $-\lambda$) is actually λ-invariant. The proof uses Ito's formula for discontinuous processes (Protter, 1992).

Proposition 4. *Let $\psi \geq 0$ satisfy $L\psi = -\lambda\psi$, and suppose that ψ^2 belongs to the domain of L. Then the square bracket process of the local martingale*

$$M_t^\psi = \psi(X_t) - \psi(X_0) + \lambda \int_0^t \psi(X_s)ds$$

is given by the formula

$$[M^\psi, M^\psi]_t = \int_0^t \Big(L(\psi^2)(X_s) + 2\lambda\psi(X_s)^2 \Big) ds + \sum_{s \leq t} \big(\Delta\psi(X_s)\big)^2,$$

where $\Delta\psi(X_s) = \psi(X_s) - \lim_{u\uparrow s} \psi(X_u)$ is the jump in the process $\psi(X_s)$ at the time s.

In case that $\mathbb{E}[M^\psi, M^\psi]_t < \infty$ for all t, the process $e^{\lambda t}\psi(X_t)$ is also an $L^2(\Omega, \mathcal{F}_t, \mathbb{P})$ martingale, and (8) *holds, thus implying the convergence stated in Theorem 3.*

Proof. Expanding the process $\psi(X_t)^2$ alternately in terms of the operator L and using Ito's formula, we find

$$\begin{aligned}
\psi(X_t)^2 &= \psi(X_0)^2 + \int_0^t L(\psi^2)(X_s)ds + M_t^{\psi^2} \\
&= \psi(X_0)^2 + 2\int_0^t \psi(X_s)_- d\psi(X_s) + [\psi(X), \psi(X)]_t \\
&= \psi(X_0)^2 - 2\lambda \int_0^t \psi(X_s)_- \psi(X_s)ds + 2\int_0^t \psi(X_s)_- dM_s^\psi + [M^\psi, M^\psi]_t,
\end{aligned}$$

where $\psi(X_s)_- = \lim_{u\uparrow s} \psi(X_u)$, and $[M^\psi, M^\psi]_t = [\psi(X), \psi(X)]_t$ since the processes differ by a continuous process of finite variation. Regrouping terms, we arrive at

$$[M^\psi, M^\psi]_t = \int_0^t \Big(L(\psi^2)(X_s) + 2\lambda\psi(X_s)_- \psi(X_s) \Big) ds + R_t,$$

where $R_t = 2\int_0^t \psi(X_s)_- dM_s^\psi - M_t^{\psi^2}$ is a local martingale whose square bracket $[R, R]_t$ has zero continuous part. Hence it must be totally discontinuous, and may be written

$$R_t = \sum_{s \leq t} \big(\Delta M_s^\psi\big)^2 = \sum_{s \leq t} \big(\Delta\psi(X_s)\big)^2.$$

Finally, since the process $\psi(X_t)$ has right continuous paths with left limits, it can only jump a finite number of times in each finite time interval. Hence we have $\int_0^t \psi(X_s)_- \psi(X_s)ds = \int_0^t \psi(X_s)^2 ds$, which gives the formula for $[M^\psi, M^\psi]_t$ as stated.

We now prove the last part. Since ψ is an eigenfunction of L, we know that $N_t = e^{\lambda t}\psi(X_t)$ is a positive local martingale, and it is given by (6). Consequently,

$$\mathbb{E}[N, N]_t = \mathbb{E}\int_0^t e^{2\lambda s} d[M^\psi, M^\psi]_s \le e^{2(\lambda\vee 0)t}\mathbb{E}[M^\psi, M^\psi]_t < \infty$$

holds for every t, so by (Protter, 1992, p.66, Corollary 3) the process N is an L^2 martingale and satisfies $\mathbb{E}[\sup_{s\le t} |N_t|^2] < \infty$, which is clearly stronger than (8). Hence Theorem 3 holds. □

Remark. In case that the Markov process X_t admits a Lévy system, the sum over the jumps of X can be written in a more convenient form. For example, let X_t be the minimal Markov chain on a countable state space $E = \{1, 2, 3, \dots\}$ with a given q-matrix $q(x, y)$:

$$q(x, y) = \lim_{t\to 0} \frac{1}{t}\Big(\mathbb{P}\big(X_t = y \,|\, X_0 = x\big) - \delta(x, y)\Big), \quad q(x) := -q(x, x) \ge 0.$$

The generator L can now be identified with the q-matrix: the function $f : E \to \mathbb{R}$ belongs to the domain of L and then $Lf(x) = \sum_y q(x, y)f(y)$ provided this sum is finite. Since the entries $q(x, y)$ are nonnegative for $x \ne y$, we can define the Lévy kernel $Q(x, A) = \sum_{y\in A\setminus\{x\}} q(x, y)$. Then we have the fundamental result (Rogers and Williams, IV.21, p.36) that

$$\mathbb{E}\sum_{s\le t}\big(\Delta\psi(X_s)\big)^2 = \mathbb{E}\int_0^t \sum_b Q(X_{s-}, b)\big(\psi(b) - \psi(X_{s-})\big)^2 ds, \tag{11}$$

and of course

$$\mathbb{E}\int_0^t \Big(L(\psi^2)(X_s) + 2\lambda\psi(X_s)^2\Big)ds = \mathbb{E}\int \Big(\sum_y q(X_s, y)\psi(y)^2 + 2\lambda\psi(X_s)^2\Big)ds. \tag{12}$$

If both (11) and (12) are finite for all t, then Theorem 3 applies.

Acknowledgement. I thank Andrew Barron for a very interesting past conversation. This research was supported by the European Union TMR network ERB-FMRX-CT96-0096.

References

[1] Bolthausen, E. (1993) Stochastic processes with long range interactions of the paths. Doeblin and modern probability (Blaubeuren, 1991), 297–319, *Contemp. Math.*, 149, AMS, Providence, RI.

[2] Csiszar, I. (1975) I-divergence geometry of probability distributions and minimization problems. *Ann. Probab.* Vol. 3, No. 1, 146-158.

[3] Csiszar, I. (1984) Sanov property, generalized I-projection and a conditional limit theorem. *Ann. Probab.* Vol. 12, No. 3, 768-793.

[4] Dembo, A. and Zeitouni, O. (1998) *Large Deviations. Techniques and Applications.* Springer Verlag. Second edition.

[5] Jacka, S.D. and Roberts, G.O. (1995). Weak convergence of conditioned processes on a countable state space. *J. Appl. Probab.* **32**, 902-916.

[6] Pollett, P.K. (1995) Modelling the long-term behaviour of evanescent ecological systems. *Ecological Modelling* **86**, 135-139.

[7] Protter, P. (1990) *Stochastic Integration and Differential Equations. A New Approach* Springer Verlag.

[8] *RESIM'99: Second International Workshop on Rare Event Simulation, March 11-12, 1999* University of Twente, Enschede, The Netherlands. Workshop organized by V.F. Nicola and J.H.A. de Smit.

[9] Rogers, C. and Williams, D. (1987) *Diffusions, Markov Processes and Martingales. Volume 2: Ito Calculus.* Wiley.

Dipartimento di Matematica, Università degli Studi Roma Tre, Largo S. Leonardo Murialdo 1, 00146 Roma, Italy

E-mail address: breyer@mat.uniroma3.it