On Hitting Times and Fastest Strong Stationary Times for Birth-and-Death, Other Skip-Free, and More General Chains

Jim Fill

Department of Applied Mathematics and Statistics
The Johns Hopkins University

May 18, 2009

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*

$$\mathbf{E}\, e^{-uT} = \prod_{j=0}^{d-1} \frac{\nu_j}{\nu_j + u}.$$

- *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* $= \{0, \dots, d\}$; *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* $\nu_0, \dots, \nu_{d-1}$ *be the* d *nonzero eigenvalues of* $-G$ *(known to have positive real parts).*
- *Then the hitting time* T *for state* d *has Laplace transform*

$$\mathbf{E}\, e^{-uT} = \prod_{j=0}^{d-1} \frac{\nu_j}{\nu_j + u}.$$

- *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* $= \{0, \dots, d\}$; *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* $\nu_0, \dots, \nu_{d-1}$ *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 (2009)].

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 (2009)].

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

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

- *state space* $= \{0, \dots, d\}$; *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* $\nu_0, \dots, \nu_{d-1}$ *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 (2009)].

Fastest SSTs for continuous-time skip-free chains

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

Theorem

- *state space* $= \{0, \dots, d\}$; *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* $\nu_0, \dots, \nu_{d-1}$ *be the* d *nonzero eigenvalues of* $-G$ *(known to have positive real parts).*
- *Then the hitting time* T *for state* d *has Laplace transform*

$$\mathbf{E}\, e^{-uT} = \prod_{j=0}^{d-1} \frac{\nu_j}{\nu_j + u}.$$

Fastest SSTs for continuous-time skip-free chains

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

Theorem

- *state space* $= \{0, \dots, d\}$; *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* $\nu_0, \dots, \nu_{d-1}$ *be the* d *nonzero eigenvalues of* $-G$ *(known to have positive real parts).*
- *Then the hitting time* T *for state* d *has Laplace transform*

$$\mathbf{E}\, e^{-uT} = \prod_{j=0}^{d-1} \frac{\nu_j}{\nu_j + u}.$$

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*

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

- *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*
$$\mathbf{E}\, u^T = \prod_{j=0}^{d-1} \frac{(1-\theta_j)u}{1-\theta_j u}.$$
- *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).

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).

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).

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 $\{0, \ldots, d\}$ 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" $\Lambda P^t = \widehat{P}^t \Lambda$ follows immediately.

"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.

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

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

     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 $k = 0, \dots, d$,

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

with the natural convention $Q_0 := I$.

- Note the recurrence relation

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

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

- Define

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

1.(a) Definition of Λ

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

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

with the natural convention $Q_0 := I$.

- Note the recurrence relation

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

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

- Define

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

1.(a) Definition of Λ

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

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

  with the natural convention $Q_0 := I$.

- Note the recurrence relation

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

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

- Define

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

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 $Q_.$'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 $k = 0, \ldots, d$ 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 $\widehat{P}$.

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 $\widehat{P}$.

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

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

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

- 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 $\max\{|\theta_0|, \dots, |\theta_{d-1}|\} < 1$, 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, $\widehat{P}^t(0, d) \to 1$.

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

We know $\mathbf{P}(T \leq t) = \widehat{P}^t(0, d)\Lambda(d, d)$ 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 $I - u\widehat{P}$ 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 $I - u\widehat{P}$ 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

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

- 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

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

- 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

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

- 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

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

- 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 $\widehat{X}$.

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 $X_t \leq \widehat{X}_t$) at all times t. In particular, $X_t \leq X_{t-1} + 1 \leq \widehat{X}_{t-1} + 1$.
- The value we construct for $\widehat{X}_t$ depends only on the values $\widehat{X}_{t-1} = \hat{x}$ and $X_t = y$ (with $y \leq \hat{x} + 1$) and independent randomness. There are two cases.
- *Case 1:* If $y \leq \hat{x}$ 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 $X_t \leq \widehat{X}_t$) at all times t. In particular, $X_t \leq X_{t-1} + 1 \leq \widehat{X}_{t-1} + 1$.
- The value we construct for $\widehat{X}_t$ depends only on the values $\widehat{X}_{t-1} = \hat{x}$ and $X_t = y$ (with $y \leq \hat{x} + 1$) and independent randomness. There are two cases.
- *Case 1:* If $y \leq \hat{x}$ 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 $X_t \leq \widehat{X}_t$) at all times t. In particular, $X_t \leq X_{t-1} + 1 \leq \widehat{X}_{t-1} + 1$.
- The value we construct for $\widehat{X}_t$ depends only on the values $\widehat{X}_{t-1} = \hat{x}$ and $X_t = y$ (with $y \leq \hat{x} + 1$) and independent randomness. There are two cases.
- *Case 1:* If $y \leq \hat{x}$ 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.

From hitting times to occupation times [Kent (1983)]

- For the next two slides, we switch to continuous time. A proof that's one line from Karlin–McGregor theorem shows that the hitting time T_d for an irred. B&D chain with gen. G has L.T.

$$\mathbf{E}\, e^{-uT} = \frac{\det(-G_0)}{\det(-G_0 + uI)},$$

with G_0 obtained from G by leaving off the last row & column.

- From this it's easy to show [using a nice remark of Kent (1983)] that the occupation-time vector $\mathbf{T} = (T_0, T_1, \ldots, T_{d-1})$ prior to hitting d has L.T.

$$\mathbf{E}\, e^{-\langle \mathbf{u}, \mathbf{T} \rangle} = \frac{\det(-G_0)}{\det(-G_0 + U)},$$

where $U := \operatorname{diag}(u_0, \ldots, u_{d-1})$. So what?

From occupation times to the Ray–Knight Theorem

- The occupation-time vector $\mathbf{T} = (T_0, T_1, \ldots, T_{d-1})$ for the chain prior to hitting d has L.T.

$$\mathbf{E}\, e^{-\langle \mathbf{u}, \mathbf{T} \rangle} = \det(-G_0) / \det(-G_0 + U).$$

So what?

- The matrix $S := D(-G_0)D^{-1}$ is positive definite, where π is G-stationary and $D := \text{diag}(\sqrt{\pi})$. Let $\Sigma := \frac{1}{2} S^{-1}$. By direct calculation, $\mathbf{T}$ has the same law as $\mathbf{Y} + \mathbf{Z}$, where $\mathbf{Y}$ and $\mathbf{Z}$ are independent random vectors with the same law and $\mathbf{Y}$ is the coord.-wise square of a Gaussian random vector $\mathbf{V} \sim \mathrm{N}(0, \Sigma)$.
- Kent (1983) uses/extends this “double derivation” of $\mathcal{L}(\mathbf{T})$ to prove the Ray (1963)–Knight (1963) theorem expressing the local time of Brownian motion as the sum of two independent 2-dimensional Bessel processes (i.e., as the sum of two independent squared Brownian motions).

Absorption for general chains: Three questions

Our proof of the central theorem for the absorption time of a discrete-time skip-free chain rested on the construction of a matrix Λ having the properties (i)–(iii) (lower triangular, rows sum to 1, "intertwining"). Three questions:

(a) Can the theorem be extended to general absorbing chains?

(b) Is the spectral-polynomials construction of Λ inevitable? I.e., is the matrix Λ uniquely determined by properties (i)–(iii)?

(c) If the eigenvalues and spectral polynomials of a general absorbing chain are all nonnegative, can our stochastic construction be extended?

Answers: yes, yes, yes

Absorption for general chains: Three questions

Our proof of the central theorem for the absorption time of a discrete-time skip-free chain rested on the construction of a matrix Λ having the properties (i)–(iii) (lower triangular, rows sum to 1, "intertwining"). Three questions:

(a) Can the theorem be extended to general absorbing chains?

(b) Is the spectral-polynomials construction of Λ inevitable? I.e., is the matrix Λ uniquely determined by properties (i)–(iii)?

(c) If the eigenvalues and spectral polynomials of a general absorbing chain are all nonnegative, can our stochastic construction be extended?

Answers: yes, yes, yes

Absorption for general chains: Three questions

Our proof of the central theorem for the absorption time of a discrete-time skip-free chain rested on the construction of a matrix Λ having the properties (i)–(iii) (lower triangular, rows sum to 1, "intertwining"). Three questions:

(a) Can the theorem be extended to general absorbing chains?

(b) Is the spectral-polynomials construction of Λ inevitable? I.e., is the matrix Λ uniquely determined by properties (i)–(iii)?

(c) If the eigenvalues and spectral polynomials of a general absorbing chain are all nonnegative, can our stochastic construction be extended?

Answers: yes, yes, yes

Absorption for general chains: Three questions

Our proof of the central theorem for the absorption time of a discrete-time skip-free chain rested on the construction of a matrix Λ having the properties (i)–(iii) (lower triangular, rows sum to 1, "intertwining"). Three questions:

(a) Can the theorem be extended to general absorbing chains?

(b) Is the spectral-polynomials construction of Λ inevitable? I.e., is the matrix Λ uniquely determined by properties (i)–(iii)?

(c) If the eigenvalues and spectral polynomials of a general absorbing chain are all nonnegative, can our stochastic construction be extended?

Answers: yes, yes, yes

Absorption for general chains: Three questions

Our proof of the central theorem for the absorption time of a discrete-time skip-free chain rested on the construction of a matrix Λ having the properties (i)–(iii) (lower triangular, rows sum to 1, "intertwining"). Three questions:

(a) Can the theorem be extended to general absorbing chains?

(b) Is the spectral-polynomials construction of Λ inevitable? I.e., is the matrix Λ uniquely determined by properties (i)–(iii)?

(c) If the eigenvalues and spectral polynomials of a general absorbing chain are all nonnegative, can our stochastic construction be extended?

Answers: yes, yes, yes

Wildly different constructions of the same Λ

Question (b) about the inevitability of the spectral-polynomials link Λ arises naturally because (in the continuous-time B&D setting) our proof and that of **Diaconis and Miclo (2009)** both rely on construction of a link Λ such that $\Lambda G = \widehat{G}\Lambda$, where $\widehat{G}$ (the analogue of $\widehat{P}$ in continuous time) is the pure-birth "spectral" generator. The two methods of construction are strikingly different, so it is interesting that **the end-product Λ is the same**.

General setting for the three questions

General setting and notation:

- discrete-time Markov chain X; state space $= \{0, \ldots, d\}$
- general init. distn. m_0 (row vector) and transition matrix P
- State d is absorbing and accessible from each other state.
- $\theta_0, \ldots, \theta_{d-1}$ are the d non-unit eigenvalues of P (in nondecreasing order if real and we care); $\theta_d = 1$
- $\widehat{P}$ as before ($\hat{p}_{ii} = \theta_i = 1 - \hat{p}_{i,i+1}$)
- normalized spectral polynomials $Q_0, \ldots, Q_d$ as before

General setting for the three questions

General setting and notation:

- discrete-time Markov chain X; state space $= \{0, \ldots, d\}$
- general init. distn. m_0 (row vector) and transition matrix P
- State d is absorbing and accessible from each other state.
- $\theta_0, \ldots, \theta_{d-1}$ are the d non-unit eigenvalues of P (in nondecreasing order if real and we care); $\theta_d = 1$
- $\widehat{P}$ as before ($\hat{p}_{ii} = \theta_i = 1 - \hat{p}_{i,i+1}$)
- normalized spectral polynomials $Q_0, \ldots, Q_d$ as before

General setting for the three questions

General setting and notation:

- discrete-time Markov chain X; state space $= \{0, \dots, d\}$
- general init. distn. m_0 (row vector) and transition matrix P
- State d is absorbing and accessible from each other state.
- $\theta_0, \dots, \theta_{d-1}$ are the d non-unit eigenvalues of P (in nondecreasing order if real and we care); $\theta_d = 1$
- $\widehat{P}$ as before ($\hat{p}_{ii} = \theta_i = 1 - \hat{p}_{i,i+1}$)
- normalized spectral polynomials $Q_0, \dots, Q_d$ as before

General setting for the three questions

General setting and notation:

- discrete-time Markov chain X; state space $= \{0, \dots, d\}$
- general init. distn. m_0 (row vector) and transition matrix P
- State d is absorbing and accessible from each other state.
- $\theta_0, \dots, \theta_{d-1}$ are the d non-unit eigenvalues of P (in nondecreasing order if real and we care); $\theta_d = 1$
- $\widehat{P}$ as before ($\hat{p}_{ii} = \theta_i = 1 - \hat{p}_{i,i+1}$)
- normalized spectral polynomials $Q_0, \dots, Q_d$ as before

General setting for the three questions

General setting and notation:

- discrete-time Markov chain X; state space $= \{0, \ldots, d\}$
- general init. distn. m_0 (row vector) and transition matrix P
- State d is absorbing and accessible from each other state.
- $\theta_0, \ldots, \theta_{d-1}$ are the d non-unit eigenvalues of P (in nondecreasing order if real and we care); $\theta_d = 1$
- $\widehat{P}$ as before $(\hat{p}_{ii} = \theta_i = 1 - \hat{p}_{i,i+1})$
- normalized spectral polynomials $Q_0, \ldots, Q_d$ as before

General setting for the three questions

General setting and notation:

- discrete-time Markov chain X; state space $= \{0, \dots, d\}$
- general init. distn. m_0 (row vector) and transition matrix P
- State d is absorbing and accessible from each other state.
- $\theta_0, \dots, \theta_{d-1}$ are the d non-unit eigenvalues of P (in nondecreasing order if real and we care); $\theta_d = 1$
- $\widehat{P}$ as before ($\hat{p}_{ii} = \theta_i = 1 - \hat{p}_{i,i+1}$)
- normalized spectral polynomials $Q_0, \dots, Q_d$ as before

General setting for the three questions

General setting and notation:

- discrete-time Markov chain X; state space $= \{0, \dots, d\}$
- general init. distn. m_0 (row vector) and transition matrix P
- State d is absorbing and accessible from each other state.
- $\theta_0, \dots, \theta_{d-1}$ are the d non-unit eigenvalues of P (in nondecreasing order if real and we care); $\theta_d = 1$
- $\widehat{P}$ as before ($\hat{p}_{ii} = \theta_i = 1 - \hat{p}_{i,i+1}$)
- normalized spectral polynomials $Q_0, \dots, Q_d$ as before

Inevitability of spectral-polynomials "link" (general setting)

Lemma

The unique matrix Λ (with rows denoted by $\lambda_0, \ldots, \lambda_d$) satisfying the two conditions

$$m_0 = \lambda_0 \qquad \text{and} \qquad \Lambda P = \widehat{P}\Lambda \tag{1}$$

is given by

$$\lambda_i = m_0 Q_i, \quad i = 0, \ldots, d. \tag{2}$$

Proof. It is easy to check that the choice (2) satisfies (1). Conversely, the ith row of $\Lambda P = \widehat{P}\Lambda$ ($i = 0, \ldots, d-1$) requires

$$\lambda_i P = \theta_i \lambda_i + (1-\theta_i)\lambda_{i+1}, \quad \text{i.e.,} \quad \lambda_{i+1} = \lambda_i \left[(1-\theta_i)^{-1}(P - \theta_i I)\right],$$

and so (by induction) (1) implies (2). □

Inevitability of spectral-polynomials "link" (general setting)

Lemma

The unique matrix Λ (with rows denoted by $\lambda_0, \dots, \lambda_d$) satisfying the two conditions

$$m_0 = \lambda_0 \qquad \text{and} \qquad \Lambda P = \widehat{P}\Lambda \tag{1}$$

is given by

$$\lambda_i = m_0 Q_i, \quad i = 0, \dots, d. \tag{2}$$

Proof. It is easy to check that the choice (2) satisfies (1). Conversely, the ith row of $\Lambda P = \widehat{P}\Lambda$ ($i = 0, \dots, d-1$) requires

$$\lambda_i P = \theta_i \lambda_i + (1-\theta_i)\lambda_{i+1}, \quad \text{i.e.,} \quad \lambda_{i+1} = \lambda_i \left[(1-\theta_i)^{-1}(P - \theta_i I)\right],$$

and so (by induction) (1) implies (2). □

Absorption time theorem for general chains: Notation

Coming next is the absorption-time theorem in the general setting. We need some conventions, notation, and observations:

- convention: An empty sum vanishes.
- notation: If the eigenvalues θ_i are all real and nonnegative, then $\mathcal{G}(\theta_0, \dots, \theta_{k-1})$ denotes the convolution of geometrics with success probabilities $1 - \theta_0, \dots, 1 - \theta_{k-1}$.
- notation: $\Lambda(-1, d) := 0$ and $\Lambda(d+1, d) := 1$
- notation: $a_k := \Lambda(k, d) - \Lambda(k-1, d), \quad k = 0, \dots, d+1$
- observation: The a_k's sum to unity.
- observation (to come): We always have $a_{d+1} = 0$.
- observation (easy to check): If the spectral polynomials are all nonnegative, then $\mathbf{a} := (a_0, \dots, a_d)$ is a pmf.

Absorption time theorem for general chains: Notation

Coming next is the absorption-time theorem in the general setting. We need some conventions, notation, and observations:

- convention: An empty sum vanishes.
- notation: If the eigenvalues θ_i are all real and nonnegative, then $\mathcal{G}(\theta_0, \dots, \theta_{k-1})$ denotes the convolution of geometrics with success probabilities $1 - \theta_0, \dots, 1 - \theta_{k-1}$.
- notation: $\Lambda(-1, d) := 0$ and $\Lambda(d + 1, d) := 1$
- notation: $a_k := \Lambda(k, d) - \Lambda(k - 1, d), \quad k = 0, \dots, d + 1$
- observation: The a_k's sum to unity.
- observation (to come): We always have $a_{d+1} = 0$.
- observation (easy to check): If the spectral polynomials are all nonnegative, then $\mathbf{a} := (a_0, \dots, a_d)$ is a pmf.

Absorption time theorem for general chains: Notation

Coming next is the absorption-time theorem in the general setting. We need some conventions, notation, and observations:

- convention: An empty sum vanishes.
- notation: If the eigenvalues θ_i are all real and nonnegative, then $\mathcal{G}(\theta_0, \ldots, \theta_{k-1})$ denotes the convolution of geometrics with success probabilities $1 - \theta_0, \ldots, 1 - \theta_{k-1}$.
- notation: $\Lambda(-1, d) := 0$ and $\Lambda(d+1, d) := 1$
- notation: $a_k := \Lambda(k, d) - \Lambda(k-1, d), \quad k = 0, \ldots, d+1$
- observation: The a_k's sum to unity.
- observation (to come): We always have $a_{d+1} = 0$.
- observation (easy to check): If the spectral polynomials are all nonnegative, then $\mathbf{a} := (a_0, \ldots, a_d)$ is a pmf.

Absorption time theorem for general chains: Notation

Coming next is the absorption-time theorem in the general setting. We need some conventions, notation, and observations:

- convention: An empty sum vanishes.
- notation: If the eigenvalues θ_i are all real and nonnegative, then $\mathcal{G}(\theta_0, \ldots, \theta_{k-1})$ denotes the convolution of geometrics with success probabilities $1 - \theta_0, \ldots, 1 - \theta_{k-1}$.
- notation: $\Lambda(-1, d) := 0$ and $\Lambda(d+1, d) := 1$
- notation: $a_k := \Lambda(k, d) - \Lambda(k-1, d), \quad k = 0, \ldots, d+1$
- observation: The a_k's sum to unity.
- observation (to come): We always have $a_{d+1} = 0$.
- observation (easy to check): If the spectral polynomials are all nonnegative, then $\mathbf{a} := (a_0, \ldots, a_d)$ is a pmf.

Absorption time theorem for general chains: Notation

Coming next is the absorption-time theorem in the general setting. We need some conventions, notation, and observations:

- convention: An empty sum vanishes.
- notation: If the eigenvalues θ_i are all real and nonnegative, then $\mathcal{G}(\theta_0, \ldots, \theta_{k-1})$ denotes the convolution of geometrics with success probabilities $1 - \theta_0, \ldots, 1 - \theta_{k-1}$.
- notation: $\Lambda(-1, d) := 0$ and $\Lambda(d + 1, d) := 1$
- notation: $a_k := \Lambda(k, d) - \Lambda(k - 1, d), \quad k = 0, \ldots, d + 1$
- observation: The a_k's sum to unity.
- observation (to come): We always have $a_{d+1} = 0$.
- observation (easy to check): If the spectral polynomials are all nonnegative, then $\mathbf{a} := (a_0, \ldots, a_d)$ is a pmf.

Absorption time theorem for general chains: Notation

Coming next is the absorption-time theorem in the general setting. We need some conventions, notation, and observations:

- convention: An empty sum vanishes.
- notation: If the eigenvalues θ_i are all real and nonnegative, then $\mathcal{G}(\theta_0, \ldots, \theta_{k-1})$ denotes the convolution of geometrics with success probabilities $1 - \theta_0, \ldots, 1 - \theta_{k-1}$.
- notation: $\Lambda(-1, d) := 0$ and $\Lambda(d + 1, d) := 1$
- notation: $a_k := \Lambda(k, d) - \Lambda(k - 1, d), \quad k = 0, \ldots, d + 1$
- observation: The a_k's sum to unity.
- observation (to come): We always have $a_{d+1} = 0$.
- observation (easy to check): If the spectral polynomials are all nonnegative, then $\mathbf{a} := (a_0, \ldots, a_d)$ is a pmf.

Absorption time theorem for general chains: Notation

Coming next is the absorption-time theorem in the general setting. We need some conventions, notation, and observations:

- convention: An empty sum vanishes.
- notation: If the eigenvalues θ_i are all real and nonnegative, then $\mathcal{G}(\theta_0, \ldots, \theta_{k-1})$ denotes the convolution of geometrics with success probabilities $1 - \theta_0, \ldots, 1 - \theta_{k-1}$.
- notation: $\Lambda(-1, d) := 0$ and $\Lambda(d+1, d) := 1$
- notation: $a_k := \Lambda(k, d) - \Lambda(k-1, d), \quad k = 0, \ldots, d+1$
- observation: The a_k's sum to unity.
- observation (to come): We always have $a_{d+1} = 0$.
- observation (easy to check): If the spectral polynomials are all nonnegative, then $\mathbf{a} := (a_0, \ldots, a_d)$ is a pmf.

Absorption time theorem for general chains: Notation

Coming next is the absorption-time theorem in the general setting. We need some conventions, notation, and observations:

- convention: An empty sum vanishes.
- notation: If the eigenvalues θ_i are all real and nonnegative, then $\mathcal{G}(\theta_0, \ldots, \theta_{k-1})$ denotes the convolution of geometrics with success probabilities $1 - \theta_0, \ldots, 1 - \theta_{k-1}$.
- notation: $\Lambda(-1, d) := 0$ and $\Lambda(d+1, d) := 1$
- notation: $a_k := \Lambda(k, d) - \Lambda(k-1, d), \quad k = 0, \ldots, d+1$
- observation: The a_k's sum to unity.
- observation (to come): We always have $a_{d+1} = 0$.
- observation (easy to check): If the spectral polynomials are all nonnegative, then $\mathbf{a} := (a_0, \ldots, a_d)$ is a pmf.

Absorption time theorem for general chains

Theorem

In the general abs.-chain setting, the absorption time T satisfies

$$\mathbf{P}(T \le t) = \sum_{k=0}^{d} a_k \sum_{j=k}^{d} \widehat{P}^t(0,j), \quad t = 0, 1, 2, \ldots,$$

with probability generating function

$$\mathbf{E}\, u^T = \sum_{k=0}^{d} a_k \prod_{j=0}^{k-1} \left[\frac{(1-\theta_j)u}{1-\theta_j u} \right].$$

In particular, if the spectral polynomials are all nonnegative matrices, then T is distributed as the mixture

$$\sum_{k=0}^{d} a_k \, \mathcal{G}(\theta_0, \ldots, \theta_{k-1}).$$

Absorption time theorem for general chains

Theorem

In the general abs.-chain setting, the absorption time T satisfies

$$\mathbf{P}(T \le t) = \sum_{k=0}^{d} a_k \sum_{j=k}^{d} \widehat{P}^t(0,j), \quad t = 0, 1, 2, \dots,$$

with probability generating function

$$\mathbf{E}\, u^T = \sum_{k=0}^{d} a_k \prod_{j=0}^{k-1} \left[\frac{(1-\theta_j)u}{1-\theta_j u} \right].$$

In particular, if the spectral polynomials are all nonnegative matrices, then T is distributed as the mixture

$$\sum_{k=0}^{d} a_k\, \mathcal{G}(\theta_0, \dots, \theta_{k-1}).$$

Absorption time thm. for general chains: Outline of proof

Proof (outline). Using our general "intertwining" lemma as in the skip-free case and then summation by parts, one finds

$$\mathbf{P}(T \le t) = \sum_{j=0}^{d} \widehat{P}^t(0,j)\Lambda(j,d) = \sum_{k=0}^{d} a_k \sum_{j=k}^{d} \widehat{P}^t(0,j).$$

We have already studied $\widehat{P}$ and know $\widehat{P}^t(0,j) \to \delta_{d,j}$ as $t \to \infty$; thus $1 = \sum_{k=0}^{d} a_k$ and so $a_{d+1} = 0$. The displayed equation is all that is needed to establish the formula for $\mathbf{E}\, u^T$ when the eigenvalues of P are nonnegative real numbers; in general one can use the explicit formula for $(I - u\widehat{P})^{-1}$ as done in the skip-free case. □

Absorption time thm. for general chains: Outline of proof

Proof (outline). Using our general "intertwining" lemma as in the skip-free case and then summation by parts, one finds

$$\mathbf{P}(T \le t) = \sum_{j=0}^{d} \widehat{P}^t(0,j)\Lambda(j,d) = \sum_{k=0}^{d} a_k \sum_{j=k}^{d} \widehat{P}^t(0,j).$$

We have already studied $\widehat{P}$ and know $\widehat{P}^t(0,j) \to \delta_{d,j}$ as $t \to \infty$; thus $1 = \sum_{k=0}^{d} a_k$ and so $a_{d+1} = 0$. The displayed equation is all that is needed to establish the formula for $\mathbf{E}\, u^T$ when the eigenvalues of P are nonnegative real numbers; in general one can use the explicit formula for $(I - u\widehat{P})^{-1}$ as done in the skip-free case. □

Absorption time thm. for general chains: Pleasant cases

Pleasant cases:

- **skip-free:** If the chain is upward skip-free and $m_0 = \delta_0$, then $a_k \equiv \delta_{d,k}$ and the earlier theorem is recovered.
- **detailed balance** [treated differently by Miclo (2008)]: If there exists π satisfying $\pi_i p_{ij} = \pi_j p_{ji}$ for all non-absorbing i, j, then arguing as for B&D chains, the eigenvalues of P are nonnegative reals and the spectral polynomials are nonnegative. Thus the absorption time is distributed as a mixture of convolutions of geometric distributions. Miclo also shows that $a_d > 0$ when $0, \ldots, d-1$ all communicate.
- **upper triangular:** If P is upper triangular, then of course the eigenvalues are nonnegative reals, and He and Zhang (2006) prove that the spectral polynomials are nonnegative.

Absorption time thm. for general chains: Pleasant cases

Pleasant cases:

- **skip-free:** If the chain is upward skip-free and $m_0 = \delta_0$, then $a_k \equiv \delta_{d,k}$ and the earlier theorem is recovered.
- **detailed balance** [treated differently by Miclo (2008)]: If there exists π satisfying $\pi_i p_{ij} = \pi_j p_{ji}$ for all non-absorbing i, j, then arguing as for B&D chains, the eigenvalues of P are nonnegative reals and the spectral polynomials are nonnegative. Thus the absorption time is distributed as a mixture of convolutions of geometric distributions. Miclo also shows that $a_d > 0$ when $0, \ldots, d-1$ all communicate.
- **upper triangular:** If P is upper triangular, then of course the eigenvalues are nonnegative reals, and He and Zhang (2006) prove that the spectral polynomials are nonnegative.

Absorption time thm. for general chains: Pleasant cases

Pleasant cases:

- **skip-free:** If the chain is upward skip-free and $m_0 = \delta_0$, then $a_k \equiv \delta_{d,k}$ and the earlier theorem is recovered.
- **detailed balance** [treated differently by Miclo (2008)]: If there exists π satisfying $\pi_i p_{ij} = \pi_j p_{ji}$ for all non-absorbing i, j, then arguing as for B&D chains, the eigenvalues of P are nonnegative reals and the spectral polynomials are nonnegative. Thus the absorption time is distributed as a mixture of convolutions of geometric distributions. Miclo also shows that $a_d > 0$ when $0, \dots, d-1$ all communicate.
- **upper triangular:** If P is upper triangular, then of course the eigenvalues are nonnegative reals, and He and Zhang (2006) prove that the spectral polynomials are nonnegative.

Absorption time thm. for general chains: Pleasant cases

Pleasant cases:

- **skip-free:** If the chain is upward skip-free and $m_0 = \delta_0$, then $a_k \equiv \delta_{d,k}$ and the earlier theorem is recovered.
- **detailed balance** [treated differently by Miclo (2008)]: If there exists π satisfying $\pi_i p_{ij} = \pi_j p_{ji}$ for all non-absorbing i, j, then arguing as for B&D chains, the eigenvalues of P are nonnegative reals and the spectral polynomials are nonnegative. Thus the absorption time is distributed as a mixture of convolutions of geometric distributions. Miclo also shows that $a_d > 0$ when $0, \dots, d-1$ all communicate.
- **upper triangular:** If P is upper triangular, then of course the eigenvalues are nonnegative reals, and He and Zhang (2006) prove that the spectral polynomials are nonnegative.

Absorption time thm. for general chains: Two extensions

Two extensions:

- **continuous time:** An analogous theorem holds for continuous-time chains. See Miclo (2008) for a discussion of connections with the extensive literature on so-called "phase-type" distributions (Neuts; O'Cinneide; He and Zhang; Botta, Harris, and Marchal; Commault and Mocanu; etc.).
- **fastest SSTs:** A similar theorem—except that now a_k needs to be defined as $[\Lambda(k,d) - \lambda(k-1,d)]/\pi(d)$—holds (with a similar proof) for the distn. of a fastest SST of a general ergodic chain X with general init. distn. m_0 and stat. distn. π, provided $\mathbf{P}(X_t = x)/\pi(x)$ is minimized for every t by the choice $x = d$. [Suff. cond. for this: the time-reversal of P is stochastically monotone (wrt natural linear order on $\{0,\ldots,d\}$) and $m_0(x)/\pi(x)$ is decreasing in x (for example, $m_0 = \delta_0$).]

Absorption time thm. for general chains: Two extensions

Two extensions:

- **continuous time:** An analogous theorem holds for continuous-time chains. See Miclo (2008) for a discussion of connections with the extensive literature on so-called "phase-type" distributions (Neuts; O'Cinneide; He and Zhang; Botta, Harris, and Marchal; Commault and Mocanu; etc.).
- **fastest SSTs:** A similar theorem—except that now a_k needs to be defined as $[\Lambda(k, d) - \lambda(k-1, d)]/\pi(d)$—holds (with a similar proof) for the distn. of a fastest SST of a general ergodic chain X with general init. distn. m_0 and stat. distn. π, provided $\mathbf{P}(X_t = x)/\pi(x)$ is minimized for every t by the choice $x = d$. [Suff. cond. for this: the time-reversal of P is stochastically monotone (wrt natural linear order on $\{0, \dots, d\}$) and $m_0(x)/\pi(x)$ is decreasing in x (for example, $m_0 = \delta_0$).]

Absorption time thm. for general chains: Two extensions

Two extensions:

- **continuous time:** An analogous theorem holds for continuous-time chains. See Miclo (2008) for a discussion of connections with the extensive literature on so-called "phase-type" distributions (Neuts; O'Cinneide; He and Zhang; Botta, Harris, and Marchal; Commault and Mocanu; etc.).
- **fastest SSTs:** A similar theorem—except that now a_k needs to be defined as $[\Lambda(k,d) - \lambda(k-1,d)]/\pi(d)$—holds (with a similar proof) for the distn. of a fastest SST of a general ergodic chain X with general init. distn. m_0 and stat. distn. π, provided $\mathbf{P}(X_t = x)/\pi(x)$ is minimized for every t by the choice $x = d$. [Suff. cond. for this: the time-reversal of P is stochastically monotone (wrt natural linear order on $\{0,\ldots,d\}$) and $m_0(x)/\pi(x)$ is decreasing in x (for example, $m_0 = \delta_0$).]

Stochastic construction perhaps, but . . .

- Whenever the eigenvalues and spectral polynomials are all nonnegative, the spectral link Λ provides an intertwining of the semigroups $(P^t)_{t\geq 0}$ and $(\widehat{P}^t)_{t\geq 0}$, and again (as for B&D chains) a chain $\widehat{X}$ with kernel $\widehat{P}$ can be constructed such that $\widehat{X}_0 = 0$ and
$$\text{distn}(X_t \,|\, \widehat{X}_0, \ldots, \widehat{X}_t) = \Lambda(\widehat{X}_t, \cdot) \quad \text{for all } t.$$
- In that case, although we do have $\Lambda(d, d) = 1$ (this simply restates our earlier observation that $a_{d+1} = 0$), unlike for skip-free chains there is no guarantee that the link Λ is lower triangular; so all we can say with certainty is that $T \leq \widehat{T}$.
- We will rectify this situation by modifying the link Λ; this will not contradict the uniqueness of Λ proven earlier, because we will also substitute a different "dual" kernel $\overline{P}$ for the spectral kernel $\widehat{P}$.

Stochastic construction perhaps, but . . .

- Whenever the eigenvalues and spectral polynomials are all nonnegative, the spectral link Λ provides an intertwining of the semigroups $(P^t)_{t \geq 0}$ and $(\widehat{P}^t)_{t \geq 0}$, and again (as for B&D chains) a chain $\widehat{X}$ with kernel $\widehat{P}$ can be constructed such that $\widehat{X}_0 = 0$ and
$$\text{distn}(X_t \,|\, \widehat{X}_0, \ldots, \widehat{X}_t) = \Lambda(\widehat{X}_t, \cdot) \quad \text{for all } t.$$
- In that case, although we do have $\Lambda(d, d) = 1$ (this simply restates our earlier observation that $a_{d+1} = 0$), unlike for skip-free chains there is no guarantee that the link Λ is lower triangular; so all we can say with certainty is that $T \leq \widehat{T}$.
- We will rectify this situation by modifying the link Λ; this will not contradict the uniqueness of Λ proven earlier, because we will also substitute a different "dual" kernel $\overline{P}$ for the spectral kernel $\widehat{P}$.

Stochastic construction perhaps, but . . .

- Whenever the eigenvalues and spectral polynomials are all nonnegative, the spectral link Λ provides an intertwining of the semigroups $(P^t)_{t \geq 0}$ and $(\widehat{P}^t)_{t \geq 0}$, and again (as for B&D chains) a chain $\widehat{X}$ with kernel $\widehat{P}$ can be constructed such that $\widehat{X}_0 = 0$ and
$$\text{distn}(X_t \,|\, \widehat{X}_0, \dots, \widehat{X}_t) = \Lambda(\widehat{X}_t, \cdot) \quad \text{for all } t.$$
- In that case, although we do have $\Lambda(d, d) = 1$ (this simply restates our earlier observation that $a_{d+1} = 0$), unlike for skip-free chains there is no guarantee that the link Λ is lower triangular; so all we can say with certainty is that $T \leq \widehat{T}$.
- We will rectify this situation by modifying the link Λ; this will not contradict the uniqueness of Λ proven earlier, because we will also substitute a different "dual" kernel $\overline{P}$ for the spectral kernel $\widehat{P}$.

A modified "link" $\overline{\Lambda}$

- We define a modified "link" $\overline{\Lambda}$ without yet assuming nonnegativity of eigenvalues or spectral polynomials.
- Define the matrix $\overline{\Lambda}$ with rows $\bar{\lambda}_0, \ldots, \bar{\lambda}_d$ by setting

$$\bar{\lambda}_i := \begin{cases} [1 - \lambda_i(d)]^{-1}[\lambda_i - \lambda_i(d)\delta_d] & \text{if } \lambda_i(d) \neq 1 \\ \delta_d & \text{if } \lambda_i(d) = 1 \end{cases}$$

  where δ_d is the coordinate row vector $(0, 0, \ldots, 0, 1)$.
- Note that the rows of $\overline{\Lambda}$, like those of Λ, sum to unity.

A modified "link" $\overline{\Lambda}$

- We define a modified "link" $\overline{\Lambda}$ without yet assuming nonnegativity of eigenvalues or spectral polynomials.
- Define the matrix $\overline{\Lambda}$ with rows $\bar{\lambda}_0, \ldots, \bar{\lambda}_d$ by setting

$$\bar{\lambda}_i := \begin{cases} [1 - \lambda_i(d)]^{-1}[\lambda_i - \lambda_i(d)\delta_d] & \text{if } \lambda_i(d) \neq 1 \\ \delta_d & \text{if } \lambda_i(d) = 1 \end{cases}$$

  where δ_d is the coordinate row vector $(0, 0, \ldots, 0, 1)$.
- Note that the rows of $\overline{\Lambda}$, like those of Λ, sum to unity.

A modified "link" $\overline{\Lambda}$

- We define a modified "link" $\overline{\Lambda}$ without yet assuming nonnegativity of eigenvalues or spectral polynomials.
- Define the matrix $\overline{\Lambda}$ with rows $\bar{\lambda}_0, \ldots, \bar{\lambda}_d$ by setting

$$\bar{\lambda}_i := \begin{cases} [1 - \lambda_i(d)]^{-1}[\lambda_i - \lambda_i(d)\delta_d] & \text{if } \lambda_i(d) \neq 1 \\ \delta_d & \text{if } \lambda_i(d) = 1 \end{cases}$$

  where δ_d is the coordinate row vector $(0, 0, \ldots, 0, 1)$.
- Note that the rows of $\overline{\Lambda}$, like those of Λ, sum to unity.

A modified dual init. distn. $\overline{m}_0$ and "dual kernel" $\overline{P}$

- We will define a modified dual initial distribution $\overline{m}_0$ (modified from δ_0) and a modified "dual kernel" $\overline{P}$, both still without yet assuming nonnegativity of eigenvalues or spectral polynomials.
- We define $\overline{m}_0$ to be the probability row vector

$$\overline{m}_0 := (1 - m_0(d), 0, \ldots, 0, m_0(d)) = (1 - a_0, 0, \ldots, 0, a_0).$$

- We will define $\overline{P} := B + R$ to be the sum of a bidiagonal upper triangular matrix B and a matrix R with rank at most 1 vanishing in all columns except for the last. The definitions of B and R follow on the next slide.
- It will be clear that the rows of $\overline{P}$ sum to unity.

A modified dual init. distn. $\overline{m}_0$ and "dual kernel" $\overline{P}$

- We will define a modified dual initial distribution $\overline{m}_0$ (modified from δ_0) and a modified "dual kernel" $\overline{P}$, both still without yet assuming nonnegativity of eigenvalues or spectral polynomials.
- We define $\overline{m}_0$ to be the probability row vector

$$\overline{m}_0 := (1 - m_0(d), 0, \ldots, 0, m_0(d)) = (1 - a_0, 0, \ldots, 0, a_0).$$

- We will define $\overline{P} := B + R$ to be the sum of a bidiagonal upper triangular matrix B and a matrix R with rank at most 1 vanishing in all columns except for the last. The definitions of B and R follow on the next slide.
- It will be clear that the rows of $\overline{P}$ sum to unity.

A modified dual init. distn. $\overline{m}_0$ and "dual kernel" $\overline{P}$

- We will define a modified dual initial distribution $\overline{m}_0$ (modified from δ_0) and a modified "dual kernel" $\overline{P}$, both still without yet assuming nonnegativity of eigenvalues or spectral polynomials.
- We define $\overline{m}_0$ to be the probability row vector

$$\overline{m}_0 := (1 - m_0(d), 0, \ldots, 0, m_0(d)) = (1 - a_0, 0, \ldots, 0, a_0).$$

- We will define $\overline{P} := B + R$ to be the sum of a bidiagonal upper triangular matrix B and a matrix R with rank at most 1 vanishing in all columns except for the last. The definitions of B and R follow on the next slide.
- It will be clear that the rows of $\overline{P}$ sum to unity.

A modified dual init. distn. $\overline{m}_0$ and "dual kernel" $\overline{P}$

- We will define a modified dual initial distribution $\overline{m}_0$ (modified from δ_0) and a modified "dual kernel" $\overline{P}$, both still without yet assuming nonnegativity of eigenvalues or spectral polynomials.
- We define $\overline{m}_0$ to be the probability row vector

$$\overline{m}_0 := (1 - m_0(d), 0, \dots, 0, m_0(d)) = (1 - a_0, 0, \dots, 0, a_0).$$

- We will define $\overline{P} := B + R$ to be the sum of a bidiagonal upper triangular matrix B and a matrix R with rank at most 1 vanishing in all columns except for the last. The definitions of B and R follow on the next slide.
- It will be clear that the rows of $\overline{P}$ sum to unity.

Definitions of B and R (with $\overline{P} = B + R$)

- If $\lambda_i(d) \neq 1$, define

$$b_{ij} := \begin{cases} \theta_i & \text{if } j = i \\ \frac{1-\lambda_{i+1}(d)}{1-\lambda_i(d)}(1-\theta_i) & \text{if } j = i+1 \\ 0 & \text{otherwise,} \end{cases}$$

and

$$r_{id} := \left[1 - \frac{1-\lambda_{i+1}(d)}{1-\lambda_i(d)}\right](1-\theta_i) = \frac{a_{i+1}}{1-\lambda_i(d)}(1-\theta_i).$$

- If $\lambda_i(d) = 1$, define

$$b_{ij} := \delta_{ij}, \qquad r_{id} := 0.$$

Definitions of B and R (with $\overline{P} = B + R$)

- If $\lambda_i(d) \neq 1$, define

$$b_{ij} := \begin{cases} \theta_i & \text{if } j = i \\ \frac{1-\lambda_{i+1}(d)}{1-\lambda_i(d)}(1-\theta_i) & \text{if } j = i+1 \\ 0 & \text{otherwise,} \end{cases}$$

  and

$$r_{id} := \left[1 - \frac{1-\lambda_{i+1}(d)}{1-\lambda_i(d)}\right](1-\theta_i) = \frac{a_{i+1}}{1-\lambda_i(d)}(1-\theta_i).$$

- If $\lambda_i(d) = 1$, define

$$b_{ij} := \delta_{ij}, \qquad r_{id} := 0.$$

"Intertwining" for the modified matrices

The following key "intertwining" equations for $\overline{m}_0$ and $\overline{P}$ follow by straightforward calculations from the corresponding equations for $\widehat{m}_0 := \delta_0$ and $\widehat{P}$.

Lemma

In the general setting and the above notation,

$$m_0 = \overline{m}_0 \overline{\Lambda} \qquad \text{and} \qquad \overline{\Lambda} P = \overline{P}\, \overline{\Lambda}.$$

This lemma can be used to give another proof of the general absorption-time theorem. But much more is possible when the eigenvalues and spectral polynomials of P are all nonnegative, as we shall henceforth assume.

"Intertwining" for the modified matrices

The following key "intertwining" equations for $\overline{m}_0$ and $\overline{P}$ follow by straightforward calculations from the corresponding equations for $\widehat{m}_0 := \delta_0$ and $\widehat{P}$.

Lemma

In the general setting and the above notation,

$$m_0 = \overline{m}_0 \overline{\Lambda} \qquad \text{and} \qquad \overline{\Lambda} P = \overline{P}\, \overline{\Lambda}.$$

This lemma can be used to give another proof of the general absorption-time theorem. But much more is possible when the eigenvalues and spectral polynomials of P are all nonnegative, as we shall henceforth assume.

"Intertwining" for the modified matrices

The following key "intertwining" equations for $\overline{m}_0$ and $\overline{P}$ follow by straightforward calculations from the corresponding equations for $\widehat{m}_0 := \delta_0$ and $\widehat{P}$.

Lemma

In the general setting and the above notation,

$$m_0 = \overline{m}_0 \overline{\Lambda} \qquad \text{and} \qquad \overline{\Lambda} P = \overline{P}\,\overline{\Lambda}.$$

This lemma can be used to give another proof of the general absorption-time theorem. But much more is possible when the eigenvalues and spectral polynomials of P are all nonnegative, as we shall henceforth assume.

Intertwining

Assume henceforth that the eigenvalues and spectral polynomials of P are all nonnegative. In that case we have the following conclusions (with all proofs entirely routine):

- $\overline{\Lambda}$ and $\overline{P}$ are both stochastic, and so the semigroups $(P^t)_{t\geq 0}$ and $(\overline{P}^t)_{t\geq 0}$ are intertwined (no quotes!) by the link $\overline{\Lambda}$.
- The set $\overline{A}$ of absorbing states for a chain $\overline{X}$ with kernel $\overline{P}$ satisfies $\overline{A} = \{\overline{d}, \dots, d\}$, where

$$\overline{d} := \min\{i : \lambda_i(d) = 1\} = \min\{i : a_i = 0\} - 1 \in \{0, \dots, d\},$$

  and $a_i = 0$ if and only if $i \geq \overline{d} + 1$.
- For a $\overline{P}$-chain, from each state in $\{0, \dots, \overline{d} - 1\}$ the states $\overline{d}$ and d are each accessible but none of the other states in $\overline{A}$ is.

Intertwining

Assume henceforth that the eigenvalues and spectral polynomials of P are all nonnegative. In that case we have the following conclusions (with all proofs entirely routine):

- $\overline{\Lambda}$ and $\overline{P}$ are both stochastic, and so the semigroups $(P^t)_{t \geq 0}$ and $(\overline{P}^t)_{t \geq 0}$ are intertwined (no quotes!) by the link $\overline{\Lambda}$.
- The set $\overline{A}$ of absorbing states for a chain $\overline{X}$ with kernel $\overline{P}$ satisfies $\overline{A} = \{\overline{d}, \dots, d\}$, where

$$\overline{d} := \min\{i : \lambda_i(d) = 1\} = \min\{i : a_i = 0\} - 1 \in \{0, \dots, d\},$$

and $a_i = 0$ if and only if $i \geq \overline{d} + 1$.
- For a $\overline{P}$-chain, from each state in $\{0, \dots, \overline{d} - 1\}$ the states $\overline{d}$ and d are each accessible but none of the other states in $\overline{A}$ is.

Intertwining

Assume henceforth that the eigenvalues and spectral polynomials of P are all nonnegative. In that case we have the following conclusions (with all proofs entirely routine):

- $\overline{\Lambda}$ and $\overline{P}$ are both stochastic, and so the semigroups $(P^t)_{t \geq 0}$ and $(\overline{P}^t)_{t \geq 0}$ are intertwined (no quotes!) by the link $\overline{\Lambda}$.
- The set $\overline{A}$ of absorbing states for a chain $\overline{X}$ with kernel $\overline{P}$ satisfies $\overline{A} = \{\overline{d}, \dots, d\}$, where

$$\overline{d} := \min\{i : \lambda_i(d) = 1\} = \min\{i : a_i = 0\} - 1 \in \{0, \dots, d\},$$

and $a_i = 0$ if and only if $i \geq \overline{d} + 1$.
- For a $\overline{P}$-chain, from each state in $\{0, \dots, \overline{d} - 1\}$ the states $\overline{d}$ and d are each accessible but none of the other states in $\overline{A}$ is.

Intertwining

Assume henceforth that the eigenvalues and spectral polynomials of P are all nonnegative. In that case we have the following conclusions (with all proofs entirely routine):

- $\overline{\Lambda}$ and $\overline{P}$ are both stochastic, and so the semigroups $(P^t)_{t \geq 0}$ and $(\overline{P}^t)_{t \geq 0}$ are intertwined (no quotes!) by the link $\overline{\Lambda}$.
- The set $\overline{A}$ of absorbing states for a chain $\overline{X}$ with kernel $\overline{P}$ satisfies $\overline{A} = \{\overline{d}, \dots, d\}$, where

$$\overline{d} := \min\{i : \lambda_i(d) = 1\} = \min\{i : a_i = 0\} - 1 \in \{0, \dots, d\},$$

  and $a_i = 0$ if and only if $i \geq \overline{d} + 1$.
- For a $\overline{P}$-chain, from each state in $\{0, \dots, \overline{d} - 1\}$ the states $\overline{d}$ and d are each accessible but none of the other states in $\overline{A}$ is.

Stochastic construction via the modified link

- The Diaconis–Fill construction allows us to build, from X and independent randomness, a chain $\overline{X}$ with initial distribution $\overline{m}$ and kernel $\overline{P}$ such that

$$\mathcal{L}(X_t \,|\, \overline{X}_0, \dots, \overline{X}_t) = \overline{\Lambda}(\overline{X}_t, \cdot) \quad \text{for all } t.$$

- The time T to absorption in state d for X is the same (sample-pathwise) as the time to absorption (call it $\overline{T}$) in $\overline{A}$ (i.e., in $\{\overline{d}, d\}$).
- Let L be the largest value reached by $\overline{X}$ prior to absorption, with the convention $L := -1$ if the initial state of $\overline{X}$ is d. Then $\mathbf{P}(L = k - 1) = a_k, \; k = 0, \dots, d$. Further, conditionally given $L = k - 1$, the amounts of time it takes for $\overline{X}$ to move up from 0 to 1, from 1 to 2, …, from $k-2$ to $k-1$, and from $k-1$ to $\overline{A}$ are independent geometric random variables with success probabilities $1 - \theta_0, 1 - \theta_1, \dots, 1 - \theta_{k-2}, 1 - \theta_{k-1}$.

Stochastic construction via the modified link

- The Diaconis–Fill construction allows us to build, from X and independent randomness, a chain $\overline{X}$ with initial distribution $\overline{m}$ and kernel $\overline{P}$ such that

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

- The time T to absorption in state d for X is the same (sample-pathwise) as the time to absorption (call it $\overline{T}$) in $\overline{A}$ (i.e., in $\{\overline{d}, d\}$).
- Let L be the largest value reached by $\overline{X}$ prior to absorption, with the convention $L := -1$ if the initial state of $\overline{X}$ is d. Then $\mathbf{P}(L = k-1) = a_k, \; k = 0, \ldots, d$. Further, conditionally given $L = k-1$, the amounts of time it takes for $\overline{X}$ to move up from 0 to 1, from 1 to 2, ..., from $k-2$ to $k-1$, and from $k-1$ to $\overline{A}$ are independent geometric random variables with success probabilities $1-\theta_0, 1-\theta_1, \ldots, 1-\theta_{k-2}, 1-\theta_{k-1}$.

Stochastic construction via the modified link

- The Diaconis–Fill construction allows us to build, from X and independent randomness, a chain $\overline{X}$ with initial distribution $\overline{m}$ and kernel $\overline{P}$ such that

$$\mathcal{L}(X_t \,|\, \overline{X}_0, \dots, \overline{X}_t) = \overline{\Lambda}(\overline{X}_t, \cdot) \quad \text{for all } t.$$

- The time T to absorption in state d for X is the same (sample-pathwise) as the time to absorption (call it $\overline{T}$) in $\overline{A}$ (i.e., in $\{\overline{d}, d\}$).
- Let L be the largest value reached by $\overline{X}$ prior to absorption, with the convention $L := -1$ if the initial state of $\overline{X}$ is d. Then $\mathbf{P}(L = k-1) = a_k$, $k = 0, \dots, d$. Further, conditionally given $L = k-1$, the amounts of time it takes for $\overline{X}$ to move up from 0 to 1, from 1 to 2, ..., from $k-2$ to $k-1$, and from $k-1$ to $\overline{A}$ are independent geometric random variables with success probabilities $1-\theta_0, 1-\theta_1, \dots, 1-\theta_{k-2}, 1-\theta_{k-1}$.

Stochastic construction via the modified link

Thus in the case of nonnegative eigenvalues and spectral polynomials we have enriched the conclusion of the general absorption-time theorem by means of a **stochastic construction** identifying

1. a random variable (namely, $L + 1$) having probability mass function (a_k); and, conditionally given $L + 1 = k$,
2. individual geometric random variables whose distributions appear in the convolution $\mathcal{G}(\theta_0, \ldots, \theta_{k-1})$.

Remark: In the Miclo (2008) case of both detailed balance and complete communication within $\{0, \ldots, d-1\}$ we have $\overline{d} = d$ and thus $\overline{A} = \{d\}$ is a singleton.

Stochastic construction via the modified link

Thus in the case of nonnegative eigenvalues and spectral polynomials we have enriched the conclusion of the general absorption-time theorem by means of a **stochastic construction** identifying

1. a random variable (namely, $L + 1$) having probability mass function (a_k); and, conditionally given $L + 1 = k$,
2. individual geometric random variables whose distributions appear in the convolution $\mathcal{G}(\theta_0, \ldots, \theta_{k-1})$.

Remark: In the Miclo (2008) case of both detailed balance and complete communication within $\{0, \ldots, d-1\}$ we have $\overline{d} = d$ and thus $\overline{A} = \{d\}$ is a singleton.

Stochastic construction via the modified link

Thus in the case of nonnegative eigenvalues and spectral polynomials we have enriched the conclusion of the general absorption-time theorem by means of a **stochastic construction** identifying

1. a random variable (namely, $L + 1$) having probability mass function (a_k); and, conditionally given $L + 1 = k$,
2. individual geometric random variables whose distributions appear in the convolution $\mathcal{G}(\theta_0, \ldots, \theta_{k-1})$.

Remark: In the Miclo (2008) case of both detailed balance and complete communication within $\{0, \ldots, d-1\}$ we have $\overline{d} = d$ and thus $\overline{A} = \{d\}$ is a singleton.

Stochastic construction via the modified link

Thus in the case of nonnegative eigenvalues and spectral polynomials we have enriched the conclusion of the general absorption-time theorem by means of a **stochastic construction** identifying

1. a random variable (namely, $L + 1$) having probability mass function (a_k); and, conditionally given $L + 1 = k$,
2. individual geometric random variables whose distributions appear in the convolution $\mathcal{G}(\theta_0, \ldots, \theta_{k-1})$.

Remark: In the Miclo (2008) case of both detailed balance and complete communication within $\{0, \ldots, d-1\}$ we have $\overline{d} = d$ and thus $\overline{A} = \{d\}$ is a singleton.