Chapter 18

Invariant Measures

If $p(t,x,dy)$ are the transition probabilities of a Markov Process on a Polish space X, then an invariant probability distribution for the process is a distribution μ on X that satisfies

$$\int p(t,x,A)d\mu(x) = \mu(A)$$

for all Borel sets A and all $t > 0$. In general μ need not be unique. But if for every pair $x, y \in X$, there exists a $t = t(x,y) < \infty$ such that $\sup_A |p(t,x,A) - p(t,y,A)| < 1$ then we will be able to show that μ is unique. To begin with we construct a translation invariant measure $P = P_\mu$ on the space of maps $(-\infty,\infty) \to X$, that has μ as the stationary marginal distribution. P_μ may or may not be ergodic with respect to time translations.

Theorem 18.1. *The only way P_μ can fail to be ergodic is if there is a decomposition of X in to Borel sets X_1, X_2 such that $X = X_1 \cup X_2$ with $X_1 \cap X_2 = \emptyset$ with the property that for $i = 1,2$, we have $0 < \mu(X_i) < 1$ and*

$$P_\mu[x(t) \in X_i \ \forall t \in (-\infty,\infty)] = \mu(X_i)$$

i.e. $\mu = \mu(X_1)\mu_1 + \mu(X_2)\mu_2$ where the restrictions μ_1, μ_2 of μ to X_1, X_2 are themselves invariant and $p(t,x,X_i) = 1$ for almost all x with respect to μ_i.

Remark 18.1. In particular if $\{p(t,x,dy)\}$ has two distinct invariant measures μ_1, μ_2, then $P_\mu = \frac{1}{2}[P_{\mu_1} + P_{\mu_2}]$ is not ergodic and admits a decomposition of the type stated in the theorem, which can not happen if $p(t,x,\dot{\,})$ and $p(t,y,\cdot)$ are not mutually orthogonal when x and y come from X_1 and X_2.

Proof. Let us recall the notation $\mathcal{F}_t^s = \sigma\{x(u) : s \le u \le t\}$. $\mathcal{F}_{-\infty} = \cap_t \mathcal{F}_t^{-\infty}$. If P_μ is not ergodic, there is an invariant set E with $0 < P_\mu(E) < 1$. Given an invariant set E and given any $\delta > 0$, there is a t_δ and a set $E_\delta \in \mathcal{F}_{t_\delta}^{-t_\delta}$ such that $P(E \Delta E_\delta) < \delta$. The translation invariance invariance allows us to replace E_δ by $T_{t_\delta} E_\delta$ and this means E can be well approximated by sets from $\mathcal{F}_0^{-\infty}$

and therefore $E \in \mathcal{F}_0^{-\infty}$. Now again by invariance $E \in \mathcal{F}_t^{-\infty}$ for every t and therefore in the remote past $\mathcal{F}_{-\infty}$. A similar argument places E in the remote future. But for a Markov process past and future are independent given the present. An invariant set is therefore independent of itself given the present. But such an event is trivial. Therefore an invariant set is trivial given the present, i.e. it is a function of the present. It is of the form $\{x(0) \in X_1\}$ for some $X_1 \subset X$. □

To verify that a given μ is invariant for $p(t,x,dy)$ is often difficult. The transition probabilities may not be explicit. They are often defined through the infinitesimal generator. Let us suppose for simplicity we are dealing with a diffusion on a compact manifold X, given in local coordinates as a second order strictly elliptic differential operator with continuous coefficients. We know how $\mathcal{L}$ acts on smooth functions and we have a unique family of processes $\{P_x;\, x \in X\}$ on $\Omega = C[[0,\infty);X]$ satisfying $P_x[x(0) = x] = 1$ and

$$f(x(t)) - f(x(0)) - \int_0^t (\mathcal{L}f)(x(s))ds$$

is a martingale for smooth f, with respect to $(\Omega, \mathcal{F}_t^0, P_x)$, where $\mathcal{F}_t^0$ is the the natural filtration $\sigma\{x(s) : 0 \le s \le t\}$. Formally we need to check that $\mathcal{L}^*\mu = 0$. This can be checked "weakly" by testing

$$\int (\mathcal{L}f)(x)d\mu(x) = 0$$

for all smooth f. Is that enough? Functional analysis does not answer the question. With continuous coefficients, the relevant PDE's can be solved only in Sobolev spaces $W_{2,p}$ and unless $\mu \in L_q$ for some $q > 1$ we can not complete the proof. However it is enough and that is the main thrust of the following theorem.

Theorem 18.2. *Let μ on X satisfy*

$$\int (\mathcal{L}f)(x)d\mu(x) = 0$$

for all smooth $f \in \mathcal{S}$. We can then construct a family $P_{h,x}$ of Markov chains taking values in X with transition probabilities $\pi_h(x,dy)$, and having μ as an invariant distribution for the chain, i.e.

$$\int \pi_h(x,A)\mu(dx) = \mu(A)$$

for all Borel subsets of X. The piecewise constant processes $\{P_{h,x} : x \in X\}$ on $D[[0,T],X]$ starting from the initial distribution μ, with transition probability π_h and time step h will converge weakly to the process $P_\mu = \int P_x\, d\mu(x)$. In particular μ being the common marginal of $P_{h,x}$, it will also be the common marginal of P_μ. This implies μ is invariant for the Markov process with generator $\mathcal{L}$.

Proof. We start with the operator $\mathcal{L}_h = (I - h\mathcal{L})$ mapping $\mathcal{S} \to \mathcal{M}_h$. According to Lemma 18.3 there is an inverse $T_h : \mathcal{M}_h \to \mathcal{S}$ that is a contraction, mapping nonnegative functions to nonnegative functions and mapping $1 \to 1$. Fix $h > 0$. We look at functions in $C(X \times X)$ of the form

$$\phi(x, y) = \sum_{i=1}^{k} F_i(x) G_i(y) + H(y)$$

where $G_i = \mathcal{L}_h H_i \in \mathcal{M}_h$, and $F_i, H \in C(X)$. We define a linear functional $\Lambda_h(\phi)$ on such ϕ

$$\Lambda_h(\phi) = \int [H(x) + \sum_i F_i(x)(T_h H_i)(x)] d\mu(x) = \int [H(x) + \sum_i F_i(x) G_i(x)] d\mu(x)$$

According to Lemma 18.4 this is a well defined nonnegative linear function which according to Lemma 18.5 can be extended to all of $C(X \times X)$ yielding a probability distribution λ on $X \times X$ that has both marginals equal to μ. This is then used to define a Markov process $P_{h,\mu}$ that has μ for an invariant distribution. Finally in Lemma 18.6 it is shown that $P_{h,\mu}$ has a limit as $h \to$ which is shown to be the diffusion with generator $\mathcal{L}$ and it has μ as the stationary distribution and this will complete the proof.

Lemma 18.3. *$T_h : \mathcal{M}_h \to C(X)$ is positivity preserving and $\sup_x |(T_h F(x)| \leq \sup_x |F(x)|$.*

Proof. If $\mathcal{L}_h F = (I - h\mathcal{L})F = G \geq 0$, then at x_0 where the infimum of F is attained, $(\mathcal{L}F)(x_0) \geq 0$ by the maximum principle. Therefore $F(x_0) = G(x_0) + h(\mathcal{L}F)(x_0) \geq G(x_0)$ and $G \geq 0$ implies $F \geq 0$. In particular $G = 0$ implies $F = 0$. $(I - h\mathcal{L})$ therefore has an inverse. Moreover $\mathcal{L}1 = 0$ which implies that $T_h 1 = 1$. A linear positivity preserving map that maps 1 to 1 is a contraction. □

Lemma 18.4. *If $\phi(x, y) \geq 0$ on $X \times X$, then $\Lambda_h(\phi) \geq 0$ and it is well defined.*

Proof. If $\phi(x, y) = H(y) + \sum_{i=1}^{k} F_i(x) G_i(y) \geq 0$ for all x, in particular with $\Psi(G_1, \ldots, G_k) = \inf_{x \in X} \sum F_i(x) G_i$, a concave function of $G_1, \ldots, G_k$

$$\begin{aligned}
H(y) &+ \inf_{x \in X} \sum_{i=1}^{k} F_i(x) G_i(y) \\
&= H(y) + \Psi(G_1(y), \ldots, G_k(y)) \\
&= H(y) + \Psi((\mathcal{L}_h H_1)(y), \ldots, ((\mathcal{L}_h H_k)(y)) \\
&\geq 0
\end{aligned}$$

Consider the integral

$$I(t) = \int \big[H(x) + \Psi((\mathcal{L}_t H_1)(x), \ldots, ((\mathcal{L}_t H_k)(x))\big] d\mu(x)$$

for $0 \le t \le h$. Let Ψ be any concave function, then $I(t)$ will be a concave function of t. If we can show that $I'(0) \le 0$ then it will follow that $I(0) \ge I(h) \ge 0$. This in turn will imply that

$$\begin{aligned} 0 \le I(0) &= \int \big[H(x) + \Psi(H_1(x), \dots, H_k(x))\big]\, d\mu(x) \\ &\le \int \big[H(x) + \sum_{i=1}^{k} F_i(x) H_i(x)\big]\, d\mu(x) \\ &= \Lambda_h(\phi) \end{aligned}$$

In particular if $\phi_1 \equiv \phi_2$, then $\Lambda(\phi_1) = \Lambda(\phi_2)$.

To show that $I'(0) \le 0$ we can assume that Ψ is smooth. From the maximum principle, or in the case of diffusions by direct calculation, it follows that for concave Ψ

$$\mathcal{L}\Psi(u_1, u_2, \dots, u_k) \le \sum_{j=1}^{k} \Psi_j(u_1, \dots, u_k) \mathcal{L} u_j$$

Therefore from

$$\int \mathcal{L}\Psi(u_1, u_2, \dots, u_k)(x) d\mu(x) = 0$$

it follows that

$$\int [\sum_{j=1}^{k} \Psi_j(u_1, \dots, u_k) \mathcal{L} u_j] d\mu(x) \ge 0$$

Hence

$$I'(0) = -\int \sum_{j=1}^{k} \frac{\partial \Psi}{\partial H_j}(H_1(x), \dots, H_k(x))(\mathcal{L} H_j)(x) \le 0$$

$\square$

Lemma 18.5. *$\Lambda(\cdot)$ extends to all of $C(X \times X)$ as a continuous linear functional. By Riesz's representation theorem there is a $\lambda(dx, dy)$ such that*

$$\Lambda(\phi) = \int_{X \times X} \phi(x, y) \lambda(dx, dy)$$

The measure $\lambda(A \times X) = \lambda(X \times A) = \mu(A)$. In particular if we write

$$\lambda(dx, dy) = \mu(dx) \pi_h(x, dy)$$

then $\mu(dx)\pi(x, A) = \mu(A)$

Proof. The extension is just Hahn-Banach theorem. Any nonnegative linear functional Λ defined on a subspace of the space of bounded continuous functions of a compact metric space $C(X)$ that contains constants and has the property $\lambda(1) = 1$ is a bounded linear functional with norm 1. For any f, $\|f\| \pm f \ge 0$

and therefore $\|f\|\Lambda(1) \pm \lambda(f) \geq 0$. Conversely if $\Lambda(1) = 1$ and $|\Lambda(f)| \leq \|f\|$ then if nonnegative, then $\|f - \frac{1}{2}\|f\|\| = \frac{1}{2}\|f\|$. Therefore

$$\Lambda(f) = \frac{1}{2}\|f\|\Lambda(1) - \frac{1}{2}\|f\| \geq 0$$

Our Λ on $C(X \times X)$ satisfies $\Lambda(h(x)) = \Lambda(h(y)) = \int h(x)\mu(dx)$ and this identifies both marginals of $\lambda(dx, dy)$ as μ. □

Lemma 18.6. *The piecewise constant Markov process with transition probability $\pi_h(x, dy)$, initial distribution μ and time step h, is tight in the space $D[[0,T], X]$ and any limit point P is a stationary Markov process with marginal μ and satisfies the Martingale relations for $\mathcal{L}$. Therefore $P = \int P_x d\mu(x)$*

Proof. If we consider $g_h = f - h\mathcal{L}f$, then under $P_{h,x}$, $\pi_h g_h = f$ and

$$g_h(X(nh) - g_h(x) - \sum_{j=1}^{n}(f - g_h)(X((j-1)h)$$

is a martingale. To prove compactness we only need to estimate the exit time from a small ball. Consider such a ball S of size δ with center at x. Let f be a smooth function with $f = 1$ on S^c and $f(x) = 0$. Its smoothness will depend only on δ. In particular $\|\mathcal{L}f\| \leq C_\delta$ uniformly in x. g_h will be uniformly close to f and for small h, again uniformly in x, $g_h(x_0) \leq C_\delta h$ and $g_h(x) \geq 1 - C_\delta h$ on $S(x,\delta)^c$. $\|f - g_h\| = h\|\mathcal{L}f\| \leq hC_\delta$. By Doob's stopping theorem, if τ is the exit time from $S(x,\delta)$, then $P_{h,x}[\tau \leq t]$ can be estimated from

$$E[g_h(x(\tau \wedge t))] - g_h(x) \leq C_\delta t$$

which leads to

$$P_{h,x}[\tau \leq t] \leq C_\delta(t + h)$$

which is enough to prove tightness. In the limit

$$f(x(t)) - f(x) - \int_0^t (\mathcal{L}f)(x(s))ds$$

is a martingale and μ is the common distribution of $x(t)$ under $\int P_x d\mu(x)$.

□

This completes the proof of the theorem.

□

Remark 18.2. We do not really need compactness. If we had R^d for the state space, nothing really would change, except for Reisz representation theorem. We would need instead the following version of it. Let $\mathcal{C} = C(X \times Y)$ be the space of bounded continuous functions on $X \times Y$ where X, Y are locally compact

Polish spaces. Let Λ be a nonnegative linear functional on $\mathcal{C}$ with $\Lambda(1) = 1$. If there are probability measures α, β on X and Y such that

$$\int f(x)\alpha(dx) = \Lambda(f) \text{ for } f \in C(X) \subset \mathcal{C}$$

and

$$\int g(y)\beta(dy) = \Lambda(g) \text{ for } g \in C(Y) \subset \mathcal{C}$$

then there is a probability measure λ on $X \times Y$ with marginals α and β such that

$$\Lambda(h) = \int h(x,y)\lambda(dx,dy) \text{ for } h \in \mathcal{C}$$

The proof requires us to show that Λ is σ-smooth, i.e if $h_n(x,y) \downarrow 0$, then $\Lambda(h_n) \downarrow 0$. Given any $\epsilon > 0$ we can find continuous functions $f(x)$ and $g(y)$ with compact support such that $0 \le f(x) \le 1, 0 \le g(y) \le 1$, $\int(1-f(x))\alpha(dx) + \int(1-g(y))\beta(dy) \le \epsilon$. We can then verify that

$$\begin{aligned}\Lambda(h_n) &= \Lambda(h_n(x,y)f(x)g(y)) + \Lambda(h_n(x,y)(1 - f(x)g(y)) \\ &\le \sup_{x,y} h_n(x,y)f(x)g(y) + \Lambda(1 - f(x)) + \Lambda(1 - g(y)) \\ &\le \sup_{x,y} h_n(x,y)f(x)g(y) + \epsilon\end{aligned}$$

By Dini's theorem the first term goes to 0. ϵ can be made arbitrarily small.

Having established that a measure is invariant for a Markov process with transition probabilities $p(t,x,dy)$ two natural questions arise. Is it unique and does $p(t,x,\cdot)$ converge to it as $t \to \infty$ for every x? In the context of elliptic diffusions on R^d the answer to both questions is yes. We will prove it in a few steps. They depend on certain estimates from PDE. We state it without proof.

For any $t > 0$, $p(t,x,\cdot)$ is continuous as a map $(0,\infty) \times R^d$ into $L_1(R^d)$. In particular for any Borel set A, $p(t,x,A)$ is continuous as a function of $t > 0$ and x.

Lemma 18.7. *We have a diffusion process with transition probability $p(t,x,A)$ that corresponds to a uniformly elliptic operator*

$$\mathcal{L} = \sum_{i,j} a_{i,j}(x)D_{ij} + \sum_j b_j(x)D_j$$

with continuous coefficients. $\{a_{i,j}(x)\}$ is bounded and uniformly positive definite. $b_j(x)$ are have at most linear growth. This is enough to yield the continuity of $p(t,x,A)$ for $t > 0$ in (t,x) for every A. If G is any nonempty open set $p(t,x,G) > 0$ for all $x \in R^d$ and $t > 0$. If for some $t > 0$ and $x \in R^d$, $p(t,x,A) = 0$, then for any $x \in R^d$ and $t > 0$, $p(t,x,A) = 0$.

Proof. Assume $p(t_0, x_0, G) = 0$ for some non empty open set, i.e. $P_x[x(t_0) \in G] = 0$. Since $\{P : P[x(t_0) \in G] = 0\}$ is a closed set, this is true for any process which is a limit of processes that are absolutely continuous with respect to P_x. In particular the process with generator

$$\tilde{\mathcal{L}} = \sum_{i,j} a_{i,j}(x) D_{ij} + \lambda \sum_j c_j(x) D_j$$

where $c_j(x) = -\frac{x-x_0}{\|x-x_0\|}$ and λ is large will at time t_0 stay away from a neighborhood of x_0. Computing with $f(x) = \|x - x_0\|^2$

$$(\tilde{\mathcal{L}} f)(x) \leq -2\lambda \|x - x_0\|^2 + C$$

If $g(t) = E[\|x(t) - x_0\|^2]$ then $g'(t) \leq -2\lambda g(t) + C$. It is easy to see that $g(t_0)$ will be small if λ is large. If $p(t, x, A)$ is positive for some $t > 0, x$ and A, then $p(t, x, A) = \int p(\frac{t}{2}, y, A) p(\frac{t}{2}, x, dy)$ and therefore $p(\frac{t}{2}, y, A) > 0$ for some y. But $p(\frac{t}{2}, y, A)$ is continuous in y and therefore it is uniformly positive on some open set G around y. Since $p(t, x, A) \geq \int_G p(\frac{t}{2}, y, A) p(\frac{t}{2}, x, dy)$ and $p(\frac{t}{2}, x, G) > 0$ the lemma is proved. □

Lemma 18.8. *The invariant probability measure if it exists is unique.*

Proof. Since either $p(t, x, A) \equiv 0$ for all $x \in R^d, t > 0$ or strictly positive for all $x \in R^d, t > 0$, the set of A 's for which any invariant measure is positive are exactly the same. There can not be two mutually orthogonal invariant measures. Therefore it is unique. □

Lemma 18.9. *If μ is an invariant probability measure then $p(t, x, \cdot) \to \mu$ as $t \to \infty$ for every x.*

Proof.

$$\begin{aligned}
\|p(t+s, x, \cdot) - p(t+s, y, \cdot)\|_{L_1} &= \int |\int [p(t, x, z) - p(t, y, z)] p(s, z, z') dz | dz' \\
&\leq \int \int |p(t, x, z) - p(t, y, z)| p(s, z, z') dz dz' \\
&\leq \int |p(t, x, z) - p(t, y, z)| dz
\end{aligned}$$

Hence $\Delta(t, x, y) = \|p(t, x, \cdot) - p(t, y, \cdot)\|$ is $\downarrow$ as a function of t.Let us denote by $\Delta(x, y)$ its limit as $t \to \infty$. It is sufficient to prove that $\Delta(x, y)$ is identically zero. It is upper semi continuous in x and y and is identically 0 on the diagonal

$x = y$.

$$\begin{aligned}
&\int p(s,x,y)p(s,x',y')\Delta(y,y')dydy' \\
&\quad = \lim_{t\to\infty} \int p(s,x,y)p(s,x',y')|p(t,y,z) - p(t,y',z)|dzdydy' \\
&\quad \geq \lim_{t\to\infty} \int |\int p(s,x,y)p(s,x',y')(p(t,y,z) - p(t,y',z))dydy'|dz \\
&\quad = \lim_{t\to\infty} \int |\int p(s+t,x,z) - p(s+t,x',z)|dz \\
&\quad = \Delta(x,x')
\end{aligned}$$

If we consider the process $Z(t) = \Delta(x(t), y(t))$ where $x(t), y(t)$ are independent copies of the Markov process then $Z(t)$ is a nonnegative sub-martingale that is bounded by 2. Hence $Z(t)$ has a limit as $t \to \infty$. Since the process is recurrent on $R^d \times R^d$ and comes arbitrarily close to the diagonal infinitely often this can only mean $\Delta(x,y) \equiv 0$. $\square$