Stationary distributions of continuous time Markov chains

Jonathon Peterson

April 13, 2012

The following are some notes containing the statement and proof of some theorems I covered in class regarding explicit formulas for the stationary distribution and interpretations of the stationary distribution as the limiting fraction of time spent in states.

1 Stationary measures in continuous time

The following theorem is an analog of the explicit formula for stationary measures for discrete time Markov chains (Theorem 1.20).

**Theorem 1.** If $X_t$ is an irreducible continuous time Markov process and all states are recurrent, then for any $x \in I$ the measure $\mu_x$ defined by

$$
\mu_x(y) = \int_0^\infty P_x(X_t = y, t < T_x) \, dt
$$

is a stationary measure.

**Remark 2.** Note that we can also write

$$
\mu_x(y) = \int_0^\infty E[1_{X_t=y, t<T_x}] \, dt
$$

$$
= E \left[ \int_0^{T_x} 1_{X_t=y, t<T_x} \, dt \right]
$$

$$
= E \left[ \int_0^{T_x} 1_{X_t=y} \, dt \right] \quad (1)
$$

That is, $\mu_x(y)$ has the interpretation of the amount of time spent in $y$ before the first return to $x$ when starting originally from $x$. 

Proof of Theorem 1. To show that $\mu_x$ is a stationary distribution, it is enough to check that $\mu_x Q = 0$. To this end, we first use the formula for $\mu_x(y)$ to write the $z$-th entry of $\mu_x Q$ as

$$
(\mu_x Q)(z) = \sum_{y \in I} \mu_x(y) Q(y, z)
= \sum_{y \in I \setminus \{z\}} \mu_x(y) q(y, z) - \lambda_z \mu_x(z)
= \sum_{y \in I \setminus \{z\}} \left( \int_0^\infty P_x(X_t = y, t < T_x) \, dt \right) q(y, z) - \lambda_z \int_0^\infty P_x(X_t = z, t < T_x) \, dt
= \int_0^\infty \left\{ \sum_{y \in I \setminus \{z\}} P_x(X_t = y, t < T_x) q(y, z) - \lambda_z P_x(X_t = z, t < T_x) \right\} \, dt \tag{2}
$$

The term inside the braces looks similar to the Kolmogorov forward differential equation, but the presence of the condition $t < T_x$ inside the probabilities makes it different. To rectify this we will create a new Markov process $\hat{X}_t$ on state space $\hat{I} = I \cup \{\zeta\}$ (where $\zeta$ is a new state not originally in $I$). The new Markov chain will have jump rates $\hat{q}$ that are given by

$$
\hat{q}(y, z) = \begin{cases} 
q(y, z) & y = x \text{ or } z \in I \setminus \{x\} \\
q(y, x) & y \in I \setminus \{x\} \text{ and } z = \zeta \\
0 & y \in I \setminus \{x\} \text{ and } z = x \\
0 & y = x \text{ and } z = \zeta \\
0 & y = \zeta \text{ and } z \in I,
\end{cases}
$$

Thus, the new Markov process $\hat{X}_t$ behaves the same as the old Markov process except all jumps into $x$ from another site are redirected to be jumps to a new absorbing state $\zeta$. With $\hat{X}_t$ defined in this way, we obtain that

$$
P_x(X_t = y, t < T_x) = P_x(\hat{X}_t = y), \quad \forall y \in I.
$$

We’ll continue our analysis of $(\mu_x Q)(z)$ depending on whether or not $z = x$.

**Case I:** $z \neq x$.

If $z \neq x$ then the terms inside the braces in (2) are

$$
\sum_{y \in I \setminus \{z\}} P_x(\hat{X}_t = y) q(y, z) - \lambda_z P_x(\hat{X}_t = z)
= \sum_{y \in I \setminus \{z\}} P_x(\hat{X}_t = y) \hat{q}(y, z) - \lambda_z P_x(\hat{X}_t = z)
= \frac{d}{dt} P_x(\hat{X}_t = z),
$$
where the last equality follows from the Kolmogorov forward equations for the Markov process $\hat{X}_t$. Note also that the in the first equality above we were able to extend the sum from $y \in I \setminus \{x\}$ to $y \in \hat{I} \setminus \{x\}$ since $\hat{q}(\zeta, z) = 0$.

Therefore, putting this back into (2) we obtain that

$$(\mu_x Q)(z) = \int_0^\infty \frac{d}{dt} P_x(\hat{X}_t = z) \, dt$$

$$= \lim_{s \to \infty} P_x(\hat{X}_s = z) - P_x(\hat{X}_0 = z)$$

$$= 0,$$

where the last equality holds because $P_x(\hat{X}_0 = z) = 0$ (since $z \neq x$) and $P_x(\hat{X}_s = z) \leq P_x(T_x > s) \to 0$ as $s \to \infty$ (since $X_t$ is recurrent). Thus, we’ve shown that $(\mu_x Q)(z) = 0$ whenever $z \neq x$ and it remains to show this is also true when $z = x$.

**Case I: $z = x$.**

If $z = x$ then the terms inside the braces in (2) are

$$\sum_{y \in I \setminus \{x\}} P_x(\hat{X}_t = y)q(y, x) - \lambda_x P_x(\hat{X}_t = x)$$

$$= \sum_{y \in I \setminus \{x\}} P_x(\hat{X}_t = y)\hat{q}(y, \zeta) - \lambda_x P_x(\hat{X}_t = x)$$

$$= \sum_{y \in I} P_x(\hat{X}_t = y)\hat{q}(y, \zeta) - \hat{\lambda}_\zeta P_x(\hat{X}_t = \zeta) - \lambda_x P_x(\hat{X}_t = x)$$

$$= \frac{d}{dt} \left\{ P_x(\hat{X}_t = \zeta) \right\} - \lambda_x P_x(\hat{X}_t = x).$$

where in the second equality is true because $\hat{q}(x, \zeta) = 0$ and $\hat{\lambda}_\zeta = 0$ and the last equality follows from the Kolmogorov forward equation for $\hat{X}_t$. Also, note that since $\hat{X}_t$ can never return to $x$ once it leaves, $P_x(\hat{X}_t = x) = e^{-\lambda x t}$ since this is just the probability that the Markov process hasn’t left $x$ yet by time $t$.

Putting all of this back into (2) we obtain that

$$(\mu_x Q)(x) = \int_0^\infty \frac{d}{dt} \left\{ P_x(\hat{X}_t = \zeta) \right\} \, dt - \int_0^\infty \lambda_x e^{-\lambda x t} \, dt$$

$$= \lim_{s \to \infty} P_x(\hat{X}_s = \zeta) - P_x(\hat{X}_0 = \zeta) - 1,$$

$$= 0,$$

where the last equality holds since $P_x(\hat{X}_s = \zeta) = P_x(T_x < s) \to 1$ as $s \to \infty$ since $X_t$ is recurrent.

Thus, we have shown that $(\mu_x Q)(z) = 0$ for all $z \in I$ and so $\mu_x$ is a stationary measure. We should also check that the definition of $\mu_x$ isn’t trivial. That is, $\mu_x(y) \in (0, \infty)$. First, note
that $\mu_x(x) = 1/\lambda_x$ since the holding time at $x$ is $\operatorname{Exp}(\lambda_x)$. Secondly, since $\mu_x$ is stationary we know that $\mu_x p_t = \mu_x$ for any $t > 0$ and so

$$\frac{1}{\lambda_x} = \mu_x(x) = (\mu_x p_t)(x) \geq \mu_x(y)p_t(y, x).$$

Since $X_t$ is irreducible we know that $p_t(y, x) > 0$ for any $t > 0$ and so $\mu_x(y) \leq \frac{1}{\lambda_x p_t(y, x)} < \infty$.

To show that $\mu_x(y) > 0$, note that since (1) shows that $\mu_x(y)$ is the expected amount of time spent in $y$ between visits to $x$, the strong Markov property implies that this is at least the probability of reaching $y$ before returning to $x$ times the expected amount of time spent in $y$ before the first jump out of $y$. That is,

$$\mu_x(y) = E_x \left[ \int_0^{T_x} 1_{X_t = y} \, dt \right] = E_x \left[ \int_0^{T_x} 1_{X_t = y} \, dt \right] \geq E_y \left[ \int_0^{T_y} 1_{X_t = y} \, dt \right] \geq E_y \left[ \int_{\min_{x \neq y} T_x}^{T_y} 1_{X_t = y} \, dt \right] = E_x \left[ T_x \right].$$

Since $X_t$ is irreducible we have that $P_x(V_y < \infty) > 0$ and so $\mu_x(y) > 0$.

As in the case of discrete time Markov chains we can normalize $\mu_x$ to obtain a stationary distribution (at least when $E_x[T_x] < \infty$).

**Corollary 2.1.** If $X_t$ is irreducible and positive recurrent, then a stationary distribution can be defined by

$$\pi(y) = \frac{\mu_x(y)}{E_x[T_x]}$$

for any $x \in I$.

**Proof.** This follows Theorem 1 and the fact that

$$\sum_y \mu_x(y) = \int_0^\infty \sum_y P_x(X_t = y, t < T_x) \, dt = \int_0^\infty P_x(t < T_x) \, dt = E_x[T_x].$$
2 Interpretations of the limiting distribution

If we know that a limiting distribution $\pi$ exists then $\pi(j)$ is approximately the probability that the Markov process will be in state $j$ at some very large time $t$. We will show in this section that $\pi(j)$ also is equal to the limiting fraction of time that the Markov process spends in state $j$. To make this precise, define

$$L_t(y) = \int_0^t 1_{X_s = y} ds.$$  

**Theorem 3.** If $X_t$ is irreducible and positive recurrent, then

$$\lim_{t \to \infty} \frac{L_t(y)}{t} = \frac{\mu_x(y)}{E_x[T_x]},$$

where the above limit holds with probability one, for any $x \in I$, and irrespective of the starting location. Moreover, since the limit is the same for any $x \in I$ we have that

$$\lim_{t \to \infty} \frac{L_t(y)}{t} = \frac{1}{\lambda_y E_y[T_y]},$$

To relate this limit to stationary distributions we need the following theorem.

**Theorem 4.** If $X_t$ is irreducible and positive recurrent, then there is a unique stationary distribution $\pi$. Moreover, $\pi$ has the formula

$$\pi(y) = \frac{1}{\lambda_y E_y[T_y]},$$

**Proof of Theorem 3.** Suppose that the Markov chain starts at $X_0 = x$. Let $T_x = \tau_1 < \tau_2 < \tau_3 < \ldots$ be the times of successive returns to $x$ and let

$$r_i = \int_{\tau_{i-1}}^{\tau_i} 1_{X_s = y} ds$$

be the amount of time spent in state $y$ between times $\tau_{i-1}$ and $\tau_i$ (here we let $\tau_0 = 0$ by convention). Also, let

$$N(t) = \max\{n \geq 0 : \tau_n \leq t\}$$

be the number of returns to $x$ by time $t$.

The law of large numbers from renewal theory (Theorem 3.1 in the book) implies that

$$\lim_{t \to \infty} \frac{N(t)}{t} = \frac{1}{E[\tau_1]} = \frac{1}{E_x[T_x]},$$

(3)
with probability one. Also, since the $r_i$ are i.i.d., the law of large numbers implies that
\[
\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} r_i = E[r_1] = E_x \left[ \int_0^{T_x} 1_{X_t = y} \, dt \right] = \mu_x(y). \tag{4}
\]

Now, since the definition of $N(t)$ implies that $\tau_{N(t)} \leq t < \tau_{N(t)+1}$ we have that
\[
\int_0^{\tau_{N(t)}} 1_{X_s = y} \, ds = \sum_{i=1}^{N(t)} r_i \leq L_t(y) \leq \sum_{i=1}^{N(t)+1} r_i = \int_0^{\tau_{N(t)+1}} 1_{X_s = y} \, ds.
\]
Therefore,
\[
\frac{N(t)}{t} \frac{1}{N(t)} \sum_{i=1}^{N(t)} r_i \leq \frac{L_t(y)}{t} \leq \frac{N(t) + 1}{t} \frac{1}{N(t)+1} \sum_{i=1}^{N(t)+1} r_i,
\]
and so with probability one $L_t(y)/t$ is sandwiched between two terms that both converge to $\mu_x(y)/E_x[T_x]$.

We still need to show that the limit is the same even if the Markov process doesn’t start at $X_0 = x$. To this end, suppose the Markov process starts at $X_0 = z \neq x$. Then define $\tau_1 = T_x$ to be the first visit to $x$ and as above $\tau_2 < \tau_3 < \ldots$ are the successive visits to $x$. In this case, it is still true that (3) holds (this is the law of large numbers for a delayed renewal process) and that (4) holds as well (this is because $r_1 < \infty$ and $r_2, r_3, \ldots$ are i.i.d.).

The final claim of the Theorem that the limit is equal to $1/(\lambda_y E_y[T_y])$ follows by choosing $x = y$ and noticing that $\mu_y(y) = 1/\lambda_y$. \hfill \Box

**Proof of Theorem 4.** Theorem 3 implies that $L_t(y)/t \to \frac{1}{\lambda_y E_y[T_y]}$ for any initial distribution of $X_0$. In particular, it holds if $X_0$ has an initial distribution $\pi$ that is stationary. Therefore, the bounded convergence theorem implies that
\[
\lim_{t \to \infty} E_\pi \left[ \frac{L_t(y)}{t} \right] = \frac{1}{\lambda_y E_y[T_y]}.
\]

On the other hand, the definition of $L_t(y)$ implies that
\[
E_\pi \left[ \frac{L_t(y)}{t} \right] = \frac{1}{t} E_\pi \left[ \int_0^t 1_{X_s = y} \, ds \right] = \frac{1}{t} \int_0^t E_{\pi}(X_s = y) \, ds = \frac{1}{t} \int_0^t \pi(y) = \pi(y),
\]
where we used that $\pi$ was a stationary distribution in the second to last equality. Thus, we have shown that any stationary distribution $\pi$ must satisfy $\pi(y) = \frac{1}{\lambda_y E_y[T_y]}$, and so $\pi$ is unique. \hfill \Box