Invited talk

for

Workshop celebrating Tony Pakes' 60th Birthday

by

Phil Pollett

The University of Queensland

ERGODICITY AND RECURRENCE

Pakes, A.G. (1969) Some conditions for ergodicity and recurrence of Markov chains. *Operat. Res.* 17, 1058–1061.

Let $(X_n, n = 0, 1, \dots)$ be an irreducible aperiodic Markov chain taking values in the non-negative integers and let

$$\gamma_i = \mathsf{E}(X_{n+1} - X_n | X_n = i).$$

Then, $\gamma_i \leq 0$ for all i sufficiently large is enough to guarantee recurrence, while $|\gamma_i| < \infty$ and $\limsup_{i \to \infty} \gamma_i < 0$ is sufficient for ergodicity.

This result has been used by many authors in a variety of contexts, for example, in the control of random access broadcast channels: slotted Aloha and CSMA/CD (Carrier sense multiple access with collision detect) protocol.

The Aloha Scheme

The following description is based on (Kelly, 1985)*.

Several stations use the same channel (assume infinitely many stations). Packets arrive for transmission as a Poisson stream with rate ν (< 1). Time is broken down into "slots" $(0, 1]$, $(1, 2]$, ... Let Y_t be the number of packets to arrive in the slot $(t-1, t]$ $(\mathrm{E}(Y_t) = \nu)$. Their transmission will first be attempted in the next slot $(t, t+1]$. Let Z_t represent the output of the channel at time t:

$$Z_t = \begin{cases} 0 & \text{if 0 transmissions attempted} \\ 1 & \text{if 1 transmission attempted} \\ * & \text{if } > 1 \text{ transmissions attempted} \end{cases}$$

*Kelly, F.P. (1985) Stochastic models of computer communication systems. *J. Royal Stat. Soc., Ser. B* 47, 379–395 (with discussion, 415–428).

If $Z_t = *$, a "collision" has occurred, and retransmission will be attempted in later slots, independently in each slot with probability f until successful. Thus, the transmission delay (measured in slots) has a geometric distribution with parameter $1 - f$.

The backlog (N_t) is a Markov chain with

$$N_{t+1} = N_t + Y_t - I[Z_t = 1].$$

Thus,

$$\begin{aligned}\gamma_n &:= \mathsf{E}(N_{t+1} - N_t | N_t = n) \\ &= \nu - \Pr(Z_t = 1 | N_t = n)\end{aligned}$$

and

$$\begin{aligned}&\Pr(Z_t = 1 | N_t = n) \\ &\qquad = e^{-\nu} n f (1 - f)^{n-1} + \nu e^{-\nu} (1 - f)^n.\end{aligned}$$

We deduce that $\gamma_n > 0$ for all n sufficiently large. Indeed the chain is *transient* (Kleinrock (1983), Fayolle, Gelenbe and Labetoulle (1977), Rosenkrantz and Towsley (1983)).

State-dependent Retransmission

Now suppose that the retransmission probability is allowed to depend on the backlog: $f = f_n$ when $N_t = n$. Then, $\Pr(Z_t = 1 | N_t = n)$ is maximized by

$$f_n = \frac{1-\nu}{n-\nu},$$

and, with this choice,

$$\begin{aligned} \gamma_n &:= \mathsf{E}(N_{t+1} - N_t | N_t = n, f = f_n) \\ &= \nu - e^{-\nu} \left(\frac{n-1}{n-\nu} \right)^{n-1}. \end{aligned}$$

Thus, $|\gamma_n| < \infty$ and $\gamma_n \to \nu - e^{-1}$. Thus, (N_t) is ergodic, that is, *the backlog is eventually cleared*, if $\nu < e^{-1} \simeq 0.368$.

But, users of the channel do not know the backlog, and thus cannot determine the optimal retransmission probability.

Towards a Better Control Scheme

It would be better to choose the retransmission probability $f_t = f(Z_1, Z_2, \ldots, Z_{t-1})$ based on the observed channel output. Several schemes have been suggested by Mikhailov (1979) and Hajek and van Loon (1982). For example, suppose each station maintains a counter S_t, updated as follows: $S_0 = 1$ and

$$\begin{aligned} S_{t+1} = \max\{1, S_t + aI[Z_t = 0] \\ + bI[Z_t = 1] + cI[Z_t = *]\}, \end{aligned}$$

where a, b and c are to be specified. For example, $(a, b, c) = (-1, 0, 1)$ is an obvious choice. Suppose that $f_t = 1/S_t$. Then, (N_t, S_t) is a Markov chain. We would like S_t to "track" the backlog, at least when N_t is large. Consider the drift in (S_t):

$$\begin{aligned} \phi_{n,s} &:= \mathsf{E}(S_{t+1} - S_t | N_t = n, S_t = s) \\ &= (a - c)\left(1 - \frac{1}{s}\right)^n + (b - c)\frac{n}{s}\left(1 - \frac{1}{s}\right)^n + c. \end{aligned}$$

Let $n \to \infty$ with $\kappa = n/s$ held fixed. Then,

$$\phi_{n,s} \to (a-c)e^{-\kappa} + (b-c)\kappa e^{-\kappa}.$$

The choice $(a,b,c) = ((2-e)\alpha, 0, \alpha)$, where $\alpha > 0$, makes the drift in (S_t) negative if $\kappa < 1$ and positive if $\kappa > 1$. Thus, if the backlog were held steady at a large value, then the counter would approach that value. Also,

$$\begin{aligned}\gamma_{n,s} &:= \mathsf{E}(N_{t+1} - N_t | N_t = n, S_t = s) \\ &= \nu - \frac{n}{s}\left(1 - \frac{1}{s}\right)^{n-1} \to \nu - \kappa e^{-\kappa}.\end{aligned}$$

Mikhailov (1979) showed that the choice (a,b,c) $= (2-e, 0, 1)$ ensures that (N_t, S_t) is ergodic whenever $\nu < e^{-1}$.

Question. For an irreducible aperiodic Markov chain (N_t, S_t), can one infer anything about its ergodicity and recurrence from the marginal drifts?

THE BIRTH-DEATH AND CATASTROPHE PROCESS

Pakes, A.G. (1987) Limit theorems for the population size of a birth and death process allowing catastrophes. *J. Math. Biol.* 25, 307–325.

An appropriate model for populations that are subject to crashes (dramatic losses can occur in animal populations due to disease, food shortages, significant changes in climate).

Such populations can exhibit *quasi-stationary behaviour*: they may survive for long periods before extinction occurs and can settle down to an apparently stationary regime. This behaviour can be modelled using a *limiting conditional (or quasi-stationary) distribution*.

The Model

It is a continuous-time Markov chain $(X(t), t \geq 0)$, where $X(t)$ represents the population size at time t, with transition rates $(q_{jk}, j, k \geq 0)$ given by

$$\begin{aligned} q_{j,j+1} &= j\rho a, && j \geq 0, \\ q_{j,j} &= -j\rho, && j \geq 0, \\ q_{j,j-i} &= j\rho b_i, && j \geq 2, 1 \leq i < j, \\ q_{j,0} &= j\rho \textstyle\sum_{i \geq j} b_i, && j \geq 1, \end{aligned}$$

with the other transition rates equal to 0. Here, $\rho > 0$, $a > 0$ and $b_i > 0$ for at least one i in $C = \{1, 2, \ldots\}$, and, $a + \sum_{i \geq 1} b_i = 1$.

Interpretation. For $j \neq k$, q_{jk} is the instantaneous rate at which the population size changes from j to k, ρ is the per capita rate of change and, given a change occurs, a is the probability that this results in a birth and b_i is the probability that this results in a catastrophe of size i (corresponding to the death or emigration of i individuals).

Some Properties

The state space. Clearly 0 is an absorbing state (corresponding to population extinction) and C is an irreducible class.

Extinction probabilities. If α_i is the probability of extinction starting with i individuals, then $\alpha_i = 1$ for all $i \in C$ if and only if D (the expected increment size), given by

$$D := a - \sum_{i \geq 1} i b_i = 1 - \sum_{i \geq 1} (i+1) b_i,$$

is less than 0 (the *subcritical* case) or equal to 0 (the *critical* case).

In the *supercritical* case ($D > 0$), the extinction probabilities can be expressed in terms of the probability generating function

$$f(s) = a + \sum_{i \geq 1} b_i s^{i+1}, \quad |s| < 1.$$

We find that

$$\sum_{i \geq 1} \alpha_i s^i = s/(1-s) - Ds/b(s),$$

where $b(s) = f(s) - s$.

Limiting Conditional Distributions

In order to describe the long-term behaviour of the process, we use two types of *limiting conditional distribution* (LCD), called Type I and Type II, corresponding to the limits:

$$\lim_{t\to\infty} \Pr(X(t) = j | X(0) = i, X(t) > 0,$$
$$X(t+r) = 0 \text{ for some } r > 0),$$

$$\lim_{t\to\infty} \lim_{s\to\infty} \Pr(X(t) = j | X(0) = i, X(t+s) > 0,$$
$$X(t+s+r) = 0 \text{ for some } r > 0),$$

where $i, j \in C$. Thus, we seek the limiting probability that the population size is j, given that extinction has not occurred, or (in the second case) will not occur in the distant future, but that eventually it *will* occur; we have conditioned on eventual extinction to deal with the supercritical case, where this event has probability less than 1.

The Existence of Limiting Conditional Distributions*

Consider the two eigenvector equations

$$\sum_{i\in C} m_i q_{ij} = -\mu m_j, \quad j \in C,$$
$$\sum_{j\in C} q_{ij} x_j = -\mu x_i, \quad i \in C,$$

where $\mu \geq 0$ and C is the irreducible class.

In order that both types of LCD exist, it is *necessary* that these equations have strictly positive solutions for some $\mu > 0$, these being the positive left and right eigenvectors of Q_C (the transition-rate matrix restricted to C) corresponding to a strictly negative eigenvalue $-\mu$.

Let λ be the *maximum* value of μ for which positive eigenvectors exist (λ is known to be finite), and denote the corresponding eigenvectors by $m = (m_j, j \in C)$ and $x = (x_j, j \in C)$.

*PKP technology

The Existence of Limiting Conditional Distributions

Proposition.* Suppose that Q is regular.

(i) If $\sum m_k x_k$ converges, and either $\sum m_k$ converges or $\{x_k\}$ is bounded, then the Type II LCD exists and defines a proper probability distribution $\pi^{(2)} = (\pi_j^{(2)}, j \in C)$ over C, given by

$$\pi_j^{(2)} = \frac{m_j x_j}{\sum m_k x_k}, \qquad j \in C.$$

(All unmarked sums are over k in C.)

(ii) If *in addition* $\sum m_k \alpha_k$ converges, then the Type I LCD exists and defines a proper probability distribution $\pi^{(1)} = (\pi_j^{(1)},\ j \in C)$ over C, given by

$$\pi_j^{(1)} = \frac{m_j \alpha_j}{\sum m_k \alpha_k}, \qquad j \in C.$$

*Pollett, P. (1988) Reversibility, invariance and μ-invariance. *Adv. Appl. Probab.* 20, 600–621.

Try to use PKP Technology

We need the fact that $b(s) = 0$ has a unique solution σ on $[0, 1]$, and that $\sigma = 1$ or $0 < \sigma < 1$ according as $D \geq 0$ or $D < 0$.

Setting $x_0 = m_0 = 0$, the eigenvector equations can be written (for $j \in C$) as

$$(j-1)\rho a m_{j-1} + \sum_{k=j+1}^{\infty} k\rho b_{k-j} m_k = (j\rho - \mu) m_j,$$

$$j\rho a x_{j+1} + \sum_{k=0}^{j} j\rho b_{j-k} x_k = (j\rho - \mu) x_j.$$

What is the maximum value of μ for which a positive solution exists? If $x = (x_j, j \in C)$ is *any* solution to the second, then its generating function $X(s) = \sum x_j s^j$ satisfies

$$X(s) = \frac{s}{b(s)} \exp(-\mu B(s)), \qquad s < \sigma,$$

where, for $s < \sigma$, $B(s) = \rho^{-1} \int_0^s dy/b(y)$.

Using this approach, we cannot really avoid the question: when is $X(s)$ a power series with *non-negative* coefficients? The function $C(s) = (\mu/\rho)\sum(x_j/j)s^j$ satisfies

$$C(s) = 1 - \exp(-\mu B(s)).$$

So, equivalently, we ask: when does $C(s)$ have non-negative coefficients?

This is answered in the following paper (assuming, as we have here, that $B(s)$ is a power series with non-negative coefficients):

> Pakes, A.G. (1997) On the recognition and structure of probability generating functions. In (Eds. K.B. Athreya and P. Jagers) *Classical and Modern Branching Processes*, IMA Vols. Math. Appl. 84, Springer, New York, pp. 263–284.

Lemon. The maximum value of μ for which a positive right eigenvector exists is $\lambda = -\rho b'(\sigma-)$. When $\mu = \lambda$, the left eigenvector is given by $m_j = \sigma^j$, $j \in C$.

The Subcritical Case

We have $D := -b'(1-) < 0$ and $\sigma < 1$. Since $m_j = \sigma^j$, $j \in C$, we have also $\sum m_k < \infty$ and $\sum m_k x_k = X(\sigma-) < \infty$.

The combination of technologies thus yields:

Theorem. In the *subcritical case* both types of LCD exist. The Type I LCD is given by

$$\pi_j^{(1)} = (1-\sigma)\sigma^{j-1},$$

and the Type II LCD has pgf

$$\Pi^{(2)}(s) = X(\sigma s)/X(\sigma-),$$

where

$$X(s) = \frac{s}{b(s)} \exp(-\lambda B(s)), \qquad s < \sigma,$$

and, for $s < \sigma$, $B(s) = \rho^{-1} \int_0^s dy/b(y)$.

This result is contained in Theorems 5.1 and 6.2 of Pakes (1987).

The Supercritical Case

We have $D > 0$ and $\sigma = 1$, and the absorption probabilities have generating function

$$\sum_{i\geq 1} \alpha_i s^i = s/(1-s) - Ds/b(s).$$

Since $m_j = 1$, $j \in C$, we have $\sum m_k \alpha_k = \sum \alpha_k$ and $\sum m_k x_k = X(1-)$. When do these series converge?

Condition (A). The catastrophe-size distribution has *finite second moment*, that is, $f''(1-) < \infty$ (equivalently $b''(1-) < \infty$).

Condition (B). The function b can be written

$$b(s) = D(1-s) + (1-s)^2 L((1-s)^{-1}),$$

where L is *slowly varying*, that is, $L(xt) \sim L(x)$ for large t.

Theorem. In the *supercritical case*, the Type I LCD exists under (A), and is given by $\pi_j^{(1)} = \alpha_j / \sum \alpha_k$. If in addition (B) holds, then the Type II LCD exists and has pgf $\Pi^{(2)}(s) = X(s)/X(1-)$.

This first part (Type I LCD) is contained in:

Pakes, A.G. and Pollett, P.K. (1989) The supercritical birth, death and catastrophe process: limit theorems on the set of extinction. *Stochastic Process. Appl.* 32, 161–170.

The second part (Type II LCD) is contained in Theorem 6.2 of Pakes (1987).

Other papers important to my work:

Pakes, A.G. (1971) A branching process with a state dependent immigration component. *Adv. Appl. Probab.* 3, 301–314.

Pakes, A.G. (1975) On the tails of waiting-time distributions. *J. Appl. Probab.* 12, 555–564.

Pakes, A.G. (1992) Divergence rates for explosive birth processes. *Stochastic Process. Appl.* 41, 91–99.

Pakes, A.G. (1993) Explosive Markov branching processes: entrance laws and limiting behaviour. *Adv. Appl. Probab.* 25, 737–756.

Pakes, A.G. (1993) Absorbing Markov and branching processes with instantaneous resurrection. *Stochastic Process. Appl.* 48, 85–106.

Pakes, A.G. (1995) Quasi-stationary laws for Markov processes: examples of an always proximate absorbing state. *Adv. Appl. Probab.* 27, 120–145.