On Hitting Times and Fastest Strong Stationary Times for Birth-and-death and Other Skip-Free Chains

Jim Fill

Department of Applied Mathematics and Statistics
The Johns Hopkins University

March 26, 2008

Hitting times for continuous-time skip-free chains

Theorem (Brown and Shao (1987), slightly extended)

- *state space* $= \{0, \ldots, d\}$; *continuous time*
- $X =$ *upward-skip-free chain with generator* G; $X(0) = 0$
- *Assume* $g_{i,i+1} > 0$ *for* $i < d$ *and* d *is absorbing.*
- *Let* $\nu_0, \ldots, \nu_{d-1}$ *be the* d *nonzero eigenvalues of* $-G$ *(known to have positive real parts).*
- *Then the hitting time* T *for state* d *has Laplace transform*

$$Ee^{-uT} = \prod_{j=0}^{d-1} \frac{\nu_j}{\nu_j + u}.$$

- *In particular, if the* ν_j*'s are real, then the hitting time distribution is the convolution of Exponential*(ν_j) *distributions.*

Hitting times for continuous-time skip-free chains

Theorem (Brown and Shao (1987), slightly extended)

- *state space* $= \{0, \ldots, d\}$*; continuous time*
- $X =$ *upward-skip-free chain with generator* G*;* $X(0) = 0$
- *Assume* $g_{i,i+1} > 0$ *for* $i < d$ *and* d *is absorbing.*
- *Let* $\nu_0, \ldots, \nu_{d-1}$ *be the* d *nonzero eigenvalues of* $-G$ *(known to have positive real parts).*
- *Then the hitting time* T *for state* d *has Laplace transform*

$$Ee^{-uT} = \prod_{j=0}^{d-1} \frac{\nu_j}{\nu_j + u}.$$

- *In particular, if the* ν_j*'s are real, then the hitting time distribution is the convolution of Exponential(*ν_j*) distributions.*

Hitting times for continuous-time birth-and-death chains

Theorem (Karlin and McGregor (1959), often attrib. to Keilson)

- *state space* $= \{0, \ldots, d\}$; *continuous time*
- $X =$ *B&D chain with generator* G; $X(0) = 0$
- *Assume* $\lambda_i := g_{i,i+1} > 0$ *for* $i < d$ *and* d *is absorbing.*
- *Let* $\nu_0, \ldots, \nu_{d-1}$ *be the* d *nonzero eigenvalues of* $-G$ *(known to be distinct and positive real).*
- *Then the distribution of the hitting time* T *for state* d *is the convolution of Exponential*(ν_j) *distributions.*

I will give

- another proof for skip-free chains [previous proofs: analytic]
- for B&D: a new simple explicit representation of T as a sum of independent Exp rv's. [cf. Diaconis & Miclo (2007)].

Hitting times for continuous-time birth-and-death chains

Theorem (Karlin and McGregor (1959), often attrib. to Keilson)

- *state space* $= \{0, \ldots, d\}$; *continuous time*
- $X =$ *B&D chain with generator* G; $X(0) = 0$
- *Assume* $\lambda_i := g_{i,i+1} > 0$ *for* $i < d$ *and* d *is absorbing.*
- *Let* $\nu_0, \ldots, \nu_{d-1}$ *be the* d *nonzero eigenvalues of* $-G$ *(known to be distinct and positive real).*
- *Then the distribution of the hitting time* T *for state* d *is the convolution of Exponential*(ν_j) *distributions.*

I will give

- another proof for skip-free chains [previous proofs: analytic]
- for B&D: a new simple explicit representation of T as a sum of independent Exp rv's. [cf. Diaconis & Miclo (2007)].

Hitting times for continuous-time birth-and-death chains

Theorem (Karlin and McGregor (1959), often attrib. to Keilson)

- *state space* $= \{0, \ldots, d\}$; *continuous time*
- $X =$ *B&D chain with generator* G; $X(0) = 0$
- *Assume* $\lambda_i := g_{i,i+1} > 0$ *for* $i < d$ *and* d *is absorbing.*
- *Let* $\nu_0, \ldots, \nu_{d-1}$ *be the* d *nonzero eigenvalues of* $-G$ *(known to be distinct and positive real).*
- *Then the distribution of the hitting time* T *for state* d *is the convolution of Exponential*(ν_j) *distributions.*

I will give

- another proof for skip-free chains [previous proofs: analytic]
- for B&D: a new simple explicit representation of T as a sum of independent Exp rv's. [cf. Diaconis & Miclo (2007)].

Fastest SSTs for continuous-time skip-free chains

Similar result (and proof) for fastest strong stationary times:

Theorem

- *state space* $= \{0, \ldots, d\}$; *continuous time*
- X = *ergodic upward-skip-free MC with generator* G; $X(0) = 0$
- *Assume* X *has stoch. monotone time-reversal* [true for B&D].
- *Let* $\nu_0, \ldots, \nu_{d-1}$ *be the* d *nonzero eigenvalues of* $-G$ *(known to have positive real parts).*
- *Then the hitting time* T *for state* d *has Laplace transform*

$$Ee^{-uT} = \prod_{j=0}^{d-1} \frac{\nu_j}{\nu_j + u}.$$

Fastest SSTs for continuous-time skip-free chains

Similar result (and proof) for fastest strong stationary times:

Theorem

- *state space* $= \{0, \ldots, d\}$; *continuous time*
- $X =$ *ergodic upward-skip-free MC with generator* G; $X(0) = 0$
- *Assume* X *has stoch. monotone time-reversal* [true for B&D].
- *Let* $\nu_0, \ldots, \nu_{d-1}$ *be the* d *nonzero eigenvalues of* $-G$ *(known to have positive real parts).*
- *Then the hitting time* T *for state* d *has Laplace transform*

$$Ee^{-uT} = \prod_{j=0}^{d-1} \frac{\nu_j}{\nu_j + u}.$$

Hitting times for discrete-time skip-free chains

Theorem

- *state space* $= \{0, \dots, d\}$*; discrete time*
- X = *upward-skip-free chain with kernel* P*;* $X_0 = 0$
- *Assume* $p_{i,i+1} > 0$ *for* $i < d$ *and* d *is absorbing.*
- *Let* $\theta_0, \dots, \theta_{d-1}$ *be the* d *non-unit eigenvalues of* P*.*
- *Then the hitting time* T *for state* d *has probability generating function*

$$Eu^T = \prod_{j=0}^{d-1} \frac{(1-\theta_j)u}{1-\theta_j u}.$$

- *In particular, if the* θ_j*'s are real* [true for B&D] *and nonnegative, then the hitting time distribution is the convolution of Geometric*$(1-\theta_j)$ *distributions.*

Hitting times for discrete-time skip-free chains

Theorem

- *state space* $= \{0, \ldots, d\}$; *discrete time*
- X = *upward-skip-free chain with kernel* P; $X_0 = 0$
- *Assume* $p_{i,i+1} > 0$ *for* $i < d$ *and* d *is absorbing.*
- *Let* $\theta_0, \ldots, \theta_{d-1}$ *be the* d *non-unit eigenvalues of* P*.*
- *Then the hitting time* T *for state* d *has probability generating function*

$$Eu^T = \prod_{j=0}^{d-1} \frac{(1-\theta_j)u}{1-\theta_j u}.$$

- *In particular, if the* θ_j*'s are real* [true for B&D] *and nonnegative, then the hitting time distribution is the convolution of Geometric*$(1-\theta_j)$ *distributions.*

Motivation and notes

- Discrete-time results are sufficient (and easier to discuss!); continuous-time results follow easily. [Use similar proofs, or consider $P(\varepsilon) := I + \varepsilon G$ for sufficiently small ε.]
- Even for hitting times of B&D chains, the problem of giving a stochastic proof and interpreting the individual Exponential random variables was open for some time.
- For B&D chains, the above theorems are the starting point of an in-depth consideration of the cut-off phenomenon in separation by Diaconis and Saloff-Coste (2006).
- An in-depth consideration of the cut-off phenomenon in total variation distance for B&D chains has been carried out by Ding, Lubetzky, and Peres (2008).

Come to Eyal Lubetzky's talk after lunch!

Motivation and notes

- Discrete-time results are sufficient (and easier to discuss!); continuous-time results follow easily. [Use similar proofs, or consider $P(\varepsilon) := I + \varepsilon G$ for sufficiently small ε.]
- Even for hitting times of B&D chains, the problem of giving a stochastic proof and interpreting the individual Exponential random variables was open for some time.
- For B&D chains, the above theorems are the starting point of an in-depth consideration of the cut-off phenomenon in separation by Diaconis and Saloff-Coste (2006).
- An in-depth consideration of the cut-off phenomenon in total variation distance for B&D chains has been carried out by Ding, Lubetzky, and Peres (2008).

Come to Eyal Lubetzky's talk after lunch!

Motivation and notes

- Discrete-time results are sufficient (and easier to discuss!); continuous-time results follow easily. [Use similar proofs, or consider $P(\varepsilon) := I + \varepsilon G$ for sufficiently small ε.]
- Even for hitting times of B&D chains, the problem of giving a stochastic proof and interpreting the individual Exponential random variables was open for some time.
- For B&D chains, the above theorems are the starting point of an in-depth consideration of the cut-off phenomenon in separation by Diaconis and Saloff-Coste (2006).
- An in-depth consideration of the cut-off phenomenon in total variation distance for B&D chains has been carried out by Ding, Lubetzky, and Peres (2008).

Come to Eyal Lubetzky's talk after lunch!

Outline of proof

Focus for remainder of talk: theorem for hitting times for discrete-time skip-free chains. We outline the proof in two steps. To set up:

- Let P have eigenvalues $\theta_0, \ldots, \theta_d$.
- It's easy to check that one of these, say θ_d, equals 1 and that the others have modulus < 1.
- Order $\theta_0, \ldots, \theta_{d-1}$ arbitrarily.

"Intertwining"

1. "Intertwining": We will exhibit a (generally complex) square matrix Λ on state space $\{0, \dots, d\}$ such that
   (i) Λ is lower triangular.
   (ii) The rows of Λ sum to unity.
   (iii) $\Lambda P = \widehat{P}\Lambda$, where $\widehat{P}$ (generally complex) is defined by

$$\hat{p}_{ij} := \begin{cases} \theta_i & \text{if } j = i \\ 1 - \theta_i & \text{if } j = i + 1 \\ 0 & \text{otherwise,} \end{cases}$$

   for $i, j = 0, \dots, d$.

The "intertwining" $\Lambda P^t \equiv \widehat{P}^t \Lambda$ follows immediately.

"Intertwining"

1. "Intertwining": We will exhibit a (generally complex) square matrix Λ on state space $\{0, \ldots, d\}$ such that
   (i) Λ is lower triangular.
   (ii) The rows of Λ sum to unity.
   (iii) $\Lambda P = \widehat{P}\Lambda$, where $\widehat{P}$ (generally complex) is defined by

$$\hat{p}_{ij} := \begin{cases} \theta_i & \text{if } j = i \\ 1 - \theta_i & \text{if } j = i + 1 \\ 0 & \text{otherwise,} \end{cases}$$

   for $i, j = 0, \ldots, d$.

   The "intertwining" $\Lambda P^t \equiv \widehat{P}^t \Lambda$ follows immediately.

2. Distribution of T

2. Distribution of T:

- We will show from properties (i)–(iii) that

$$\mathbf{P}(T \le t) = \widehat{P}^t(0, d), \quad t = 0, 1, \ldots,$$

  where T is the absorption time in question.
- In the special case that the θ_j's are real and nonnegative, the theorem follows immediately.
- The general case won't take much additional work.

2. Distribution of T

2. Distribution of T:
   - We will show from properties (i)–(iii) that

$$\mathbf{P}(T \le t) = \widehat{P}^t(0, d), \quad t = 0, 1, \ldots,$$

     where T is the absorption time in question.
   - In the special case that the θ_j's are real and nonnegative, the theorem follows immediately.
   - The general case won't take much additional work.

2. Distribution of T

2. Distribution of T:
    - We will show from properties (i)–(iii) that

      $$\mathbf{P}(T \le t) = \widehat{P}^t(0, d), \quad t = 0, 1, \ldots,$$

      where T is the absorption time in question.
    - In the special case that the θ_j's are real and nonnegative, the theorem follows immediately.
    - The general case won't take much additional work.

1.(a) Definition of Λ

- Let I denote the identity matrix and define, for $k = 0, \ldots, d$,

$$Q_k := (1-\theta_0)^{-1} \cdots (1-\theta_{k-1})^{-1} (P - \theta_0 I) \cdots (P - \theta_{k-1} I)$$

  with the natural convention $Q_0 := I$.

- Note the recurrence relation

$$Q_k P = \theta_k Q_k + (1-\theta_k) Q_{k+1}, \quad k = 0, \ldots, d-1,$$

  and that the rows of each Q_k all sum to 1.

- Define

$$\Lambda(i,j) := Q_i(0,j), \quad i,j = 0, \ldots, d.$$

1.(a) Definition of Λ

- Let I denote the identity matrix and define, for $k = 0, \ldots, d$,

$$Q_k := (1-\theta_0)^{-1} \cdots (1-\theta_{k-1})^{-1} (P - \theta_0 I) \cdots (P - \theta_{k-1} I)$$

  with the natural convention $Q_0 := I$.

- Note the recurrence relation

$$Q_k P = \theta_k Q_k + (1-\theta_k) Q_{k+1}, \quad k = 0, \ldots, d-1,$$

  and that the rows of each Q_k all sum to 1.

- Define

$$\Lambda(i,j) := Q_i(0,j), \quad i,j = 0, \ldots, d.$$

1.(a) Definition of Λ

- Let I denote the identity matrix and define, for $k = 0, \ldots, d$,

$$Q_k := (1-\theta_0)^{-1} \cdots (1-\theta_{k-1})^{-1} (P - \theta_0 I) \cdots (P - \theta_{k-1} I)$$

  with the natural convention $Q_0 := I$.

- Note the recurrence relation

$$Q_k P = \theta_k Q_k + (1-\theta_k) Q_{k+1}, \quad k = 0, \ldots, d-1,$$

  and that the rows of each Q_k all sum to 1.

- Define

$$\Lambda(i,j) := Q_i(0,j), \quad i,j = 0, \ldots, d.$$

1.(b) Proof of properties (i)–(iii) for Λ

Proof of properties (i)–(iii) for Λ:

(ii) The rows of Λ sum to 1 because each of the basic factors $(1-\theta_r)^{-1}(P-\theta_r I)$ in the definition of the Q_k's has that property.

(i) Since P is skip-free, Λ is lower triangular.

(iii) Our claim is that $\Lambda P = \widehat{P}\Lambda$. Indeed, equality of kth rows is clear for $k < d$ from the recurrence relation for the $Q_\cdot$'s and for $k = d$ from the Cayley–Hamilton theorem.

1.(b) Proof of properties (i)–(iii) for Λ

Proof of properties (i)–(iii) for Λ:

(ii) The rows of Λ sum to 1 because each of the basic factors $(1 - \theta_r)^{-1}(P - \theta_r I)$ in the definition of the Q_k's has that property.

(i) Since P is skip-free, Λ is lower triangular.

(iii) Our claim is that $\Lambda P = \widehat{P}\Lambda$. Indeed, equality of kth rows is clear for $k < d$ from the recurrence relation for the $Q_\cdot$'s and for $k = d$ from the Cayley–Hamilton theorem.

1.(b) Proof of properties (i)–(iii) for Λ

Proof of properties (i)–(iii) for Λ:

(ii) The rows of Λ sum to 1 because each of the basic factors $(1-\theta_r)^{-1}(P-\theta_r I)$ in the definition of the Q_k's has that property.

(i) Since P is skip-free, Λ is lower triangular.

(iii) Our claim is that $\Lambda P = \widehat{P}\Lambda$. Indeed, equality of kth rows is clear for $k < d$ from the recurrence relation for the $Q_{\cdot}$'s and for $k = d$ from the Cayley–Hamilton theorem.

Remark

Remark:

- The lower triangular matrix Λ is nonsingular, because its diagonal entries are all nonzero: Since P is upward-skip-free, for $k = 0, \dots, d$ we have

$$\Lambda(k,k) = Q_k(0,k) = \frac{p_{0,1} \cdots p_{k-1,k}}{(1-\theta_0)\cdots(1-\theta_{k-1})} \neq 0.$$

- So $\Lambda P = \widehat{P}\Lambda$ implies that P and $\widehat{P}$ are similar. So it is no surprise that the eigenvalues of P line the diagonal of the upper triangular matrix $\widehat{P}$.

Remark

Remark:

- The lower triangular matrix Λ is nonsingular, because its diagonal entries are all nonzero: Since P is upward-skip-free, for $k = 0, \ldots, d$ we have

$$\Lambda(k,k) = Q_k(0,k) = \frac{p_{0,1} \cdots p_{k-1,k}}{(1-\theta_0)\cdots(1-\theta_{k-1})} \neq 0.$$

- So $\Lambda P = \widehat{P}\Lambda$ implies that P and $\widehat{P}$ are similar. So it is no surprise that the eigenvalues of P line the diagonal of the upper triangular matrix $\widehat{P}$.

2.(a) Distribution of T in terms of $\widehat{P}$

Proof that $\mathbf{P}(T \leq t) = \widehat{P}^t(0, d)$:

- We first claim that

$$\mathbf{P}(T \leq t) = \widehat{P}^t(0, d)\Lambda(d, d)$$

by comparing $(0, d)$-entries in $\Lambda P^t = \widehat{P}^t \Lambda$ and using lower triangularity of Λ [whence $\Lambda(0, 0) = 1$ because row 0 sums to 1]:

  - On the left:

$$(\Lambda P^t)(0, d) = \sum_k \Lambda(0, k) P^t(k, d) = P^t(0, d) = \mathbf{P}(T \leq t).$$

  - On the right:

$$(\widehat{P}^t \Lambda)(0, d) = \sum_k \widehat{P}^t(0, k)\Lambda(k, d) = \widehat{P}^t(0, d)\Lambda(d, d).$$

- Now we must show that $\Lambda(d, d) = 1$.

2.(a) Distribution of T in terms of $\widehat{P}$

Proof that $\mathbf{P}(T \le t) = \widehat{P}^t(0,d)$:

- We first claim that

$$\mathbf{P}(T \le t) = \widehat{P}^t(0,d)\Lambda(d,d)$$

  by comparing $(0,d)$-entries in $\Lambda P^t = \widehat{P}^t \Lambda$ and using lower triangularity of Λ [whence $\Lambda(0,0) = 1$ because row 0 sums to 1]:
  - On the left:

$$(\Lambda P^t)(0,d) = \sum_k \Lambda(0,k) P^t(k,d) = P^t(0,d) = \mathbf{P}(T \le t).$$

  - On the right:

$$(\widehat{P}^t \Lambda)(0,d) = \sum_k \widehat{P}^t(0,k)\Lambda(k,d) = \widehat{P}^t(0,d)\Lambda(d,d).$$

- Now we must show that $\Lambda(d,d) = 1$.

2.(a) Distribution of T in terms of $\widehat{P}$

Proof that $\mathbf{P}(T \leq t) = \widehat{P}^t(0, d)$:

- We first claim that

$$\mathbf{P}(T \leq t) = \widehat{P}^t(0, d)\Lambda(d, d)$$

  by comparing $(0, d)$-entries in $\Lambda P^t = \widehat{P}^t \Lambda$ and using lower triangularity of Λ [whence $\Lambda(0, 0) = 1$ because row 0 sums to 1]:

  - On the left:

$$(\Lambda P^t)(0, d) = \sum_k \Lambda(0, k) P^t(k, d) = P^t(0, d) = \mathbf{P}(T \leq t).$$

  - On the right:

$$(\widehat{P}^t \Lambda)(0, d) = \sum_k \widehat{P}^t(0, k)\Lambda(k, d) = \widehat{P}^t(0, d)\Lambda(d, d).$$

- Now we must show that $\Lambda(d, d) = 1$.

2.(a) Distribution of T in terms of $\widehat{P}$

Proof that $\mathbf{P}(T \le t) = \widehat{P}^t(0,d)$:

- We first claim that

$$\mathbf{P}(T \le t) = \widehat{P}^t(0,d)\Lambda(d,d)$$

  by comparing $(0,d)$-entries in $\Lambda P^t = \widehat{P}^t \Lambda$ and using lower triangularity of Λ [whence $\Lambda(0,0) = 1$ because row 0 sums to 1]:
  - On the left:

$$(\Lambda P^t)(0,d) = \sum_k \Lambda(0,k) P^t(k,d) = P^t(0,d) = \mathbf{P}(T \le t).$$

  - On the right:

$$(\widehat{P}^t \Lambda)(0,d) = \sum_k \widehat{P}^t(0,k)\Lambda(k,d) = \widehat{P}^t(0,d)\Lambda(d,d).$$

- Now we must show that $\Lambda(d,d) = 1$.

Proof that $\Lambda(d, d) = 1$

We know $\mathbf{P}(T \le t) = \widehat{P}^t(0, d)\Lambda(d, d)$ and need $\Lambda(d, d) = 1$.

- Limit as $t \to \infty$ of LHS equals 1.
- Limit as $t \to \infty$ of $\widehat{P}^t(0, d)$ in RHS equals 1, completing proof that $\Lambda(d, d) = 1$, if eigenvalues are all real and nonnegative.
- In fact, in general case $\widehat{P}^t(0, d) \to 1$ and so $\Lambda(d, d) = 1$. To see this, break off last row and column to write $\widehat{P} = \begin{bmatrix} A & b \\ 0 & 1 \end{bmatrix}$. The matrix A is upper triangular with spectral radius $\max\{|\theta_0|, \ldots, |\theta_{d-1}|\} < 1$, and $b = (I - A)\mathbf{1}$. Then

$$\widehat{P}^t = \begin{bmatrix} A^t & (I + \cdots A^{t-1})b \\ 0 & 1 \end{bmatrix} \to \begin{bmatrix} 0 & (I - A)^{-1}b \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & \mathbf{1} \\ 0 & 1 \end{bmatrix}.$$

In particular, $\widehat{P}^t(0, d) \to 1$.

Proof that $\Lambda(d, d) = 1$

We know $\mathbf{P}(T \le t) = \widehat{P}^t(0, d)\Lambda(d, d)$ and need $\Lambda(d, d) = 1$.

- Limit as $t \to \infty$ of LHS equals 1.
- Limit as $t \to \infty$ of $\widehat{P}^t(0, d)$ in RHS equals 1, completing proof that $\Lambda(d, d) = 1$, if eigenvalues are all real and nonnegative.
- In fact, in general case $\widehat{P}^t(0, d) \to 1$ and so $\Lambda(d, d) = 1$. To see this, break off last row and column to write $\widehat{P} = \begin{bmatrix} A & b \\ 0 & 1 \end{bmatrix}$. The matrix A is upper triangular with spectral radius $\max\{|\theta_0|, \dots, |\theta_{d-1}|\} < 1$, and $b = (I - A)\mathbf{1}$. Then

$$\widehat{P}^t = \begin{bmatrix} A^t & (I + \cdots A^{t-1})b \\ 0 & 1 \end{bmatrix} \to \begin{bmatrix} 0 & (I - A)^{-1}b \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & \mathbf{1} \\ 0 & 1 \end{bmatrix}.$$

In particular, $\widehat{P}^t(0, d) \to 1$.

Proof that $\Lambda(d, d) = 1$

We know $\mathbf{P}(T \le t) = \widehat{P}^t(0, d)\Lambda(d, d)$ and need $\Lambda(d, d) = 1$.

- Limit as $t \to \infty$ of LHS equals 1.
- Limit as $t \to \infty$ of $\widehat{P}^t(0, d)$ in RHS equals 1, completing proof that $\Lambda(d, d) = 1$, if eigenvalues are all real and nonnegative.
- In fact, in general case $\widehat{P}^t(0, d) \to 1$ and so $\Lambda(d, d) = 1$. To see this, break off last row and column to write $\widehat{P} = \begin{bmatrix} A & b \\ 0 & 1 \end{bmatrix}$. The matrix A is upper triangular with spectral radius $\max\{|\theta_0|, \dots, |\theta_{d-1}|\} < 1$, and $b = (I - A)\mathbf{1}$. Then

$$\widehat{P}^t = \begin{bmatrix} A^t & (I + \cdots A^{t-1})b \\ 0 & 1 \end{bmatrix} \to \begin{bmatrix} 0 & (I - A)^{-1}b \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & \mathbf{1} \\ 0 & 1 \end{bmatrix}.$$

In particular, $\widehat{P}^t(0, d) \to 1$.

Proof that $\Lambda(d, d) = 1$

We know $\mathbf{P}(T \le t) = \widehat{P}^t(0, d)\Lambda(d, d)$ and need $\Lambda(d, d) = 1$.

- Limit as $t \to \infty$ of LHS equals 1.
- Limit as $t \to \infty$ of $\widehat{P}^t(0, d)$ in RHS equals 1, completing proof that $\Lambda(d, d) = 1$, if eigenvalues are all real and nonnegative.
- In fact, in general case $\widehat{P}^t(0, d) \to 1$ and so $\Lambda(d, d) = 1$. To see this, break off last row and column to write $\widehat{P} = \begin{bmatrix} A & b \\ 0 & 1 \end{bmatrix}$. The matrix A is upper triangular with spectral radius $\max\{|\theta_0|, \dots, |\theta_{d-1}|\} < 1$, and $b = (I - A)\mathbf{1}$. Then

$$\widehat{P}^t = \begin{bmatrix} A^t & (I + \cdots A^{t-1})b \\ 0 & 1 \end{bmatrix} \to \begin{bmatrix} 0 & (I - A)^{-1}b \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & \mathbf{1} \\ 0 & 1 \end{bmatrix}.$$

In particular, $\widehat{P}^t(0, d) \to 1$.

2.(b) Pgf for T

We now know $(*)$ $\mathbf{P}(T \le t) = \widehat{P}^t(0, d)$ for $t = 0, 1, \ldots$.

- We now know that distribution of T is convolution of Geometrics, completing proof of theorem, if eigenvalues are all real and nonnegative.
- General finish to proof of theorem: By $(*)$,

$$\mathbf{E}\, u^T = (1-u)\,(I - u\widehat{P})^{-1}(0, d), \quad |u| < 1.$$

But it's easy to invert $1 - u\widehat{P}$ explicitly: the inverse is upper triangular, with

$$(I - u\widehat{P})^{-1}(i, j) = \frac{(1-\theta_i)\cdots(1-\theta_{j-1})u^{j-i}}{(1-\theta_i u)\cdots(1-\theta_j u)}, \quad 0 \le i \le j \le d.$$

Take $i = 0$ and $j = d$.

2.(b) Pgf for T

We now know $(*)$ $\mathbf{P}(T \le t) = \widehat{P}^t(0, d)$ for $t = 0, 1, \ldots$.

- We now know that distribution of T is convolution of Geometrics, completing proof of theorem, if eigenvalues are all real and nonnegative.
- General finish to proof of theorem: By $(*)$,

$$\mathbf{E}\, u^T = (1-u)\,(I - u\widehat{P})^{-1}(0, d), \quad |u| < 1.$$

But it's easy to invert $1 - u\widehat{P}$ explicitly: the inverse is upper triangular, with

$$(I - u\widehat{P})^{-1}(i, j) = \frac{(1-\theta_i)\cdots(1-\theta_{j-1})u^{j-i}}{(1-\theta_i u)\cdots(1-\theta_j u)}, \quad 0 \le i \le j \le d.$$

Take $i = 0$ and $j = d$.

Stochastic constructions for B&D chains

- Restrict attention to birth-and-death chains from now on. Stick with discrete time (for definiteness), for now.
- For a B&D chain as in our hitting-time theorem, the eigenvalues θ_j of the kernel P are all real. To see this, perturb (by arbitrarily small amount) to get an ergodic kernel, which is time-reversible and thus diagonally similar to a symmetric matrix.
- Henceforth suppose that the eigenvalues are all nonnegative (for which $p_{ii} \geq 1/2$ for all i is sufficient). We now know that the absorption time T is distributed as the convolution of Geometric$(1 - \theta_j)$ distributions.

Spectral polynomials

- Order the eigenvalues θ_j so that

$$0 \le \theta_0 \le \cdots \le \theta_{d-1} < \theta_d = 1.$$

- The polynomials

$$(P - \theta_0 I) \cdots (P - \theta_{k-1} I)$$

in P used to define the respective Q_k's (modulo scalar factors) are called *spectral polynomials*.

- Claim: The spectral polynomials are all nonnegative matrices! Then the Q_k's are stochastic, and hence so is the matrix Λ defined (we recall) by

$$\Lambda(i,j) := Q_i(0,j), \quad i,j = 0, \ldots, d.$$

Spectral polynomials

- Order the eigenvalues θ_j so that

$$0 \leq \theta_0 \leq \cdots \leq \theta_{d-1} < \theta_d = 1.$$

- The polynomials

$$(P - \theta_0 I) \cdots (P - \theta_{k-1} I)$$

in P used to define the respective Q_k's (modulo scalar factors) are called *spectral polynomials*.

- Claim: The spectral polynomials are all nonnegative matrices! Then the Q_k's are stochastic, and hence so is the matrix Λ defined (we recall) by

$$\Lambda(i, j) := Q_i(0, j), \quad i, j = 0, \ldots, d.$$

Spectral polynomials

- Order the eigenvalues θ_j so that

$$0 \le \theta_0 \le \cdots \le \theta_{d-1} < \theta_d = 1.$$

- The polynomials

$$(P - \theta_0 I) \cdots (P - \theta_{k-1} I)$$

  in P used to define the respective Q_k's (modulo scalar factors) are called *spectral polynomials*.
- Claim: The spectral polynomials are all nonnegative matrices! Then the Q_k's are stochastic, and hence so is the matrix Λ defined (we recall) by

$$\Lambda(i,j) := Q_i(0,j), \quad i,j = 0,\ldots,d.$$

Nonnegativity of spectral polynomials

Proof of claim that the spectral polynomials $(P - \theta_0 I) \cdots (P - \theta_{k-1} I)$ are nonnegative matrices: It's true

1. for nonnegative symmetric matrices, by (rather nontrivial!) theorem of Micchelli and Willoughby (1979, *Linear Algebra and its Applications*);
2. for ergodic B&D kernels, by positive diagonal similarity to a nonnegative symmetric matrix;
3. for absorbing B&D kernels as in our hitting-time theorem, by perturbation.

Nonnegativity of spectral polynomials

Proof of claim that the spectral polynomials $(P - \theta_0 I) \cdots (P - \theta_{k-1} I)$ are nonnegative matrices: It's true

1. for nonnegative symmetric matrices, by (rather nontrivial!) theorem of Micchelli and Willoughby (1979, *Linear Algebra and its Applications*);
2. for ergodic B&D kernels, by positive diagonal similarity to a nonnegative symmetric matrix;
3. for absorbing B&D kernels as in our hitting-time theorem, by perturbation.

Nonnegativity of spectral polynomials

Proof of claim that the spectral polynomials $(P - \theta_0 I) \cdots (P - \theta_{k-1} I)$ are nonnegative matrices: It's true

1. for nonnegative symmetric matrices, by (rather nontrivial!) theorem of Micchelli and Willoughby (1979, *Linear Algebra and its Applications*);
2. for ergodic B&D kernels, by positive diagonal similarity to a nonnegative symmetric matrix;
3. for absorbing B&D kernels as in our hitting-time theorem, by perturbation.

Nonnegativity of spectral polynomials

Proof of claim that the spectral polynomials $(P - \theta_0 I) \cdots (P - \theta_{k-1} I)$ are nonnegative matrices: It's true

1. for nonnegative symmetric matrices, by (rather nontrivial!) theorem of Micchelli and Willoughby (1979, *Linear Algebra and its Applications*);
2. for ergodic B&D kernels, by positive diagonal similarity to a nonnegative symmetric matrix;
3. for absorbing B&D kernels as in our hitting-time theorem, by perturbation.

Intertwining

Recap: For a B&D chain with nonnegative eigenvalues, the matrices P, $\widehat{P}$, and Λ are all stochastic, and we have the identity

$$\Lambda P^t = \widehat{P}^t \Lambda, \quad t \geq 0.$$

- One says: "The semigroups $(P^t)_{t\geq 0}$ and $(\widehat{P}^t)_{t\geq 0}$ are intertwined by the link Λ."
- *Whenever* we have such an intertwining [with $\Lambda(0,\cdot) = \delta_0$], Section 2.4 of the strong stationary duality paper

  Diaconis, P. and Fill, J. A. Strong stationary times via a new form of duality. *Ann. Probab.* **18** (1990), 1483–1522

  shows (more than) one way to construct explicitly, from X and independent randomness, another Markov chain $\widehat{X}$ with kernel $\widehat{P}$ such that

$$\text{distn}(X_t \mid \widehat{X}_0, \ldots, \widehat{X}_t) = \Lambda(\widehat{X}_t, \cdot) \quad \text{for all } t.$$

Intertwining

Recap: For a B&D chain with nonnegative eigenvalues, the matrices P, $\widehat{P}$, and Λ are all stochastic, and we have the identity

$$\Lambda P^t = \widehat{P}^t \Lambda, \quad t \geq 0.$$

- One says: "The semigroups $(P^t)_{t \geq 0}$ and $(\widehat{P}^t)_{t \geq 0}$ are intertwined by the link Λ."
- *Whenever* we have such an intertwining [with $\Lambda(0, \cdot) = \delta_0$], Section 2.4 of the strong stationary duality paper

  Diaconis, P. and Fill, J. A. Strong stationary times via a new form of duality. *Ann. Probab.* **18** (1990), 1483–1522

  shows (more than) one way to construct explicitly, from X and independent randomness, another Markov chain $\widehat{X}$ with kernel $\widehat{P}$ such that

$$\text{distn}(X_t \mid \widehat{X}_0, \ldots, \widehat{X}_t) = \Lambda(\widehat{X}_t, \cdot) \quad \text{for all } t.$$

The chain $\widehat{X}$ and Geometrics

- We can construct explicitly, from X and independent randomness, another Markov chain $\widehat{X}$ with kernel $\widehat{P}$ such that

$$\text{distn}(X_t \mid \widehat{X}_0, \ldots, \widehat{X}_t) = \Lambda(\widehat{X}_t, \cdot) \quad \text{for all } t.$$

- In particular, since the link Λ is lower triangular [property (i)] and $\Lambda(d, d) = 1$ [as we have seen], it follows that

$$\widehat{X} \text{ and } X \text{ have the same absorption time } T.$$

- Once we have built the pure birth chain $\widehat{X}$, the independent Geometric random variables summing to T are simply the waiting times between successive births in $\widehat{X}$.

The chain $\widehat{X}$ and Geometrics

- We can construct explicitly, from X and independent randomness, another Markov chain $\widehat{X}$ with kernel $\widehat{P}$ such that

$$\text{distn}(X_t \mid \widehat{X}_0, \ldots, \widehat{X}_t) = \Lambda(\widehat{X}_t, \cdot) \quad \text{for all } t.$$

- In particular, since the link Λ is lower triangular [property (i)] and $\Lambda(d, d) = 1$ [as we have seen], it follows that

$\widehat{X}$ and X have the same absorption time T.

- Once we have built the pure birth chain $\widehat{X}$, the independent Geometric random variables summing to T are simply the waiting times between successive births in $\widehat{X}$.

The chain $\widehat{X}$ and Geometrics

- We can construct explicitly, from X and independent randomness, another Markov chain $\widehat{X}$ with kernel $\widehat{P}$ such that

$$\text{distn}(X_t \mid \widehat{X}_0, \ldots, \widehat{X}_t) = \Lambda(\widehat{X}_t, \cdot) \quad \text{for all } t.$$

- In particular, since the link Λ is lower triangular [property (i)] and $\Lambda(d, d) = 1$ [as we have seen], it follows that

$\widehat{X}$ and X have the same absorption time T.

- Once we have built the pure birth chain $\widehat{X}$, the independent Geometric random variables summing to T are simply the waiting times between successive births in $\widehat{X}$.

Remark: stochastic maximality

- We can construct explicitly, from X and independent randomness, another Markov chain $\widehat{X}$ with kernel $\widehat{P}$ such that

$$\text{distn}(X_t \,|\, \widehat{X}_0, \ldots, \widehat{X}_t) = \Lambda(\widehat{X}_t, \cdot) \quad \text{for all } t.$$

- By the lower-triangularity of Λ, our construction satisfies $X_t \leq \widehat{X}_t$ for all t. Thus, among all discrete-time B&D chains on $\{0, \ldots, d\}$ started at 0 and with absorbing state d and given nonnegative eigenvalues

$$0 \leq \theta_0 \leq \theta_1 \leq \cdots \leq \theta_{d-1} < \theta_d = 1,$$

the pure-birth "spectral" kernel $\widehat{P}$ is stochastically maximal at every epoch t.

One way to construct $\widehat{X}$

One way to construct $\widehat{X}$ [from Diaconis and Fill (1990)]:

- The chain X starts with $X_0 = 0$ and we set $\widehat{X}_0 = 0$.
- Inductively, we will have $\Lambda(\widehat{X}_t, X_t) > 0$ (and so $X_t \le \widehat{X}_t$) at all times t. In particular, $X_t \le X_{t-1} + 1 \le \widehat{X}_{t-1} + 1$.
- The value we construct for $\widehat{X}_t$ depends only on the values $\widehat{X}_{t-1} = \hat{x}$ and $X_t = y$ (with $y \le \hat{x} + 1$) and independent randomness. There are two cases.
- *Case 1:* If $y \le \hat{x}$ then set $\widehat{X}_t = \hat{x} + 1$ with probability
$$\frac{\widehat{P}(\hat{x}, \hat{x}+1)\Lambda(\hat{x}+1, y)}{(\widehat{P}\Lambda)(\hat{x}, y)} = \frac{(1-\theta_{\hat{x}})\Lambda(\hat{x}+1, y)}{\theta_{\hat{x}}\Lambda(\hat{x}, y) + (1-\theta_{\hat{x}})\Lambda(\hat{x}+1, y)}$$
and $\widehat{X}_t = \hat{x}$ with the complementary probability.
- *Case 2:* If $y = \hat{x} + 1$, then we set $\widehat{X}_t = \hat{x} + 1$ with certainty.

One way to construct $\widehat{X}$

One way to construct $\widehat{X}$ [from Diaconis and Fill (1990)]:

- The chain X starts with $X_0 = 0$ and we set $\widehat{X}_0 = 0$.
- Inductively, we will have $\Lambda(\widehat{X}_t, X_t) > 0$ (and so $X_t \le \widehat{X}_t$) at all times t. In particular, $X_t \le X_{t-1} + 1 \le \widehat{X}_{t-1} + 1$.
- The value we construct for $\widehat{X}_t$ depends only on the values $\widehat{X}_{t-1} = \hat{x}$ and $X_t = y$ (with $y \le \hat{x} + 1$) and independent randomness. There are two cases.
- *Case 1:* If $y \le \hat{x}$ then set $\widehat{X}_t = \hat{x} + 1$ with probability
$$\frac{\widehat{P}(\hat{x}, \hat{x}+1)\Lambda(\hat{x}+1, y)}{(\widehat{P}\Lambda)(\hat{x}, y)} = \frac{(1-\theta_{\hat{x}})\Lambda(\hat{x}+1, y)}{\theta_{\hat{x}}\Lambda(\hat{x}, y) + (1-\theta_{\hat{x}})\Lambda(\hat{x}+1, y)}$$
and $\widehat{X}_t = \hat{x}$ with the complementary probability.
- *Case 2:* If $y = \hat{x} + 1$, then we set $\widehat{X}_t = \hat{x} + 1$ with certainty.

One way to construct $\widehat{X}$

One way to construct $\widehat{X}$ [from Diaconis and Fill (1990)]:

- The chain X starts with $X_0 = 0$ and we set $\widehat{X}_0 = 0$.
- Inductively, we will have $\Lambda(\widehat{X}_t, X_t) > 0$ (and so $X_t \le \widehat{X}_t$) at all times t. In particular, $X_t \le X_{t-1} + 1 \le \widehat{X}_{t-1} + 1$.
- The value we construct for $\widehat{X}_t$ depends only on the values $\widehat{X}_{t-1} = \hat{x}$ and $X_t = y$ (with $y \le \hat{x} + 1$) and independent randomness. There are two cases.
- *Case 1:* If $y \le \hat{x}$ then set $\widehat{X}_t = \hat{x} + 1$ with probability
$$\frac{\widehat{P}(\hat{x}, \hat{x}+1)\Lambda(\hat{x}+1, y)}{(\widehat{P}\Lambda)(\hat{x}, y)} = \frac{(1-\theta_{\hat{x}})\Lambda(\hat{x}+1, y)}{\theta_{\hat{x}}\Lambda(\hat{x}, y) + (1-\theta_{\hat{x}})\Lambda(\hat{x}+1, y)}$$
and $\widehat{X}_t = \hat{x}$ with the complementary probability.
- *Case 2:* If $y = \hat{x} + 1$, then we set $\widehat{X}_t = \hat{x} + 1$ with certainty.

One way to construct $\widehat{X}$

One way to construct $\widehat{X}$ [from Diaconis and Fill (1990)]:

- The chain X starts with $X_0 = 0$ and we set $\widehat{X}_0 = 0$.
- Inductively, we will have $\Lambda(\widehat{X}_t, X_t) > 0$ (and so $X_t \le \widehat{X}_t$) at all times t. In particular, $X_t \le X_{t-1} + 1 \le \widehat{X}_{t-1} + 1$.
- The value we construct for $\widehat{X}_t$ depends only on the values $\widehat{X}_{t-1} = \hat{x}$ and $X_t = y$ (with $y \le \hat{x} + 1$) and independent randomness. There are two cases.
- *Case 1:* If $y \le \hat{x}$ then set $\widehat{X}_t = \hat{x} + 1$ with probability
$$\frac{\widehat{P}(\hat{x}, \hat{x}+1)\Lambda(\hat{x}+1, y)}{(\widehat{P}\Lambda)(\hat{x}, y)} = \frac{(1-\theta_{\hat{x}})\Lambda(\hat{x}+1, y)}{\theta_{\hat{x}}\Lambda(\hat{x}, y) + (1-\theta_{\hat{x}})\Lambda(\hat{x}+1, y)}$$
and $\widehat{X}_t = \hat{x}$ with the complementary probability.
- *Case 2:* If $y = \hat{x} + 1$, then we set $\widehat{X}_t = \hat{x} + 1$ with certainty.

One way to construct $\widehat{X}$

One way to construct $\widehat{X}$ [from Diaconis and Fill (1990)]:

- The chain X starts with $X_0 = 0$ and we set $\widehat{X}_0 = 0$.
- Inductively, we will have $\Lambda(\widehat{X}_t, X_t) > 0$ (and so $X_t \le \widehat{X}_t$) at all times t. In particular, $X_t \le X_{t-1} + 1 \le \widehat{X}_{t-1} + 1$.
- The value we construct for $\widehat{X}_t$ depends only on the values $\widehat{X}_{t-1} = \hat{x}$ and $X_t = y$ (with $y \le \hat{x} + 1$) and independent randomness. There are two cases.
- *Case 1:* If $y \le \hat{x}$ then set $\widehat{X}_t = \hat{x} + 1$ with probability
$$\frac{\widehat{P}(\hat{x}, \hat{x}+1)\Lambda(\hat{x}+1, y)}{(\widehat{P}\Lambda)(\hat{x}, y)} = \frac{(1-\theta_{\hat{x}})\Lambda(\hat{x}+1, y)}{\theta_{\hat{x}}\Lambda(\hat{x}, y) + (1-\theta_{\hat{x}})\Lambda(\hat{x}+1, y)}$$
and $\widehat{X}_t = \hat{x}$ with the complementary probability.
- *Case 2:* If $y = \hat{x} + 1$, then we set $\widehat{X}_t = \hat{x} + 1$ with certainty.

One way to construct $\widehat{X}$: continuous-time B&D chains

One way to construct $\widehat{X}$ [briefly, using Fill (1992, *Journal of Theoretical Probability*)]:

- If the bivariate chain $(\widehat{X}, X)$ is in state $(\hat{x}, x)$ at a given jump time, then we construct an exponential random variable with rate

$$r = \nu_{\hat{x}} \Lambda(\hat{x}+1, x)/\Lambda(\hat{x}, x).$$

- If X jumps before this exponential expires, then $\widehat{X}$ holds unless X jumps to $\hat{x}+1$, in which case $\widehat{X}$ also jumps to $\hat{x}+1$.
- But if the exponential expires first, then at that expiration time X holds but $\widehat{X}$ jumps to $\hat{x}+1$.

One way to construct $\widehat{X}$: continuous-time B&D chains

One way to construct $\widehat{X}$ [briefly, using Fill (1992, *Journal of Theoretical Probability*)]:

- If the bivariate chain $(\widehat{X}, X)$ is in state $(\hat{x}, x)$ at a given jump time, then we construct an exponential random variable with rate

$$r = \nu_{\hat{x}} \Lambda(\hat{x}+1, x) / \Lambda(\hat{x}, x).$$

- If X jumps before this exponential expires, then $\widehat{X}$ holds unless X jumps to $\hat{x}+1$, in which case $\widehat{X}$ also jumps to $\hat{x}+1$.
- But if the exponential expires first, then at that expiration time X holds but $\widehat{X}$ jumps to $\hat{x}+1$.

One way to construct $\widehat{X}$: continuous-time B&D chains

One way to construct $\widehat{X}$ [briefly, using Fill (1992, *Journal of Theoretical Probability*)]:

- If the bivariate chain $(\widehat{X}, X)$ is in state $(\hat{x}, x)$ at a given jump time, then we construct an exponential random variable with rate

$$r = \nu_{\hat{x}} \Lambda(\hat{x}+1, x)/\Lambda(\hat{x}, x).$$

- If X jumps before this exponential expires, then $\widehat{X}$ holds unless X jumps to $\hat{x}+1$, in which case $\widehat{X}$ also jumps to $\hat{x}+1$.
- But if the exponential expires first, then at that expiration time X holds but $\widehat{X}$ jumps to $\hat{x}+1$.

A final remark: stochastic constructions for skip-free chains

Remark: In the general setting of our hitting-time theorem for upward-skip-free chains, we do not know any broad class of examples other than the B&D chains we have just treated for which the eigenvalues are nonnegative real numbers and the spectral polynomials are nonnegative matrices. Nevertheless, the stochastic construction we have described applies verbatim to all such chains.