Coupling and Control

Saul Jacka (and others), University of Warwick
Warwick

9^{th} December 2019

Warwick Statistics

Control and optimal (co-adapted) coupling are intimately connected. In the talk I'll give 3 examples (if time permits).

Three problems

- *Coupling regime switching diffusions*
- *Optimal coupling for the random walk on the hypercube*
- *Strong Feller property for (controlled) jump diffusions*

Coupling regime switching diffusions

Given two regime-switching martingale dynamics:

$$dX_t = \sigma_1(Z_t)dB_t \text{ and } dY_t(V) = \sigma_2(Z_t)dV_t.$$

Here: Z and B are $\mathcal{F}_t$-adapted; Z is an $\mathcal{F}_t$-Markov process and B is an $\mathcal{F}_t$-Brownian motion. Seek to couple Y, by choosing a suitable $\mathcal{F}_t$-BM, V.

Coupling regime switching diffusions

Given two regime-switching martingale dynamics:

$$dX_t = \sigma_1(Z_t)dB_t \text{ and } dY_t(V) = \sigma_2(Z_t)dV_t.$$

Here: Z and B are $\mathcal{F}_t$-adapted; Z is an $\mathcal{F}_t$-Markov process and B is an $\mathcal{F}_t$-Brownian motion. Seek to couple Y, by choosing a suitable $\mathcal{F}_t$-BM, V.

Conjecture: minimal coupling time attained by mirror-coupling: i.e. take $dV_t = -sgn(\sigma_1(Z_t))sgn(\sigma_2(Z_t))dB_t$.

"**Proof**": need only consider V of form

$$V_t^c = V_0 + \int_0^t c_t dB_t + \int_0^t \sqrt{1 - c_t^2} dW_t$$

where W is BM indep^t of B and c is $\mathcal{F}_t$-adapted process in $[-1, 1]$.

Look at the distance process, $R^c_t \stackrel{def}{=} X_t - Y^{V^c}_t$:

$$R^c_t = r + \int_0^t (\sigma_1(Z_t) - c_t\sigma_2(Z_t))dB_t - \int_0^t \sqrt{1-c_t^2}\sigma_2(Z_t)dW_t.$$

Look at R's quadratic variation

$$A^c_t = \int_0^t ((\sigma_1(Z_t)^2 - 2c_t\sigma_1(Z_t)\sigma_2(Z_t) + \sigma_2^2(Z_t))dt$$

"Clearly", the bigger the quadratic variation the faster R will hit 0 and the processes will couple. Largest qv^n is attained by choice of c above.

Look at the distance process, $R_t^c \stackrel{def}{=} X_t - Y_t^{V^c}$:

$$R_t^c = r + \int_0^t (\sigma_1(Z_t) - c_t\sigma_2(Z_t))dB_t - \int_0^t \sqrt{1 - c_t^2}\sigma_2(Z_t)dW_t.$$

Look at R's quadratic variation

$$A_t^c = \int_0^t ((\sigma_1(Z_t)^2 - 2c_t\sigma_1(Z_t)\sigma_2(Z_t) + \sigma_2^2(Z_t))dt$$

"Clearly", the bigger the quadratic variation the faster R will hit 0 and the processes will couple. Largest qv^n is attained by choice of c above.

Conjecture is, in general, false even if we assume that Z and B are independent!

However,

Theorem(*a non-coupling result*) if Z is an $\mathcal{F}_t$-Markov chain then Z is independent of any $\mathcal{F}_t$-BM (and the conjecture is true since the "proof" then works via stoch. control argument).

However,

Theorem(*a non-coupling result*) if Z is an $\mathcal{F}_t$-Markov chain then Z is independent of any $\mathcal{F}_t$-BM (and the conjecture is true since the "proof" then works via stoch. control argument).

Proof Suppose that Z is an $\mathcal{F}_t$ Markov chain and W is an $\mathcal{F}_t$-BM with transition semigroup P^W. If we can show that

$$\mathbb{E}[f(Z_T)g(W_T)] = \mathbb{E}[f(Z_T)]\mathbb{E}[g(W_T)], \qquad (2)$$

for arbitrary bounded measurable, real-valued functions f and g, then Z_T and W_T are independent. Then extend argument to get independence of finite dimensional distributions.

To prove (2), define

$$M_t = \mathbb{E}[f(Z_T)|\mathcal{F}_t] \text{ and } N_t = \mathbb{E}[g(W_T)|\mathcal{F}_t].$$

If Z has transition semigroup P^Z and $Q-$matrix Q and W has the Brownian transition semigroup P^W then $M_t = P^Z_{T-t}f(Z_t)$ and $N_t = P^W_{T-t}g(W_t)$ so that

$$dM_t = \Delta P^Z_{T-.}f(Z_.)_t - QP^Z_{T-t}f(Z_t)dt \text{ and } dN_t = s_t dW_t.$$

If Z has transition semigroup P^Z and $Q-$matrix Q and W has the Brownian transition semigroup P^W then $M_t = P^Z_{T-t}f(Z_t)$ and $N_t = P^W_{T-t}g(W_t)$ so that

$$dM_t = \Delta P^Z_{T-\cdot}f(Z_\cdot)_t - QP^Z_{T-t}f(Z_t)dt \text{ and } dN_t = s_t dW_t.$$

So M is a purely discontinuous martingale and N is a continuous martingale so their co-variation $[M, N]$ is identically zero. Stochastic calculus now tells us that MN is a martingale. Hence

$$\mathbb{E}[f(Z_T)g(W_T)] = \mathbb{E}[M_T N_T] = M_0 N_0 = \mathbb{E}[f(Z_T)]\mathbb{E}[g(W_T)]$$

as required □

Random walk on the hypercube

Label vertices of an n-dimensional hypercube by $\{0,1\}^n$.

A continuous-time random walk X on $\{0,1\}^n$ may be defined using a Poisson process Λ of rate n, at jumps of Λ choose uniformly a coordinate i to change.

We want to couple two such random walks, X and Y, starting from different states. Denote matched coordinates by M_t and unmatched coordinates by U_t.

A simple coupling technique given by Aldous:

- if $X(i)$ flips at time t, with $i \in M_t$, then also flip coordinate $Y(i)$ at time t (matched coordinates move synchronously);

A simple coupling technique given by Aldous:

- if $X(i)$ flips at time t, with $i \in M_t$, then also flip coordinate $Y(i)$ at time t (matched coordinates move synchronously);
- if $|U_t| > 1$ and $X(i)$ flips at time t, with $i \in U_t$, also flip coordinate $Y(j)$ at time t, where j is chosen uniformly from $U_t \setminus \{i\}$—this reduces $|U_t|$ by 2 ;

A simple coupling technique given by Aldous:

- if $X(i)$ flips at time t, with $i \in M_t$, then also flip coordinate $Y(i)$ at time t (matched coordinates move synchronously);
- if $|U_t| > 1$ and $X(i)$ flips at time t, with $i \in U_t$, also flip coordinate $Y(j)$ at time t, where j is chosen uniformly from $U_t \setminus \{i\}$—this reduces $|U_t|$ by 2 ;
- else, if $U_t = \{i\}$ contains only one element, allow coordinates $X(i)$ and $Y(i)$ to evolve independently until final match is made.

We search for best possible co-adapted coupling of X and Y.

Note that any co-adapted coupling (X^c, Y^c) must satisfy following three constraints:

- At any instant the number of jumps by the co-ordinates process (X^c, Y^c) cannot exceed two (one coordinate of X^c and one of Y^c);

We search for best possible co-adapted coupling of X and Y.

Note that any co-adapted coupling (X^c, Y^c) must satisfy following three constraints:

- At any instant the number of jumps by the co-ordinates process (X^c, Y^c) cannot exceed two (one coordinate of X^c and one of Y^c);
- All single and double jumps must have rates bounded above by one;

We search for best possible co-adapted coupling of X and Y.

Note that any co-adapted coupling (X^c, Y^c) must satisfy following three constraints:

- At any instant the number of jumps by the co-ordinates process (X^c, Y^c) cannot exceed two (one coordinate of X^c and one of Y^c);
- All single and double jumps must have rates bounded above by one;
- For all $i = 1, \dots, n$, the *total* rate at which $X^c(i)$ jumps must equal one.

So, a general co-adapted coupling for X and Y can be defined by (adapted, random) rates of coordinate pairs (i, j) flipping, with index 0 corresponding to no jump for relevant process, and sum of rates being 1 for each co-ordinate.

Proposed optimal coupling strategy, $\hat{c}$, is as follows:

- Use Aldous coupling if N_t, the size of U_t, has an even number of elements
- if N_t is odd, all unmatched coordinates of X and Y are made to evolve independently until N becomes even.

The only difference from Aldous's coupling is $\hat{c}$ seeks to restore parity of N at the beginning.

Proposed optimal coupling strategy, $\hat{c}$, is as follows:

- Use Aldous coupling if N_t, the size of U_t, has an even number of elements
- if N_t is odd, all unmatched coordinates of X and Y are made to evolve independently until N becomes even.

The only difference from Aldous's coupling is $\hat{c}$ seeks to restore parity of N at the beginning. It follows that $\tau^{\hat{c}}$, the coupling time under $\hat{c}$, is the sum of independent exponential r.v.s:

$$\tau^{\hat{c}} = E_{2m} + E_{2(m-1)} + \ldots + E_2 \text{ if } N_0 = 2m$$

and

$$\tau^{\hat{c}} = E_{2(2m+1)} + E_{2m} + E_{2(m-1)} + \ldots + E_2 \text{ if } N_0 = 2m+1,$$

where E_k has rate k.

Now define

$$\hat{v}(x, y, t) = \mathbb{P}\left[\hat{\tau} > t \,|\, X_0 = x,\ Y_0 = y\right] \tag{1}$$

to be the tail of the coupling time distribution under $\hat{c}$.

Theorem

For any states $x, y \in \{0, 1\}^n$ and time $t \geq 0$,

$$\hat{v}(x, y, t) = \inf_{c \in \mathcal{C}} \mathbb{P}\left[\tau^c > t \,|\, X_0 = x, Y_0 = y\right]. \tag{2}$$

In other words, $\hat{\tau}$ is the stochastic minimum of all co-adapted coupling times for the pair (X, Y).

How do we prove this? For any coupling/control c, we construct the Bellman process $S^c_t = \hat{v}(X_t, Y^c_t, T-t)$ and show that S^c is a (bounded) submartingale. It follows that

$$\mathbb{P}(\tau^{\hat{c}} > T) = \hat{v}(x, y, T) \leq S^c_0 \leq \mathbb{E}[S^c_{\tau^c \wedge T}] = \mathbb{P}(\tau^c > T).$$

How do we prove this? For any coupling/control c, we construct the Bellman process $S^c_t = \hat{v}(X_t, Y^c_t, T-t)$ and show that S^c is a (bounded) submartingale. It follows that

$$\mathbb{P}(\tau^{\hat{c}} > T) = \hat{v}(x, y, T) \leq S^c_0 \leq \mathbb{E}[S^c_{\tau^c \wedge T}] = \mathbb{P}(\tau^c > T).$$

How? We consider the dynamics of S^c:

$$dS^c_t = dZ^c_t + \left(Q^{c_t} \hat{v} - \frac{\partial \hat{v}}{\partial t} \right) dt \,,$$

where Z^c is a martingale and Q^c is the Q-matrix corresponding to the switching rates.

How do we prove this? For any coupling/control c, we construct the Bellman process $S_t^c = \hat{v}(X_t, Y_t^c, T - t)$ and show that S^c is a (bounded) submartingale. It follows that

$$\mathbb{P}(\tau^{\hat{c}} > T) = \hat{v}(x, y, T) \le S_0^c \le \mathbb{E}[S^c_{\tau^c \wedge T}] = \mathbb{P}(\tau^c > T).$$

How? We consider the dynamics of S^c:

$$dS_t^c = dZ_t^c + \left(Q^{c_t} \hat{v} - \frac{\partial \hat{v}}{\partial t} \right) dt \,,$$

where Z^c is a martingale and Q^c is the Q-matrix corresponding to the switching rates.

So to show the result, since we know that $S^{\hat{c}}$ is a martingale we need only show that $d(c, \hat{c}, t) \stackrel{def}{=} Q^c \hat{v} - Q^{\hat{c}} \hat{v} \ge 0$ for any c. Do this by showing that Laplace transform of d is totally monotone.

Strong Feller property

The problem is to construct a Markov process X with the following attributes and then establish the strong Feller property:

- In between jumps, X should behave like the Ito diffusion Y, which satisfies

$$dY_t = \sigma(Y_t)dB_t + \mu(Y_t)dt. \tag{1}$$

Strong Feller property

The problem is to construct a Markov process X with the following attributes and then establish the strong Feller property:

- In between jumps, X should behave like the Ito diffusion Y, which satisfies

$$dY_t = \sigma(Y_t)dB_t + \mu(Y_t)dt. \tag{1}$$

- When at the point $x \in \mathbb{R}^d$, X should have a jump intensity measure $\nu(x, \cdot)$ so that, conditional on being at x, X has jumps distributed as a compound Poisson process.

Assume that

- σ and μ are such that there is a unique (possibly explosive) weak solution (i.e. in law) to (1).
- σ is locally uniformly elliptic and μ is locally bounded.
- ν is, locally, a finite kernel i.e for each Borel $A \subset \mathbb{R}^d$, $\nu(\cdot, A)$ is measurable and there is a sequence of compact $K_n \uparrow \mathbb{R}^d$ such that $\sup_{x \in K_n} \nu(x, \mathbb{R}^d) = r_n < \infty$.

The construction is then fairly standard (although I've never seen it) using Poisson thinning. Since we just kill the process on explosion, we only need to construct up to the first exit from K_n for any n, so wlog ν is bounded in total mass which is then 1 wlog.

Construct a unit rate PP, N.

- Construct Y solving (1), starting at $X_0^x = x$
- Then set $X = Y$ prior to the first jump time J_1 of N
- Generate a jump proposal, Z, with conditional distribution $\frac{\nu(Y_{J_1},\cdot)}{\nu(Y_{J_1},\mathbb{R}^d)}$.
- Accept the jump with probability $\nu(Y_{J_1},\mathbb{R}^d)$, setting $X_{J_1} = X_{J_1-} + Z$, and with probability $1-\nu(Y_{J_1},\mathbb{R}^d)$ reject the jump and leave X where it was.
- Now repeat starting at $X_{J_1},\ldots$.

Now we want to show that the semigroup for X, $(P_t)_{t\geq 0}$ has the property that P_t maps bounded, measurable functions to continuous ones (for $t > 0$).

Now we want to show that the semigroup for X, $(P_t)_{t\geq 0}$ has the property that P_t maps bounded, measurable functions to continuous ones (for $t > 0$).

Coupling really helps here: if we can show that there is a coupling of X^x and $X^{x'}$ with coupling time τ and $\mathbb{P}(\tau > t) \to 0$ as $x' \to x$ then, if $|f| \leq b$,

$$|P_t f(x) - P_t f(x')| \leq \mathbb{E}[|f(X_t^x - f(X_t^{x'})|] \leq 2b\mathbb{P}(\tau > t) \to 0 \text{ as } x' \to x.$$

Trick is to observe that jumps make no big difference to argument: given $t, \delta > 0$ take E s.t. $E < t$ and $P(J_1 < E \leq \delta/2)$ then

$$\mathbb{P}(\tau(x,x') > t) \leq \mathbb{P}(\tau(x,x') > E) \leq \frac{\delta}{2} + \mathbb{P}(\tilde{\tau}(x,x') > E),$$

where $\tilde{\tau}$ is the coupling time for diffusions Y^x and $Y^{x'}$.

Trick is to observe that jumps make no big difference to argument: given $t, \delta > 0$ take E s.t. $E < t$ and $P(J_1 < E \leq \delta/2)$ then

$$\mathbb{P}(\tau(x,x') > t) \leq \mathbb{P}(\tau(x,x') > E) \leq \frac{\delta}{2} + \mathbb{P}(\tilde{\tau}(x,x') > E),$$

where $\tilde{\tau}$ is the coupling time for diffusions Y^x and $Y^{x'}$.

To couple these we need a further assumption: that σ and μ are locally Lipschitz

- We can use reflection coupling to do the job- so Y driven by "B orthogonal to line joining Y^x and $Y^{x'}$ and $-B$ on line joining Y^x and $Y^{x'}$".

Trick is to observe that jumps make no big difference to argument: given $t, \delta > 0$ take E s.t. $E < t$ and $P(J_1 < E \leq \delta/2)$ then

$$\mathbb{P}(\tau(x,x') > t) \leq \mathbb{P}(\tau(x,x') > E) \leq \frac{\delta}{2} + \mathbb{P}(\tilde{\tau}(x,x') > E),$$

where $\tilde{\tau}$ is the coupling time for diffusions Y^x and $Y^{x'}$.

To couple these we need a further assumption: that σ and μ are locally Lipschitz

- We can use reflection coupling to do the job- so Y driven by "B orthogonal to line joining Y^x and $Y^{x'}$ and $-B$ on line joining Y^x and $Y^{x'}$".
- End up (after time-change) comparing distance between Y^x and $Y^{x'}$ to a Bessel process of dimension $1+\epsilon$. Since this hits zero, processes couple with high prob. (in small time) if start close enough together and do so before exiting a ball (so local assumption doesn't matter).