Processes with reinforcement and approximation of Quasi-Stationary Distributions

Michel Benaim
Neuchâtel University

Luminy, September 2018

Talk based on recent collaborations with

Talk based on recent collaborations with

Bertrand Cloez (Montpellier)

and **Fabien Panloup** (Anger)

Talk based on recent collaborations with

Bertrand Cloez (Montpellier)

and **Fabien Panloup** (Anger)

Stochastic Approximation of Quasi-Stationary Distributions on Compact Spaces and Applications Annals of Applied Probability, 2018

and on (more) recent works by

and on (more) recent works by

• A. Wang (Oxford) , G. O Roberts (Warwick) and D. Steinsaltz (Oxford)

and on (more) recent works by

• A. Wang (Oxford) , G. O Roberts (Warwick) and D. Steinsaltz (Oxford)

An approximation Scheme for Quasistationary Distributions of Killed Diffusions arXiv August 23, 2018

and on (more) recent works by

• A. Wang (Oxford) , G. O Roberts (Warwick) and D. Steinsaltz (Oxford)

An approximation Scheme for Quasistationary Distributions of Killed Diffusions arXiv August 23, 2018

• C. Mailler (Bath) and D. Villemonais (Nancy)

and on (more) recent works by

• A. Wang (Oxford) , G. O Roberts (Warwick) and D. Steinsaltz (Oxford)

An approximation Scheme for Quasistationary Distributions of Killed Diffusions arXiv August 23, 2018

• C. Mailler (Bath) and D. Villemonais (Nancy)

Stochastic Approximation on non-compact measure spaces and application to measure-valued Polya Processes arXiv September 6, 2018

Context

Context

- $\mathcal{E}$ a metric space, ∂ a *cemetery* point

Context

- $\mathcal{E}$ a metric space, ∂ a *cemetery* point
- $(X_t)_{t \in \mathbb{Z}^+}$ a Markov chain on $\mathcal{E} \cup \partial$ eventually absorbed by ∂ :

Context

- $\mathcal{E}$ a metric space, ∂ a *cemetery* point
- $(X_t)_{t\in\mathbb{Z}^+}$ a Markov chain on $\mathcal{E}\cup\partial$ eventually absorbed by ∂ :
  (i) $\tau_\partial = \inf\{t \geq 0 : X_t = \partial\} < \infty$ a.s
  (ii) $X_t = \partial \Rightarrow X_{t+s} = \partial.$

Context

- $\mathcal{E}$ a metric space, ∂ a *cemetery* point
- $(X_t)_{t \in \mathbb{Z}^+}$ a Markov chain on $\mathcal{E} \cup \partial$ eventually absorbed by ∂ :
  - (i) $\tau_\partial = \inf\{t \geq 0 : X_t = \partial\} < \infty$ a.s
  - (ii) $X_t = \partial \Rightarrow X_{t+s} = \partial$.

Definition

A *Quasi-Stationary Distribution* for X is a probability μ on $\mathcal{E}$ such that

$$\mathbb{P}_\mu(X_t \in \cdot | t < \tau_\delta) = \mu(\cdot).$$

Context

- $\mathcal{E}$ a metric space, ∂ a *cemetery* point
- $(X_t)_{t \in \mathbb{Z}^+}$ a Markov chain on $\mathcal{E} \cup \partial$ eventually absorbed by ∂ :
  (i) $\tau_\partial = \inf\{t \geq 0 : X_t = \partial\} < \infty$ a.s
  (ii) $X_t = \partial \Rightarrow X_{t+s} = \partial.$

Definition

A *Quasi-Stationary Distribution* for X is a probability μ on $\mathcal{E}$ such that
$$\mathbb{P}_\mu(X_t \in \cdot | t < \tau_\delta) = \mu(\cdot).$$

Under appropriate assumptions, such a QSD exists and

$$\lim_{t \to \infty} \mathbb{P}(X_t \in \cdot | t < \tau_\delta) = \mu(\cdot)$$

Context

- $\mathcal{E}$ a metric space, ∂ a *cemetery* point
- $(X_t)_{t \in \mathbb{Z}^+}$ a Markov chain on $\mathcal{E} \cup \partial$ eventually absorbed by ∂ :
  (i) $\tau_\partial = \inf\{t \geq 0 : X_t = \partial\} < \infty$ a.s
  (ii) $X_t = \partial \Rightarrow X_{t+s} = \partial$.

Definition

A *Quasi-Stationary Distribution* for X is a probability μ on $\mathcal{E}$ such that

$$\mathbb{P}_\mu(X_t \in \cdot | t < \tau_\delta) = \mu(\cdot).$$

Under appropriate assumptions, such a QSD exists and

$$\lim_{t \to \infty} \mathbb{P}(X_t \in \cdot | t < \tau_\delta) = \mu(\cdot)$$

Natural object in Population Dynamics

Context

- $\mathcal{E}$ a metric space, ∂ a *cemetery* point
- $(X_t)_{t\in\mathbb{Z}^+}$ a Markov chain on $\mathcal{E}\cup\partial$ eventually absorbed by ∂ :
  - (i) $\tau_\partial = \inf\{t \geq 0 : \; X_t = \partial\} < \infty$ a.s
  - (ii) $X_t = \partial \Rightarrow X_{t+s} = \partial.$

Definition

A *Quasi-Stationary Distribution* for X is a probability μ on $\mathcal{E}$ such that

$$\mathbb{P}_\mu(X_t \in \cdot | t < \tau_\delta) = \mu(\cdot).$$

Under appropriate assumptions, such a QSD exists and

$$\lim_{t\to\infty} \mathbb{P}(X_t \in \cdot | t < \tau_\delta) = \mu(\cdot)$$

Natural object in Population Dynamics because eventually everyone gets killed...

How to compute/estimate such a QSD ?

- A Classical approach is based on *Fleming Viot Algorithms*

How to compute/estimate such a QSD ?

- A Classical approach is based on *Fleming Viot Algorithms*
  (i) Run N independent particles until one is killed,

How to compute/estimate such a QSD ?

- A Classical approach is based on *Fleming Viot Algorithms*
  - (i) Run N independent particles until one is killed,
  - (ii) replace the killed particle by an offspring whose location is randomly chosen according the spatial occupation measure of the other particles

How to compute/estimate such a QSD ?

- A Classical approach is based on *Fleming Viot Algorithms*
  - (i) Run N independent particles until one is killed,
  - (ii) replace the killed particle by an offspring whose location is randomly chosen according the spatial occupation measure of the other particles

Then (under appropriate assumptions)

$$\lim_{N\to\infty}\lim_{t\to\infty}\frac{1}{N}\sum_{i=1}^{N}\delta_{X_t^i}=\mu.$$

How to compute/estimate such a QSD ?

- A Classical approach is based on *Fleming Viot Algorithms*
  - (i) Run N independent particles until one is killed,
  - (ii) replace the killed particle by an offspring whose location is randomly chosen according the spatial occupation measure of the other particles

Then (under appropriate assumptions)

$$\lim_{N\to\infty} \lim_{t\to\infty} \frac{1}{N} \sum_{i=1}^{N} \delta_{X_t^i} = \mu.$$

$\rightsquigarrow$ Idea explored by Burdzy, Holyst & March (2000); Del Moral & Miclo (2000); Villemonais (2014); Cloez & Thai (2016)...

• Alternative approach : Use a single particle and replace the spatial occupation measure by the temporal occupation measure

$$\mu_t = \frac{1}{t} \sum_{s<t} \delta_{X_s}$$

• Alternative approach : Use a single particle and replace the spatial occupation measure by the temporal occupation measure

$$\mu_t = \frac{1}{t} \sum_{s<t} \delta_{X_s}$$

(i) Run (X_t) until it gets killed,

• Alternative approach : Use a single particle and replace the spatial occupation measure by the temporal occupation measure

$$\mu_t = \frac{1}{t} \sum_{s<t} \delta_{X_s}$$

(i) Run (X_t) until it gets killed,

(ii) Resuscitate it at a location randomly chosen according to μ_t

• Alternative approach : Use a single particle and replace the spatial occupation measure by the temporal occupation measure

$$\mu_t = \frac{1}{t} \sum_{s<t} \delta_{X_s}$$

(i) Run (X_t) until it gets killed,

(ii) Resuscitate it at a location randomly chosen according to μ_t

$\rightsquigarrow$ This original idea goes back to Aldous, Flannery & Palacios (1988) and their approach relies on branching processes.

• Alternative approach : Use a single particle and replace the spatial occupation measure by the temporal occupation measure

$$\mu_t = \frac{1}{t} \sum_{s<t} \delta_{X_s}$$

(i) Run (X_t) until it gets killed,

(ii) Resuscitate it at a location randomly chosen according to μ_t

⇝ This original idea goes back to Aldous, Flannery & Palacios (1988) and their approach relies on branching processes.

⇝ Here we will revisit it using tools from *stochastic approximation, self-reinforced processes* combined with recent ideas & results due to Champagnat and Villemonais (2015)

Notation & Hypotheses

- K the Sub-Markov Kernel of X on $\mathcal{E}$:

$$K(x,\cdot) = \mathbb{P}_x(X_1 \in \cdot)$$

Notation & Hypotheses

- K the Sub-Markov Kernel of X on $\mathcal{E}$:

$$K(x, \cdot) = \mathbb{P}_x(X_1 \in \cdot)$$

- **H1** $\mathcal{E}$ is compact and K Feller;

Notation & Hypotheses

- K the Sub-Markov Kernel of X on $\mathcal{E}$:

$$K(x, \cdot) = \mathbb{P}_x(X_1 \in \cdot)$$

- **H1** $\mathcal{E}$ is compact and K Feller;
- **H2** ∂ is *accessible* :

$$\forall x \, \exists n \geq 0 \ : K^n(x, \partial) < 1;$$

Notation & Hypotheses

- K the Sub-Markov Kernel of X on $\mathcal{E}$:

$$K(x, \cdot) = \mathbb{P}_x(X_1 \in \cdot)$$

- **H1** $\mathcal{E}$ is compact and K Feller;
- **H2** ∂ is *accessible* :

$$\forall x \, \exists n \geq 0 \; : K^n(x, \partial) < 1;$$

Remark

H1&H2 $\Rightarrow$ *Existence of (at least) one QSD*

- **H3** There exists an open set $U \subset \partial$ which is *accessible:*

- **H3** There exists an open set $U \subset \partial$ which is *accessible:*

$$\forall x \sum_n K^n(x, U) > 0.$$

• **H3** There exists an open set $U \subset \partial$ which is *accessible:*

$$\forall x \sum_n K^n(x, U) > 0.$$

and *petite* :

• **H3** There exists an open set $U \subset \partial$ which is *accessible:*

$$\forall x \sum_n K^n(x, U) > 0.$$

and *petite* :

$$\forall x \in U \sum_n K^n(x, \cdot) \geq \epsilon \Psi(\cdot)$$

where Ψ is a probability on $\mathcal{E}$

• **H3** There exists an open set $U \subset \partial$ which is *accessible:*

$$\forall x \sum_n K^n(x, U) > 0.$$

and *petite* :

$$\forall x \in U \sum_n K^n(x, \cdot) \geq \epsilon \Psi(\cdot)$$

where Ψ is a probability on $\mathcal{E}$

Remark

If K were Markov (i.e $K(x, \mathcal{E}) = 1$), $H1, H2, H3$ would ensure the uniqueness of an invariant measure μ But, this is not sufficient to ensure uniqueness of a QSD !

The Champagnat-Villemonais condition

• **H4** There exists a non increasing convex function $C : \mathbb{R}^+ \mapsto \mathbb{R}^+$ satisfying

$$\int_0^\infty C(s)ds = \infty$$

such that

$$\frac{\Psi(K^n\mathbf{1})}{\sup_{x\in\mathcal{E}} K^n\mathbf{1}(x)} \geq C(n)$$

The Champagnat-Villemonais condition

- **H4** There exists a non increasing convex function $C : \mathbb{R}^+ \mapsto \mathbb{R}^+$ satisfying

$$\int_0^\infty C(s)ds = \infty$$

such that

$$\frac{\Psi(K^n \mathbf{1})}{\sup_{x \in \mathcal{E}} K^n \mathbf{1}(x)} \geq C(n)$$

Remark

- *When $C(n) = C$, this condition is due to Champagnat and Villemonais (2015)*

The Champagnat-Villemonais condition

• **H4** There exists a non increasing convex function $C : \mathbb{R}^+ \mapsto \mathbb{R}^+$ satisfying

$$\int_0^\infty C(s)ds = \infty$$

such that

$$\frac{\Psi(K^n \mathbf{1})}{\sup_{x\in\mathcal{E}} K^n\mathbf{1}(x)} \geq C(n)$$

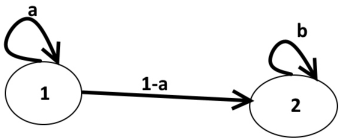
Remark

• *When $C(n) = C$, this condition is due to Champagnat and Villemonais (2015)*

• *It ensures the uniqueness of the QSD.*

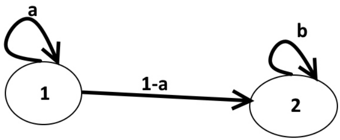
Example

$a, b \in (0, 1)$



Example

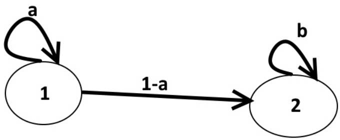
$a, b \in (0, 1)$



- If $a < b$, $C(t) = C$ and there is a unique QSD $\mu = \delta_2$

Example

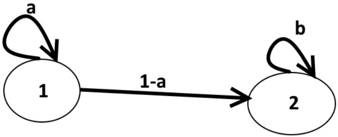
$a, b \in (0,1)$



- If $a < b$, $C(t) = C$ and there is a unique QSD $\mu = \delta_2$
- If $a = b$, $C(t) = \frac{C}{1+t}$ and there is a unique QSD $\mu = \delta_2$

Example

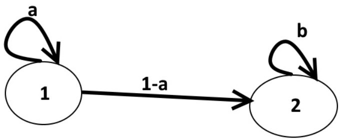
$a, b \in (0, 1)$



- If $a < b$, $C(t) = C$ and there is a unique QSD $\mu = \delta_2$
- If $a = b$, $C(t) = \frac{C}{1+t}$ and there is a unique QSD $\mu = \delta_2$
- If $a > b$ there is another QSD $\mu^\star = \frac{a-b}{1-b}\delta_1 + \frac{1-a}{1-b}\delta_2$.

Example

$a, b \in (0,1)$



- If $a < b$, $C(t) = C$ and there is a unique QSD $\mu = \delta_2$
- If $a = b$, $C(t) = \frac{C}{1+t}$ and there is a unique QSD $\mu = \delta_2$
- If $a > b$ there is another QSD $\mu^\star = \frac{a-b}{1-b}\delta_1 + \frac{1-a}{1-b}\delta_2$.

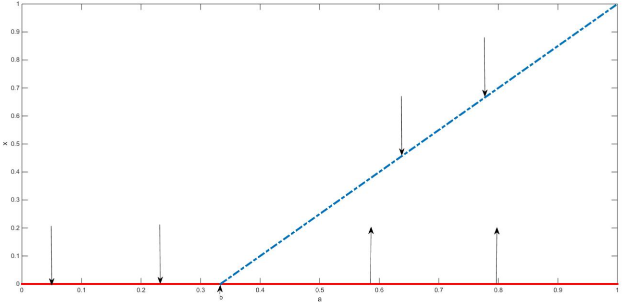


Figure: $b = 1/3; a \mapsto \mu(1), \mu^*(1)$

Results

- For each μ probability on $\mathcal{E}$

$$K_\mu(x, dy) = K(x, dy) + (1 - K(x, \mathcal{E}))\mu(dy)$$

Results

- For each μ probability on $\mathcal{E}$

$$K_\mu(x, dy) = K(x, dy) + (1 - K(x, \mathcal{E}))\mu(dy)$$

$=$ kernel of a chain which behaves like (X_t) until it dies and, then is redistributed according to μ.

Results

- For each μ probability on $\mathcal{E}$

$$K_\mu(x, dy) = K(x, dy) + (1 - K(x, \mathcal{E}))\mu(dy)$$

$=$ kernel of a chain which behaves like (X_t) until it dies and, then is redistributed according to μ.

- $\Pi(\mu)$ the invariant probability measure of K_μ.

Results

- For each μ probability on $\mathcal{E}$

$$K_\mu(x,dy) = K(x,dy) + (1 - K(x,\mathcal{E}))\mu(dy)$$

$=$ kernel of a chain which behaves like (X_t) until it dies and, then is redistributed according to μ.

- $\Pi(\mu)$ the invariant probability measure of K_μ.

$$\Pi(\mu) = \frac{\mu G}{\mu G \mathbf{1}}, G = \sum_{n \geq 0} K^n.$$

Results

- For each μ probability on $\mathcal{E}$

$$K_\mu(x, dy) = K(x, dy) + (1 - K(x, \mathcal{E}))\mu(dy)$$

$=$ kernel of a chain which behaves like (X_t) until it dies and, then is redistributed according to μ.

- $\Pi(\mu)$ the invariant probability measure of K_μ.

$$\Pi(\mu) = \frac{\mu G}{\mu G \mathbf{1}}, G = \sum_{n \geq 0} K^n.$$

- $\{\Phi_t\}_{t \in \mathbb{R}^+}$ the deterministic semiflow induced by the ODE

$$\dot{\mu} = -\mu + \Pi(\mu)$$

(in a weak sense)

Theorem

Under hypotheses H1 (Feller) and H2 (∂ accessible), the limit set of (μ_n) is almost surely a Attractor Free set of Φ

Theorem

Under hypotheses H1 (Feller) and H2 (∂ accessible), the limit set of (μ_n) is almost surely a Attractor Free set of Φ

Set L is *Attractor free* means :

Theorem

Under hypotheses H1 (Feller) and H2 (∂ accessible), the limit set of (μ_n) is almost surely a Attractor Free set of Φ

Set L is *Attractor free* means :

(i) L is compact,

Theorem

Under hypotheses H1 (Feller) and H2 (∂ accessible), the limit set of (μ_n) is almost surely a Attractor Free set of Φ

Set L is *Attractor free* means :

(i) L is compact,

(ii) connected,

Theorem

Under hypotheses H1 (Feller) and H2 (∂ accessible), the limit set of (μ_n) is almost surely a Attractor Free set of Φ

Set L is *Attractor free* means :

(i) L is compact,

(ii) connected,

(iii) invariant : $\Phi_t(L) = L$ and

Theorem

Under hypotheses H1 (Feller) and H2 (∂ accessible), the limit set of (μ_n) is almost surely a Attractor Free set of Φ

Set L is *Attractor free* means :

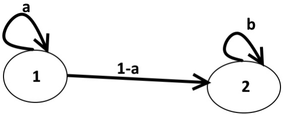
(i) L is compact,

(ii) connected,

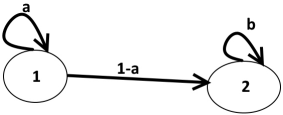
(iii) invariant : $\Phi_t(L) = L$ and

(iv) $\Phi|_L$ has no proper attractor

Example



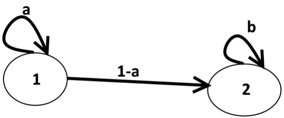
Example



Set $\mu = x\delta_1 + (1-x)\delta_2$. The ODE writes

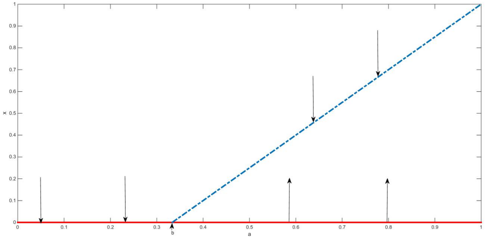
$$\dot{x} = -x + \frac{(1-b)x}{(1-a) + (1-b)x}.$$

Example



Set $\mu = x\delta_1 + (1-x)\delta_2$. The ODE writes

$$\dot{x} = -x + \frac{(1-b)x}{(1-a) + (1-b)x}.$$



Theorem

Under the additional assumptions H3 and H4 (The Champagnat Villemonais condition); Φ *has a global attractor given as the unique QSD* $\{\mu\}$*. Hence, there is only one attractor free set* $\{\mu\}$ *and*

$$\mu_n \to \mu.$$

Strategy of proof

Strategy of proof

(i) Show that $t \mapsto \hat{\mu}_t := \mu_{e^t}$ is a Asymptotic Pseudo Trajectory of Φ

Strategy of proof

(i) Show that $t \mapsto \hat{\mu}_t := \mu_{e^t}$ is a Asymptotic Pseudo Trajectory of Φ i.e

$$\lim_{t \to \infty} \sup_{0 \le s \le T} dist(\hat{\mu}_{t+s}, \Phi_s(\hat{\mu}_t)) = 0.$$

Strategy of proof

(i) Show that $t \mapsto \hat{\mu}_t := \mu_{e^t}$ is a Asymptotic Pseudo Trajectory of Φ i.e

$$\lim_{t \to \infty} \sup_{0 \leq s \leq T} dist(\hat{\mu}_{t+s}, \Phi_s(\hat{\mu}_t)) = 0.$$

(ii) Use old results from the late 90s on the dynamics of APT (B, B & Hirsch) :

Strategy of proof

(i) Show that $t \mapsto \hat{\mu}_t := \mu_{e^t}$ is a Asymptotic Pseudo Trajectory of Φ i.e

$$\lim_{t \to \infty} \sup_{0 \leq s \leq T} dist(\hat{\mu}_{t+s}, \Phi_s(\hat{\mu}_t)) = 0.$$

(ii) Use old results from the late 90s on the dynamics of APT (B, B & Hirsch) :

- The limit set of an APT is attractor free

Strategy of proof

(i) Show that $t \mapsto \hat{\mu}_t := \mu_{e^t}$ is a Asymptotic Pseudo Trajectory of Φ i.e

$$\lim_{t\to\infty} \sup_{0\leq s\leq T} dist(\hat{\mu}_{t+s}, \Phi_s(\hat{\mu}_t)) = 0.$$

(ii) Use old results from the late 90s on the dynamics of APT (B, B & Hirsch) :

- The limit set of an APT is attractor free

$\Rightarrow$ If L is attractor free and meets the basin of attraction of attractor A; Then $L \subset A$.

Strategy of proof

(i) Show that $t \mapsto \hat{\mu}_t := \mu_{e^t}$ is a Asymptotic Pseudo Trajectory of Φ i.e

$$\lim_{t\to\infty} \sup_{0\leq s\leq T} dist(\hat{\mu}_{t+s}, \Phi_s(\hat{\mu}_t)) = 0.$$

(ii) Use old results from the late 90s on the dynamics of APT (B, B & Hirsch) :
• The limit set of an APT is attractor free
$\Rightarrow$ If L is attractor free and meets the basin of attraction of attractor A; Then $L \subset A$.
$\Rightarrow$ If Φ has a global attractor A; Then $L \subset A$.

Strategy of proof

(i) Show that $t \mapsto \hat{\mu}_t := \mu_{e^t}$ is a Asymptotic Pseudo Trajectory of Φ i.e

$$\lim_{t\to\infty} \sup_{0\leq s\leq T} dist(\hat{\mu}_{t+s}, \Phi_s(\hat{\mu}_t)) = 0.$$

(ii) Use old results from the late 90s on the dynamics of APT (B, B & Hirsch) :

- The limit set of an APT is attractor free

$\Rightarrow$ If L is attractor free and meets the basin of attraction of attractor A; Then $L \subset A$.

$\Rightarrow$ If Φ has a global attractor A; Then $L \subset A$.

(iii) Show that under $H4$, Φ has the unique QSD of K as global attractor

Recent results

Strategy recently applied by

Strategy recently applied by
• **Wang, Roberts and Steinsaltz** (arXiv August 23, 2018) to diffusions on a compact manifold with soft killing:

Strategy recently applied by
• **Wang, Roberts and Steinsaltz** (arXiv August 23, 2018) to diffusions on a compact manifold with soft killing:

$$dX_t = \nabla V(X_t)dt + dW_t$$

$$\tau_\partial = \inf\{t \geq 0 : \int_0^t \kappa(X_s)ds > \xi\}$$

where ξ is an independent random variable with exponential distribution; $\kappa > 0$ and smooth

Strategy recently applied by
• **Wang, Roberts and Steinsaltz** (arXiv August 23, 2018) to diffusions on a compact manifold with soft killing:

$$dX_t = \nabla V(X_t)dt + dW_t$$

$$\tau_\partial = \inf\{t \geq 0 : \int_0^t \kappa(X_s)ds > \xi\}$$

where ξ is an independent random variable with exponential distribution; $\kappa > 0$ and smooth

• **Mailler and Villemonais** (arXiv September 6, 2018) to very general measure valued Polya processes;
↪ allows to treat the non compact situation under a Lyapunov condition

Strategy recently applied by

• **Wang, Roberts and Steinsaltz** (arXiv August 23, 2018) to diffusions on a compact manifold with soft killing:

$$dX_t = \nabla V(X_t)dt + dW_t$$

$$\tau_\partial = \inf\{t \geq 0 : \int_0^t \kappa(X_s)ds > \xi\}$$

where ξ is an independent random variable with exponential distribution; $\kappa > 0$ and smooth

• **Mailler and Villemonais** (arXiv September 6, 2018) to very general measure valued Polya processes;

$\hookrightarrow$ allows to treat the non compact situation under a Lyapunov condition