Introduction to set-valued intertwining duality for Markov processes

Laurent Miclo

Institut de Mathématiques de Toulouse
Toulouse School of Economics

Plan of the talk

1. Top-to-random shuffle [Aldous and Diaconis, 1986]
2. Strong stationary times and Markov intertwining relations
3. Set-valued duals
4. Dubins' example [1968] and Pitman's theorem [1975, Pitman and Rogers 1981]
5. One-dimensional diffusions
6. References

Plan of the talk

1. Top-to-random shuffle [Aldous and Diaconis, 1986]
2. Strong stationary times and Markov intertwining relations
3. Set-valued duals
4. Dubins' example [1968] and Pitman's theorem [1975, Pitman and Rogers 1981]
5. One-dimensional diffusions
6. References

Decks of cards

N cards labeled $1, 2, ..., N$. A deck of these cards is represented by an element σ of the symmetric group $\mathcal{S}_N$:

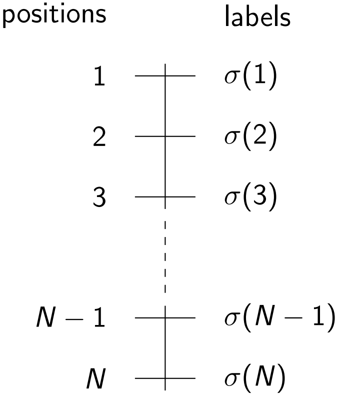

Figure: The deck of cards σ

Top-to-random shuffle

Starting from the ordered deck $(1, 2, \ldots, N)$, shuffle it as follows, where the positions I_1, I_2, I_3, ... are independent and uniformly distributed on $\{1, 2, \ldots, N\}$:

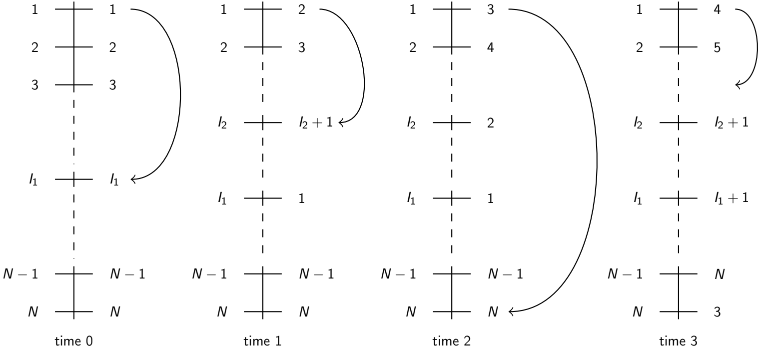

Figure: An example of top-to-random shuffle, here $I_3 = N$

Markov chains

Let $X(n) \in \mathcal{S}_N$ be the deck at time $n \in \mathbb{Z}_+$. The random sequence $X := (X(n))_{n\in\mathbb{Z}_+}$ is a **Markov chain**: to construct $X(n+1)$, only the knowledge of $X(n)$ (and of I_{n+1}, which is independent from the past) is needed, not the way we ended up with $X(n)$.

The uniform distribution $\mathcal{U}_{\mathcal{S}_N}$ on $\mathcal{S}_N$ is **invariant** for X: if $X(0) \sim \mathcal{U}_{\mathcal{S}_N}$, then for any time $n \in \mathbb{Z}_+$, $X(n) \sim \mathcal{U}_{\mathcal{S}_N}$. Furthermore, by irreducibility and aperiodicity, the law $\mathcal{L}(X(n))$ converges to $\mathcal{U}_{\mathcal{S}_N}$ for large time n.

But from a practical point of view, we want to know for how long we have to shuffle the deck to be close to $\mathcal{U}_{\mathcal{S}_N}$: we are looking for **quantitative convergence** estimates.

Quantitative convergence (1)

For the top-to-random Markov chain, there is a probabilistic way to get such an estimate. Look at the last card N at time 0. This card goes up, until it reaches the top of the deck, say at random time T. At time $\tau := T + 1$ the card N is sent to a uniformly chosen position.
At time τ the deck is uniformly distributed. Furthermore τ is a stopping time independent from the ordering of the deck $X(\tau)$.
This property implies that in total variation,

$$\forall\ n \in \mathbb{Z}_+, \qquad \|\mathcal{L}(X(n)) - \mathcal{U}_{\mathcal{S}_N}\|_{\mathrm{tv}} \leq \mathbb{P}[\tau > n]$$

To deduce a quantitative bound, observe that τ is distributed as a sum of independent geometric random variables of parameters $1/N$, $2/N$, ..., $(N-1)/N$ and 1.

Quantitative convergence (2)

In particular, we get

$$\begin{aligned} \mathbb{E}[\tau] &= N + \frac{N}{2} + \frac{N}{3} + \cdots + \frac{N}{N} \\ &\sim N \ln(N) \end{aligned}$$

which suggests that $N \ln(N)$ shuffles are needed to be at equilibrium.
Recognizing the coupon collector problem, one can be more precise: for any $c > 0$,

$$\|\mathcal{L}(X(\lfloor N \ln(N) + cN \rfloor)) - \mathcal{U}_{\mathcal{S}_N}\|_{\mathrm{tv}} \leq \exp(-c)$$

Going even further, one can deduce an instance of the **cut-off** phenomenon at time $N \ln(N)$, in separation and total variation.

Plan of the talk

The above considerations extend into a much more general theory, see in particular [Diaconis and Fill, 1990].
Consider $X := (X(n))_{n\in\mathbb{Z}_+}$ a Markov chain taking values in a finite state space V and admitting a unique invariant probability π. A finite stopping time τ (with respect to the filtration generated by X and some independent randomness) is said to be a **strong stationary time** when

$$\tau \perp\!\!\!\perp X(\tau) \quad \text{and} \quad X(\tau) \sim \pi$$

An example is provided by the first time the last card is inserted in the deck in the top-to-random model.
Strong stationary times enable probabilistic exact simulations of π.

Separation

The **separation** discrepancy between two probability measures μ and π on the same state space is defined by:

$$\mathfrak{s}(\mu,\pi) := \operatorname*{ess\,sup}_{\mu} 1 - \frac{d\mu}{d\pi} \geq \|\mu-\pi\|_{\mathrm{tv}}$$

(recall the total variation: $\|\mu-\pi\|_{\mathrm{tv}} := \sup_{A \text{ event}} |\mu(A)-\pi(A)|$, coinciding with $\frac{1}{2}\left\|\frac{d\mu}{d\pi}-1\right\|_{\mathbb{L}^1(\pi)}$ when $\mu \ll \pi$).

A general property of any strong stationary time τ:

$$\forall\ n \in \mathbb{Z}_+, \qquad \mathfrak{s}(\mathcal{L}(X(n)),\pi) \quad \leq \quad \mathbb{P}[\tau > n]$$

Furthermore, as soon as X is ergodic, there exist strong stationary times τ such that all these inequalities are equalities, and such τ are said to be a **sharp strong stationary time**.

Intertwining relations (1)

Denote P the transition matrix of X on V and consider a transition matrix Q on another finite state space W, as well a Markov kernel Λ from W to V, called the **link**. There is an **algebraic intertwining Markov relation** from Q to P when the following commutation relation holds:

$$Q\Lambda = \Lambda P \tag{1}$$

Let $Y := (Y(n))_{n\in\mathbb{Z}_+}$ be a Markov chain whose transitions are dictated by Q. When in addition to (1), we have

$$\mathcal{L}(X(0)) = \mathcal{L}(Y(0))\Lambda$$

we say there is a **algebraic intertwining Markov relation** from Y to X. The Markov chain Y is also called an **intertwining dual** for the **primal** Markov chain X.

Intertwining relations (2)

Assume there is an algebraic intertwining Markov relation from Y to X. Then there exists a coupling of X and Y such that for all times $n \in \mathbb{Z}_+$, we have for the conditional expectations:

$$\mathcal{L}(X(n)|Y([\![0, n]\!])) = \Lambda(Y(n), \cdot) \quad (2)$$

$$\mathcal{L}(Y([\![0, n]\!])|X) = \mathcal{L}(Y([\![0, n]\!])|X([\![0, n]\!])) \quad (3)$$

where $Y([\![0, n]\!])$ stands for the partial trajectory $(Y(l))_{l \in [\![0,n]\!]}$.
The equalities (3) imply that Y can be constructed from X in an "adapted" way. In particular any stopping time for Y is also a stopping for X under such a coupling.
In this situation we say that Y and X satisfies a **probabilistic intertwining Markov relation** or that they are **coupled by intertwining**.

Absorbed intertwining duals

Assume that there exists an **absorbing state** $\omega \in W$ for Q, $Q(\omega, \omega) = 1$, and that furthermore $\{\omega\}$ is the unique recurrence class for Q. In particular, whatever the initial distribution of $Y(0)$, the absorbing time

$$\tau \quad := \quad \inf\{t \in \mathbb{Z}_+ \ : \ Y(n) = \omega\}$$

is a.s. finite.
When X and Y are coupled by intertwining, then the absorbing time τ is a strong stationary time for X. Indeed, an important observation is that $\Lambda(\omega, \cdot) = \pi$, coming from (1) applied at ω.
This is probably the most common method to construct strong stationary times.

Plan of the talk

1. Top-to-random shuffle [Aldous and Diaconis, 1986]
2. Strong stationary times and Markov intertwining relations
3. Set-valued duals
4. Dubins' example [1968] and Pitman's theorem [1975, Pitman and Rogers 1981]
5. One-dimensional diffusions
6. References

Set-valued duals

A particular instance of the previous framework corresponds to

$$\begin{array}{rrcl} & W & := & \{S \subset V,\, S \neq \emptyset\} \\ \forall\, S \in W, & \Lambda(S,\cdot) & = & \dfrac{\pi(S \cap \cdot)}{\pi(S)} \end{array}$$

at least when π is positive (typically for irreducible Markov chains X). The corresponding dual chain Y is then set-valued.
The whole set V usually turns out to be absorbing for Y and we recover that $\Lambda(V,\cdot) = \pi$.
The top-to-random example can be written under this form, by considering cylindrical subsets of $\mathcal{S}_N$ corresponding to decks of cards coinciding with the current deck on the positions above the last card.
Diaconis and Fill [1990] have shown on a three-point example, that sharp set-valued duals may not exist.

Evolving sets [Morris and Peres, 2005]

Assume that π is positive and consider P^* the adjoint Markov kernel of P is $\mathbb{L}^2(\pi)$ (it is the transition kernel of the time-reversed Markov chain of X at equilibrium):

$$\forall\, x, y \in V, \qquad P^*(x, y) \;\; := \;\; \frac{\pi(y)}{\pi(x)} P(y, x)$$

A transition kernel J on $\bar{W} := W \sqcup \{\emptyset\}$ is constructed as follows: given $S \subset V$ and a random variable U uniformly distributed on $[0, 1]$, consider the random subset

$$\Phi(S) \;\; := \;\; \{y \in V \; : \; P^*(y, S) \geq U\}$$

We define

$$\forall\, S, S' \in \bar{W}, \qquad J(S, S') \;\; := \;\; \mathbb{P}[\Phi(S) = S']$$

The **evolving set processes** of Morris and Peres are the $\bar{W}$-valued Markov chains $Z := (Z(n))_{n \in \mathbb{Z}_+}$ whose transition kernel is J.

Deduction of a set-valued dual

The Markov chain Z is absorbed at $\emptyset$ and at V. The mapping $\pi : \bar{W} \ni S \mapsto \pi(S)$ is harmonic for J, in the sense that $J\pi = \pi$. It leads to consider the **Doob transform** of J by π, which is the $W \times W$ transition matrix Q given by

$$\forall\ S, S' \in W, \qquad Q(S, S') \ := \ \frac{\pi(S')}{\pi(S)} J(S, S')$$

It is the transition kernel of Z conditioned not to be absorbed at $\emptyset$.
Let $Y := (Y(n))_{n \in \mathbb{Z}_+}$ be a Markov chain whose transition kernel is Q and such that $\mathcal{L}(Y(0))\Lambda = \mathcal{L}(X(0))$.
Then Y is a set-valued intertwining dual for X.

Random mappings (1)

The previous construction can be generalised via an approach similar to the coupling-from-the-past algorithm of Propp and Wilson [1996]. For any $S \in \bar{W}$, let be given a random mapping $\psi_S : V \to V$ such that

$$\forall\, x \in V,\ \forall\, x' \in S, \qquad \mathbb{P}[\psi_S(x) = x'] \;=\; \frac{P^*(x, x')}{\zeta(S)}$$

where ζ is a positive mapping on $\bar{W}$. The family $(\psi_S)_{S\in\bar{W}}$ enables to define a random mapping Ψ from $\bar{W}$ to $\bar{W}$ via

$$\forall\, S \in \bar{W}, \qquad \Psi(S) \;:=\; \{y \in V \,:\, \psi_S(y) \in S\}$$

Consider the transition matrix K from $\bar{W}$ to $\bar{W}$ given by

$$\forall\, S, S' \in \bar{W}, \qquad K(S, S') \;:=\; \mathbb{P}[\Psi(S) = S']$$

and its modified Doob transform via

$$\forall\, S, S' \in W, \qquad Q(S, S') \;:=\; \frac{\pi(S')\zeta(S)}{\pi(S)} K(S, S')$$

Random mappings (2)

Introduce the following conditioned transition: for $x, x' \in V$ such that $P(x, x') > 0$ and any $S \in W$ containing x,

$$\forall\ S' \in W, \qquad K_{x,x'}(S, S') \ := \ \mathbb{P}[\Psi(S) = S' | \psi_S(x') = x]$$

Consider

$$\mathcal{W} \ := \ \{(x, S) \in V \times W \ : \ x \in S\}$$

and let $\mathcal{P}$ be the set of probability measures m on $\mathcal{W}$ which can be written under the form

$$\forall\ (x, S) \in W, \qquad m(x, S) \ = \ \mu(S)\Lambda(S, x)$$

where μ is the marginal of m on $\mathcal{W}$.

Random mappings (3)

Define a Markov kernel $\mathcal{Q}$ on $\mathcal{W}$ via

$$\forall\ (x,S),\ (x',S') \in \mathcal{W}, \quad \mathcal{Q}((x,S),(x',S')) \ \coloneqq\ P(x,x')K_{x,x'}(S,S')$$

Theorem

Let $(X(n),Y(n))_{n\in\mathbb{Z}_+}$ be a Markov chain on $\mathcal{W}$ whose initial distribution $\mathcal{L}(X_0,Y_0)$ belongs to $\mathcal{P}$ and whose transitions are given by $\mathcal{Q}$. Then $X \coloneqq (X(n))_{n\in\mathbb{N}}$ and $Y \coloneqq (Y(n))_{n\in\mathbb{N}}$ are Markov chains whose respective transitions are given by P and Q and Y is a set-valued intertwining dual for X.

This result was used to construct strong stopping times for random walks on discrete Heisenberg groups.

Plan of the talk

Dubins' example [1968]

Consider for X the Brownian motion on $[0, 1]$, reflected at 0 and 1 and starting from $1/2$. It is positive recurrent with the restriction of the Lebesgue measure as invariant probability. A strong stationary time can be constructed as follows: let τ_1 be the first time X hits $1/4$ or $3/4$. Next let τ_2 be the first time after τ_1 that $X(\tau_1) \pm 1/8$ is reached. Iteratively, τ_{n+1} is the first time after τ_n that $X(\tau_n) \pm 1/2^{n+2}$ is hit. The limit $\tau := \lim_{n\to\infty} \tau_n$ exists a.s. and is a strong stationary time for X.

This construction can be extended to any initial distribution, by first waiting that $1/2$ is reached (not always smart, for instance if $X(0)$ was already at equilibrium).

Pitman's transformation

Let $B := (B(t))_{t \geq 0}$ be a standard Brownian motion. Consider the Pitman's transformation $R := (R(t))_{t \geq 0}$ given by

$$R(t) \quad := \quad -B(t) + 2 \max_{s \in [0,t]} B(s)$$

Let Λ the Markov kernel from $\mathbb{R}_+$ to $\mathbb{R}$ given by

$$\forall\ r \geq 0, \qquad \Lambda(r, dx) \quad := \quad \mathcal{U}_{[-r,r]}(dx)$$

Relations (2) and (3) can be extended to this continuous setting: for all time $t \geq 0$,

$$\begin{cases} \mathcal{L}(B(t)|R([0,t])) & = & \Lambda(R(t), \cdot) \\ \mathcal{L}(R([0,t])|B) & = & \mathcal{L}(R([0,t])|B([0,t])) \end{cases} \tag{4}$$

Furthermore, the process R is a **Bessel-3 process**, i.e. it has the law of the norm of a Brownian motion in dimension 3.

Pitman's theorem in pictures

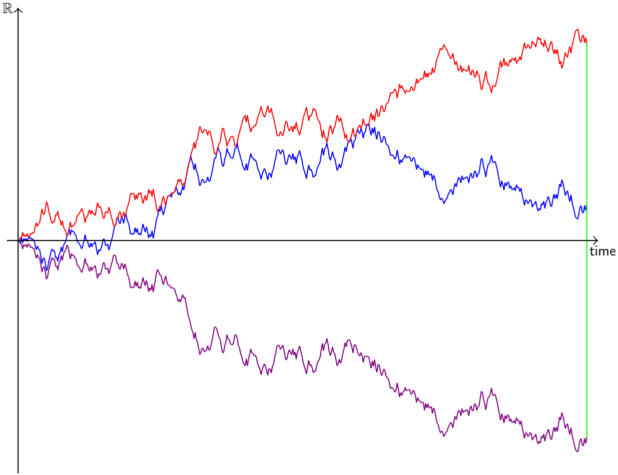

Figure: Trajectories: Brownian motion $B([0,t])$, $R([0,t])$, $-R([0,t])$, and the segment-valued dual: $[-R(t), R(t)]$

Consequently of (4), we get estimates on the convergence of the Brownian motion $W := (W(t))_{t\geq 0}$ on the circle $\mathbb{T} := \mathbb{R}/(2\pi\mathbb{Z})$: let τ be the hitting time of π by R:

$$\tau \quad := \quad \inf\{t \geq 0 \,:\, R(t) = \pi\}$$

It is a strong stationary time for W. The hitting times of Bessel processes are well-studied, we have:

$$\forall\ \lambda > 0, \qquad \mathbb{E}[\exp(-\lambda\tau)] \quad = \quad \frac{\sqrt{\pi}\sqrt[4]{\lambda}}{\sqrt[4]{2}\Gamma(3/2)} \frac{1}{I_{1/2}(\pi\sqrt{2\lambda})}$$

where $I_{1/2}$ is the modified Bessel function of index $1/2$.
A Tauberien theorem enables to deduce the behavior for large $t \geq 0$ of $\mathbb{P}[\tau > t]$ and thus of $\mathfrak{s}(\mathcal{L}(W(t)), \mathcal{U}_{\mathbb{T}})$ and $\|\mathcal{L}(W(t)) - \mathcal{U}_{\mathbb{T}}\|_{\mathrm{tv}}$.

Plan of the talk

1. Top-to-random shuffle [Aldous and Diaconis, 1986]
2. Strong stationary times and Markov intertwining relations
3. Set-valued duals
4. Dubins' example [1968] and Pitman's theorem [1975, Pitman and Rogers 1981]
5. One-dimensional diffusions
6. References

Elliptic real diffusions

On $\mathbb{R}$, consider a diffusion X whose generator is $G := a\partial^2 + b\partial$, where a, b are smooth functions and a is positive. Consider the **speed measure** $\mu(dx) := \underline{\mu}(x)\, dx$ with density

$$\underline{\mu}(x) := \frac{\exp\left(c(x)\right)}{a(x)} \quad \text{with} \quad c(x) := \int_0^x \frac{b(y)}{a(y)}\, dy$$

It is invariant for X and even **reversible**: G can be extended into a self-adjoint operator on $\mathbb{L}^2(\mu)$.
Consider

$$\mathcal{D} \quad := \quad \{[y,z] \,:\, y,z \in (-\infty,+\infty),\, y \leq z\}$$

As usual, define the associated Markov kernel Λ from $\mathcal{D}$ to $\mathbb{R}$ via

$$\forall\ [y,z] \in \mathcal{D},\ \Lambda([y,z],dx) := \begin{cases} \delta_y(dx) & \text{, if } y = z \\ \frac{\underline{\mu}(x)}{\mu([y,z])} \mathbb{1}_{[y,z]}(x)\, dx & \text{, otherwise} \end{cases}$$

Dual process

On $\mathcal{D}$ define the degenerate generator

$$\begin{aligned}\mathcal{G} \quad := \quad & (\sqrt{a(z)}\partial_z - \sqrt{a(y)}\partial_y)^2 \\ & +(a'(y)/2 - b(y))\partial_y + (a'(z)/2 - b(z))\partial_z \\ & +2\frac{\sqrt{a(y)}\underline{\mu}(y) + \sqrt{a(z)}\underline{\mu}(z)}{\mu([y,z])}(\sqrt{a(z)}\partial_z - \sqrt{a(y)}\partial_y),\end{aligned}$$

Its interest is in the intertwining relation $\mathcal{G}\Lambda = \Lambda G$.

Proposition

There exists a unique process $[Y, Z]$ whose generator is $\mathcal{G}$, up to its explosion time ζ. The diagonal is an entrance boundary for this process.

The proof is based on the fact that $(\mu([Y(t), Z(t)]))_{t\in[0,\zeta]}$ is a (stopped) Bessel-3 process, up to the time-change $(\theta(t))_{t\in[0,\varsigma]}$ given by

$$2\int_0^{\theta(t)} (\sqrt{a(Y(s))}\underline{\mu}(Y(s)) + \sqrt{a(Z(s))}\underline{\mu}(Z(s)))^2\, ds \quad = \quad t$$

Strong stationary times in 1-dimension

Application to the existence of strong stationary times:

Theorem

Assume that X is positive recurrent. There exists a strong stationary time for X, whatever its initial distribution, if and only if $-\infty$ and $+\infty$ are entrance boundaries.

Positive recurrence or **ergodicity**: μ is finite and in large time $X(t)$ converges to the renormalization of μ, analytically this is characterized by

$$\int_{-\infty}^{0} \exp(-c(y))\, dy = +\infty \quad \text{and} \quad \int_{0}^{\infty} \exp(-c(y))\, dy = +\infty$$

Entrance boundary: e.g. for $+\infty$, it means that X can be started from $+\infty$ (comes down from infinity), it amounts to:

$$\int_{0}^{+\infty} \left(\int_{0}^{x} \exp\left(-c(y)\right)\, dy \right) \mu(dx) < +\infty$$

1-dimension hypoelliptic diffusions

Consider again the generator $G := a\partial^2 + b\partial$, on $\mathbb{R}$ or $\mathbb{T}$, where a is allowed to vanish on a finite number of points, but such that $\sqrt{a}$ remains smooth. On the vanishing points, assume b does not vanish: 1-dimensional hypoellipticity in the sense of [Hörmander, 1967].

Up to some adjustments (modification of the measure used in Λ, dual processes which can disconnect, ...), the above approach is valid and enables to recover the density theorem (i.e. the law of $X(t)$ admits a density for all $t > 0$) and to obtain estimates on the speed of convergence to equilibrium (even when the invariant measure does not charge the whole state space). The Bessel-3 process is still there, the hypoellipticity is only felt at the level of the time change.

Is it possible to recover the generality of Hörmander's theorem in this probabilistic way?

Plan of the talk

References for the finite framework (1)

- David Aldous and Persi Diaconis. Shuffling cards and stopping times. *Amer. Math. Monthly*, 93(5):333–348, 1986.
- Persi Diaconis and James Allen Fill. Strong stationary times via a new form of duality. *Ann. Probab.*, 18(4):1483–1522, 1990.
- Ben Morris and Yuval Peres. Evolving sets, mixing and heat kernel bounds. *Probab. Theory Related Fields*, 133(2):245–266, 2005.
- David A. Levin, Yuval Peres, and Elizabeth L. Wilmer. *Markov chains and mixing times*. American Mathematical Society, Providence, RI, 2009. With a chapter by James G. Propp and David B. Wilson.

References for the finite framework (2)

James Gary Propp and David Bruce Wilson. Exact sampling with coupled Markov chains and applications to statistical mechanics. In *Proceedings of the Seventh International Conference on Random Structures and Algorithms (Atlanta, GA, 1995)*, volume 9, pages 223–252, 1996.

Laurent Miclo. On the construction of measure-valued dual processes. *Electron. J. Probab.*, 25:1–64, 2020.

Miclo, Laurent. Strong stationary times for finite Heisenberg walks. *ESAIM: PS*, 27:515–557, 2023.

References for the one-dimensional diffusion case (1)

- Lester E. Dubins. On a theorem of Skorohod. *Ann. Math. Statist.*, 39:2094–2097, 1968.
- James W. Pitman. One-dimensional Brownian motion and the three-dimensional Bessel process. *Advances in Appl. Probability*, 7(3):511–526, 1975.
- L. Chris G. Rogers and James W. Pitman. Markov functions. *Ann. Probab.*, 9(4):573–582, 1981.

References for the one-dimensional diffusion case (2)

- James Allen Fill and Vince Lyzinski. *Strong stationary duality for diffusion processes*, J. Theoret. Probab. **29** (2016), no. 4, 1298–1338.
- Laurent Miclo. Strong stationary times for one-dimensional diffusions. *Ann. Inst. Henri Poincaré Probab. Stat.*, 53(2):957–996, 2017.
- Lars Hörmander. Hypoelliptic second order differential equations, Acta Math. **119** (1967), 147–171.
- Laurent Miclo. Duality and hypoellipticity: one-dimensional case studies. *Electron. J. Probab.*, 22:Paper No. 91, 32, 2017.

References on Markov intertwinings

- David Aldous and Persi Diaconis. Strong uniform times and finite random walks. *Adv. in Appl. Math.*, 8(1):69–97, 1987.
- Philippe Carmona, Frédérique Petit and Marc Yor. Beta-gamma random variables and intertwining relations between certain Markov processes. *Rev. Mat. Iberoamericana*, 14(2):311–367, 1998.
- Alexei Borodin and Grigori Olshanski. Markov processes on the path space of the Gelfand-Tsetlin graph and on its boundary. *J. Funct. Anal.*, 263(1):248–303, 2012.
- Soumik Pal and Mykhaylo Shkolnikov. Intertwining diffusions and wave equations. *ArXiv e-prints*, June 2013.