Some directions in Perfect Simulation

Laird Breyer,
mailto:laird@lbreyer.com

joint work with Gareth Roberts (Lancaster)
The work presented here has involved the following institutions:


University of Lancaster


Aalborg University


Università di Roma Tre

Aims of this talk

In this talk, I shall present loosely some recent ideas related to Perfect Simulation, including:

Aims of this talk

In this talk, I shall present loosely some recent ideas related to Perfect Simulation, including:

- what Read-Once CFTP means

Aims of this talk

In this talk, I shall present loosely some recent ideas related to Perfect Simulation, including:

- what Read-Once CFTP means
- some old and new coupling constructions

Aims of this talk

In this talk, I shall present loosely some recent ideas related to Perfect Simulation, including:

- what Read-Once CFTP means
- some old and new coupling constructions
- an example of perfect simulation

Aims of this talk

In this talk, I shall present loosely some recent ideas related to Perfect Simulation, including:

- what Read-Once CFTP means
- some old and new coupling constructions
- an example of perfect simulation

PS is still at the beginning, and new methods are discovered regularly...

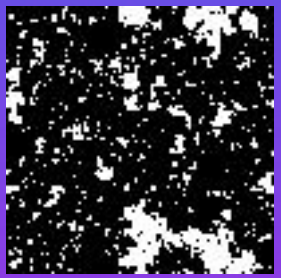
MCMC and Perfect Samples

In MCMC, we commonly have two distinct goals:

MCMC and Perfect Samples

In MCMC, we commonly have two distinct goals:

Simulation



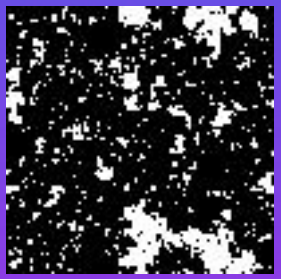
Integration

$$\mathbb{P}(A \mid B) = \int_A \mathbb{P}(B \mid \omega)\mathbb{P}(d\omega)/\mathbb{P}(B)$$

MCMC and Perfect Samples

In MCMC, we commonly have two distinct goals:

Simulation



Integration

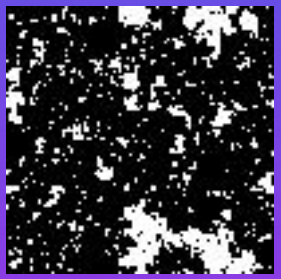
$$\mathbb{P}(A \mid B) = \int_A \mathbb{P}(B \mid \omega)\mathbb{P}(d\omega)/\mathbb{P}(B)$$

PS solves the simulation problem completely,

MCMC and Perfect Samples

In MCMC, we commonly have two distinct goals:

Simulation



Integration

$$\mathbb{P}(A \mid B) = \int_A \mathbb{P}(B \mid \omega)\mathbb{P}(d\omega)/\mathbb{P}(B)$$

PS solves the simulation problem completely, and reduces integration problems to Classical Statistics.

Building Markov chains

There are several equivalent definitions for Markov chains. For PS, we use a computational definition:

A sequence X_1, X_2,... is a Markov chain if there exists an IID series of random functions $F_1(x)$, $F_2(x)$,... such that

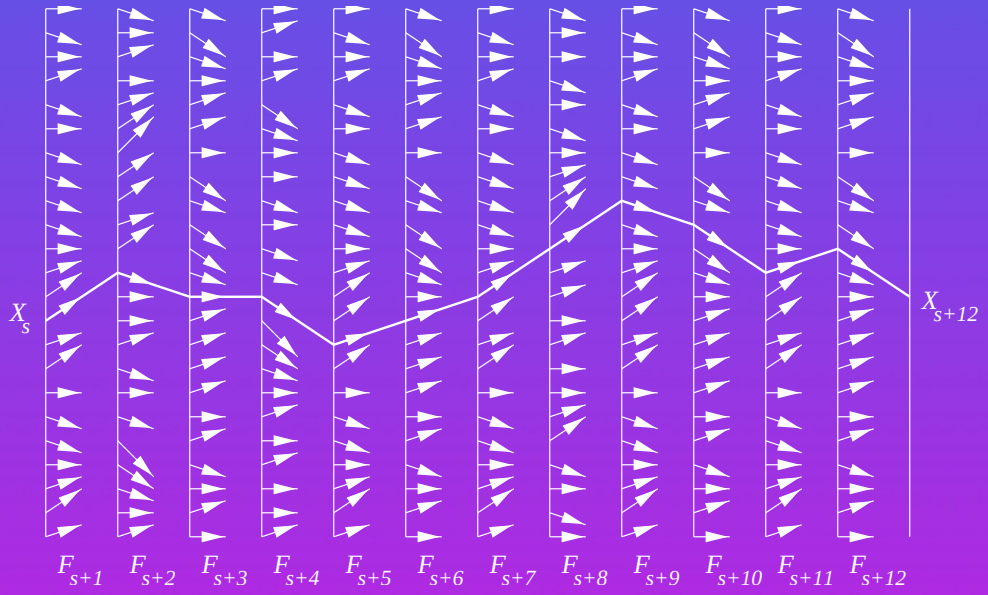
$$F_1(X_1) = X_2, \quad F_2(X_2) = X_3, \quad \ldots, \quad F_k(X_k) = X_{k+1}$$

Building Markov chains

There are several equivalent definitions for Markov chains. For PS, we use a computational definition:

A sequence X_1, X_2,... is a Markov chain if there exists an IID series of random functions $F_1(x)$, $F_2(x)$,... such that

$$F_1(X_1) = X_2, \quad F_2(X_2) = X_3, \quad \ldots, \quad F_k(X_k) = X_{k+1}$$



The stationary density

Consider the chain over a larger time scale: the stationary density π satisfies $\pi P = \pi$, where

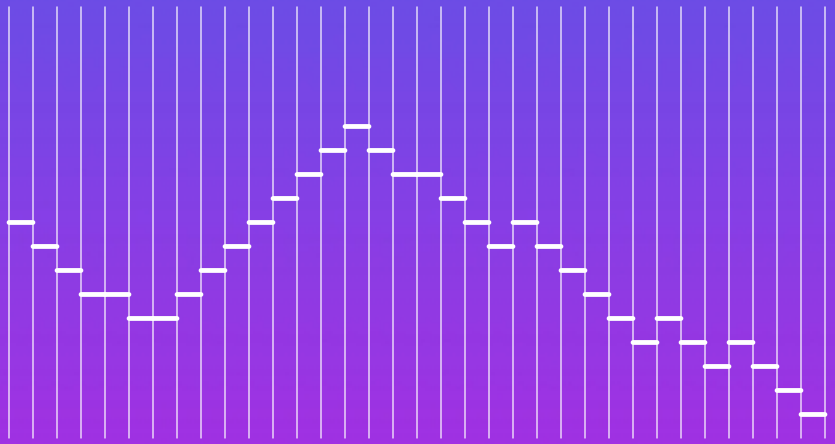
$$P(x, dy) = \mathbb{P}(F_t(x) \in dy) \quad \text{for all } t.$$

The stationary density

Consider the chain over a larger time scale: the stationary density π satisfies $\pi P = \pi$, where

$$P(x, dy) = \mathbb{P}(F_t(x) \in dy) \quad \text{for all } t.$$

Here is a path:

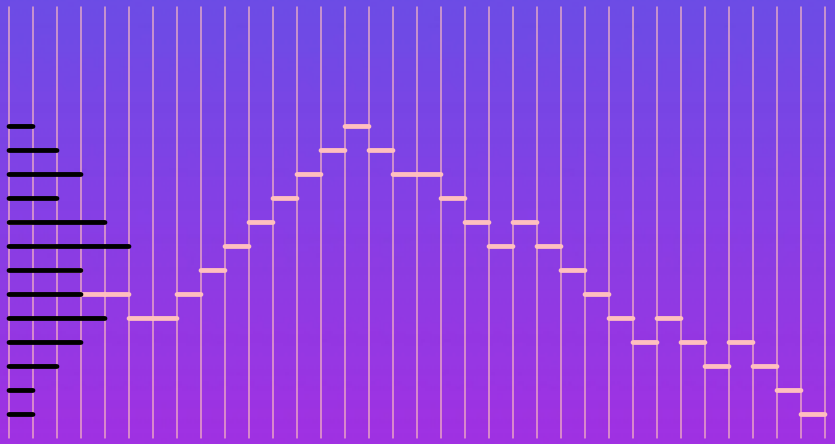


The stationary density

Consider the chain over a larger time scale: the stationary density π satisfies $\pi P = \pi$, where

$$P(x, dy) = \mathbb{P}(F_t(x) \in dy) \quad \text{for all } t.$$

After reordering the occupied states, we get:

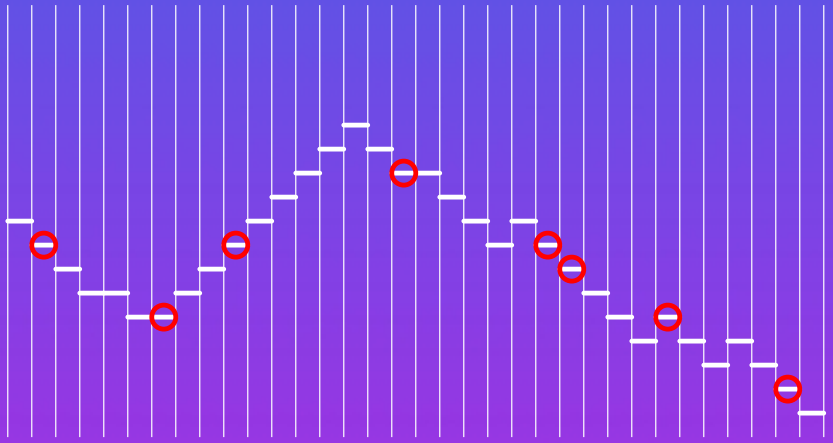


Read Once Perfect Simulation

Wilson (1999) showed how to recognize random times T_2, T_3,... when the chain is *exactly* in equilibrium:

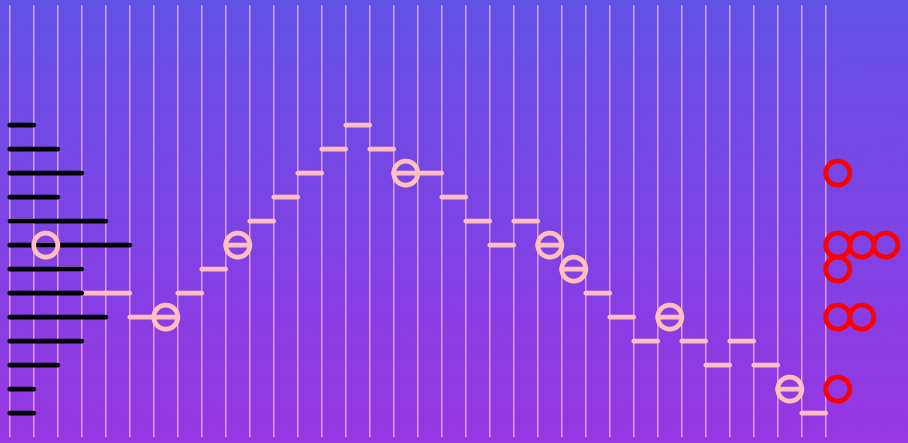
Read Once Perfect Simulation

Wilson (1999) showed how to recognize random times T_2, T_3,... when the chain is *exactly* in equilibrium:



Read Once Perfect Simulation

Wilson (1999) showed how to recognize random times T_2, T_3,... when the chain is *exactly* in equilibrium:



How to find the perfect samples

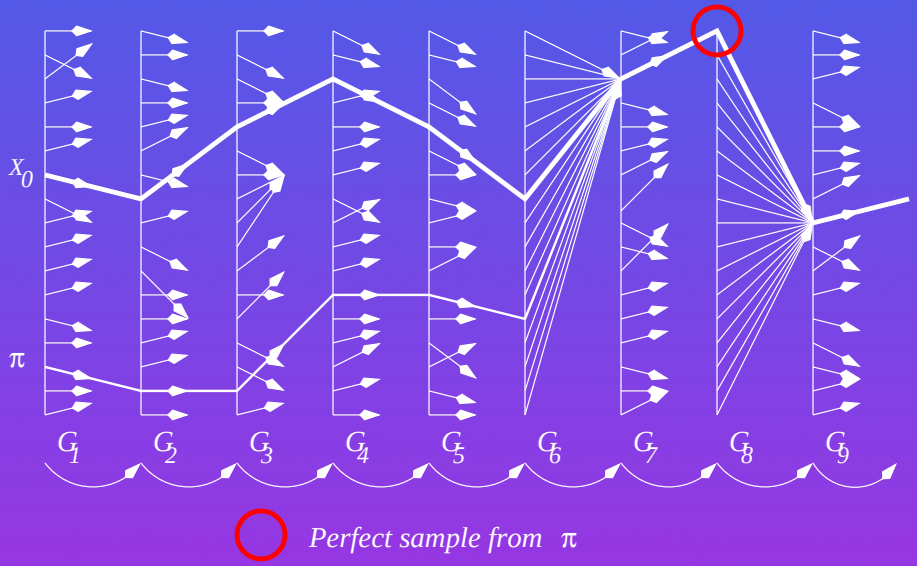
- Trace all possible paths under the mappings F_1, F_2,..., looking for *coalescence* within m steps:
- Let $G_1(x)$, $G_2(x)$,... be defined for all x by:

$$G_1(x) = F_m(F_{m-1}(\cdots F_1(x)\cdots))$$
$$G_2(x) = F_{2m}(F_{2m-1}(\cdots F_{m+1}(x)\cdots))$$

and set

$$T_1 = \min\{u \geq 0 : G_u(x) = G_u(y) \text{ for all } x, y\}$$
$$T_{k+1} = \min\{u \geq T_k : G_u(x) = G_u(y) \text{ for all } x, y\}$$

How to find the perfect samples



The fundamental identity

Let $\epsilon = \mathbb{P}(T_1 \leq 1)$ and write

$$\begin{aligned}\mathbb{P}(G_1(x) \in dy) &= (1-\epsilon)\mathbb{P}(G_1(x) \in dy \,|\, T_1 > 1) + \epsilon \mathbb{P}(G_1(x) \in dy \,|\, T_1 \leq 1) \\ &= (1-\epsilon) Q(x, dy) + \epsilon \mu(dy).\end{aligned}$$

Then $\pi = \epsilon \sum_{s=0}^{\infty} (1-\epsilon)^s \mu Q^s$.

The fundamental identity

Let $\epsilon = \mathbb{P}(T_1 \leq 1)$ and write

$$\begin{aligned}\mathbb{P}(G_1(x) \in dy) &= (1-\epsilon)\mathbb{P}(G_1(x) \in dy \,|\, T_1 > 1) + \epsilon \mathbb{P}(G_1(x) \in dy \,|\, T_1 \leq 1) \\ &= (1-\epsilon) Q(x, dy) + \epsilon \mu(dy).\end{aligned}$$

Then $\pi = \epsilon \sum_{s=0}^{\infty} (1-\epsilon)^s \mu Q^s$.

Proof. Using stationarity $\pi P = \pi$, we have

$$(1-\epsilon)^k \pi Q^k = (1-\epsilon)^{k-1} \pi (P - \epsilon\mu) Q^{k-1}$$

The fundamental identity

Let $\epsilon = \mathbb{P}(T_1 \leq 1)$ and write

$$\begin{aligned}\mathbb{P}(G_1(x) \in dy) &= (1-\epsilon)\mathbb{P}(G_1(x) \in dy \,|\, T_1 > 1) + \epsilon \mathbb{P}(G_1(x) \in dy \,|\, T_1 \leq 1) \\ &= (1-\epsilon)Q(x, dy) + \epsilon\mu(dy).\end{aligned}$$

Then $\pi = \epsilon \sum_{s=0}^{\infty} (1-\epsilon)^s \mu Q^s$.

Proof. Using stationarity $\pi P = \pi$, we have

$$\begin{aligned}(1-\epsilon)^k \pi Q^k &= (1-\epsilon)^{k-1} \pi (P - \epsilon\mu) Q^{k-1} \\ &= (1-\epsilon)^{k-1} \pi Q^{k-1} - \epsilon(1-\epsilon)^{k-1} \mu Q^{k-1}\end{aligned}$$

The fundamental identity

Let $\epsilon = \mathbb{P}(T_1 \leq 1)$ and write

$$\begin{aligned}\mathbb{P}(G_1(x) \in dy) &= (1-\epsilon)\mathbb{P}(G_1(x) \in dy \,|\, T_1 > 1) + \epsilon \mathbb{P}(G_1(x) \in dy \,|\, T_1 \leq 1) \\ &= (1-\epsilon) Q(x, dy) + \epsilon \mu(dy).\end{aligned}$$

Then $\pi = \epsilon \sum_{s=0}^{\infty} (1-\epsilon)^s \mu Q^s$.

Proof. Using stationarity $\pi P = \pi$, we have

$$\begin{aligned}(1-\epsilon)^k \pi Q^k &= (1-\epsilon)^{k-1} \pi (P - \epsilon\mu) Q^{k-1} \\ &= (1-\epsilon)^{k-1} \pi Q^{k-1} - \epsilon (1-\epsilon)^{k-1} \mu Q^{k-1} \\ &= \cdots = \\ &= \pi - \epsilon \sum_{s=1}^{k} (1-\epsilon)^{k-s} \mu Q^{k-s}.\end{aligned}$$

The fundamental identity

Let $\epsilon = \mathbb{P}(T_1 \leq 1)$ and write

$$\begin{aligned}\mathbb{P}(G_1(x) \in dy) &= (1-\epsilon)\mathbb{P}(G_1(x) \in dy \,|\, T_1 > 1) + \epsilon \mathbb{P}(G_1(x) \in dy \,|\, T_1 \leq 1) \\ &= (1-\epsilon) Q(x, dy) + \epsilon \mu(dy).\end{aligned}$$

Then $\pi = \epsilon \sum_{s=0}^{\infty} (1-\epsilon)^s \mu Q^s$.

Proof. Using stationarity $\pi P = \pi$, we have

$$\begin{aligned}(1-\epsilon)^k \pi Q^k &= (1-\epsilon)^{k-1} \pi (P - \epsilon\mu) Q^{k-1} \\ &= (1-\epsilon)^{k-1} \pi Q^{k-1} - \epsilon (1-\epsilon)^{k-1} \mu Q^{k-1} \\ &= \cdots = \\ &= \pi - \epsilon \sum_{s=1}^{k} (1-\epsilon)^{k-s} \mu Q^{k-s}.\end{aligned}$$

Let $k \to \infty$, then left side goes to zero.

How to ensure coupling occurs

Again assume that we have an IID sequence F_1, F_2,... of random maps satisfying

$$P(x, dy) = \mathbb{P}(F_t(x) \in dy) \quad \text{for all } t. \tag{1}$$

With $F_t(x)$ arbitrary, coalescence of Markov chain sample paths need not ever happen. We must modify the maps subject to preserving the condition (**??**).

Modifying the updates

- Normally at each iteration, we use a randomly generated function $F_t(x)$ such that $X_{t+1} = F_t(X_t)$.

Modifying the updates

- Normally at each iteration, we use a randomly generated function $F_t(x)$ such that $X_{t+1} = F_t(X_t)$.
- If $F_t(x) = F_t(x')$ holds for some x, x', then there is the *possibility* of coupling.

Modifying the updates

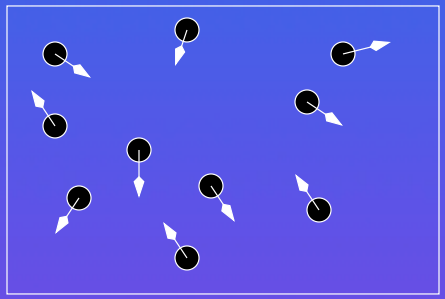
- Normally at each iteration, we use a randomly generated function $F_t(x)$ such that $X_{t+1} = F_t(X_t)$.
- If $F_t(x) = F_t(x')$ holds for some x, x', then there is the *possibility* of coupling.
- **Strategy:** From F, generate a new random function $\mathcal{C}_Y(F)$ which has a higher chance of coupling.

Modifying the updates

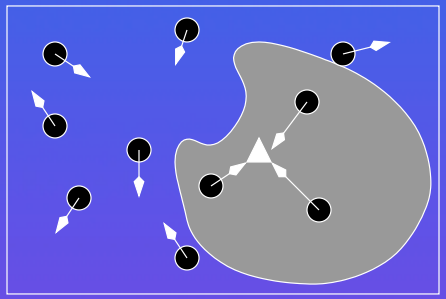
- Normally at each iteration, we use a randomly generated function $F_t(x)$ such that $X_{t+1} = F_t(X_t)$.
- If $F_t(x) = F_t(x')$ holds for some x, x', then there is the *possibility* of coupling.
- **Strategy:** From F, generate a new random function $\mathcal{C}_Y(F)$ which has a higher chance of coupling.
- **Definition:** Let Y be independent of F, with $\mathbb{P}(Y = y) = q(y)$ say.

$$\mathcal{C}_Y(F)(x) = \begin{cases} Y & \text{if } \frac{p(x,Y)q(F(x))}{p(x,F(x))q(Y)} > U[0,1] \\ F(x) & \text{otherwise.} \end{cases}$$

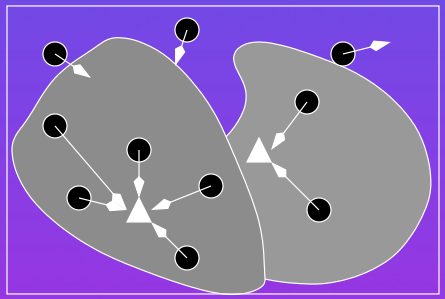
Modifying the updates



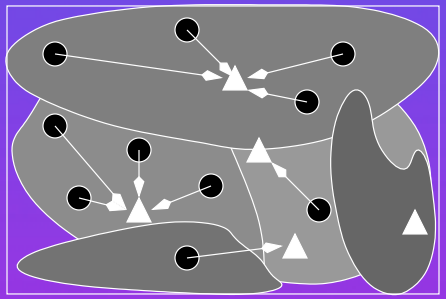
f



$G_{Y_1}(f)$

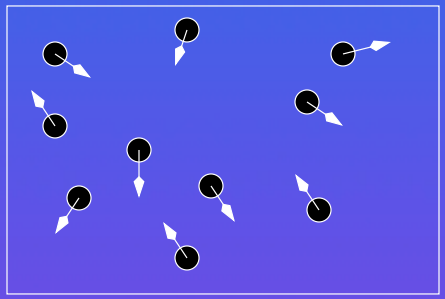


$G_{Y_2}(C_{Y_1}(f))$

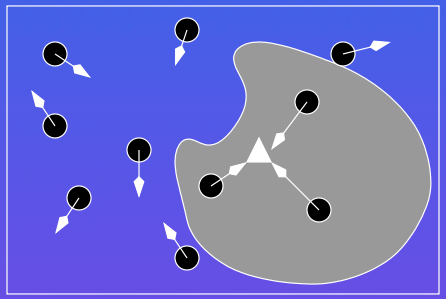


$G_{Y_5}(C_{Y_4}(G_{Y_3}(C_{Y_2}(C_{Y_1}(f)))))$

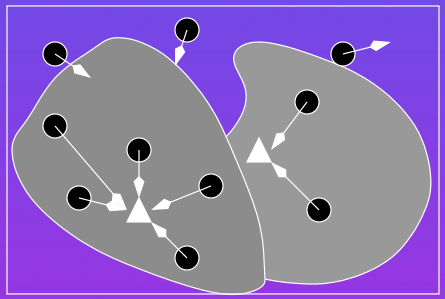
Modifying the updates



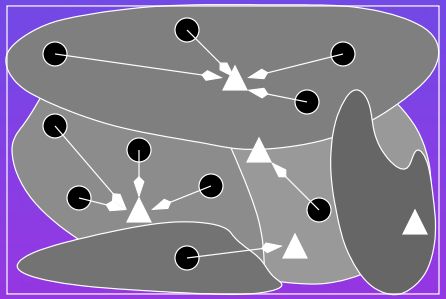
f



$G_{Y_1}(f)$



$G_{Y_2}(C_{Y_1}(f))$



$G_{Y_5}(C_{Y_4}(G_{Y_3}(C_{Y_2}(C_{Y_1}(f)))))$

We call the triangles catalysts. Click here.

Why it works

- In the formula

$$\mathcal{C}_Y(F)(x) = \begin{cases} Y & \text{if } \frac{p(x,Y)q(F(x))}{p(x,F(x))q(Y)} > U[0,1] \\ F(x) & \text{otherwise.} \end{cases}$$

fix x and define $\Pi(y) = p(x, y)$, $X = F(x)$ and $X' = \mathcal{C}_Y(F)(x)$.

Why it works

- In the formula

$$\mathcal{C}_Y(F)(x) = \begin{cases} Y & \text{if } \frac{p(x,Y)q(F(x))}{p(x,F(x))q(Y)} > U[0,1] \\ F(x) & \text{otherwise.} \end{cases}$$

  fix x and define $\Pi(y) = p(x,y)$, $X = F(x)$ and $X' = \mathcal{C}_Y(F)(x)$.

- We get

$$X' = \begin{cases} Y & \text{if } \frac{\Pi(Y)q(X)}{\Pi(X)q(Y)} > U[0,1] \\ X & \text{otherwise} \end{cases}$$

Why it works

- In the formula

$$\mathcal{C}_Y(F)(x) = \begin{cases} Y & \text{if } \frac{p(x,Y)q(F(x))}{p(x,F(x))q(Y)} > U[0,1] \\ F(x) & \text{otherwise.} \end{cases}$$

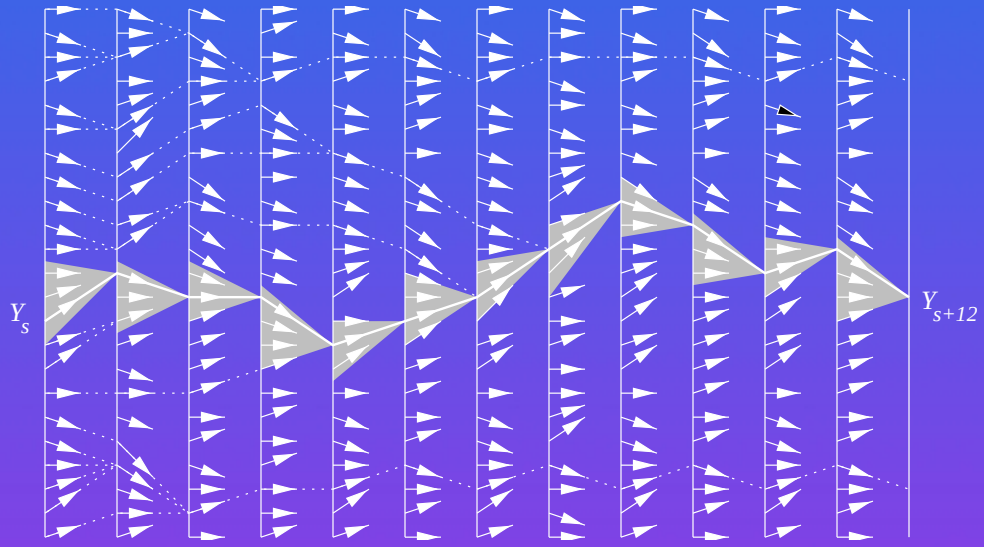
  fix x and define $\Pi(y) = p(x,y)$, $X = F(x)$ and $X' = \mathcal{C}_Y(F)(x)$.

- We get

$$X' = \begin{cases} Y & \text{if } \frac{\Pi(Y)q(X)}{\Pi(X)q(Y)} > U[0,1] \\ X & \text{otherwise} \end{cases}$$

- Therefore, since $\Pi(dy) = \mathbb{P}(X \in dy)$ by assumption, it follows that $\mathbb{P}(X' \in dy) = \Pi(dy)$ also.

Extensions



- At every update, let the catalyst Y_t come from an independent Markov chain with identical transition probabilities. Great for Gibbs samplers!
- We have the formula

$$\|\mathbb{P}(X_t \in \cdot) - \pi\|_{\text{TV}} \leq \mathbb{E} \prod_{s=0}^{t} \left(1 - \frac{p(X_s, Y_{s+1})p(Y_s, X_{s+1})}{p(X_s, X_{s+1})p(Y_s, Y_{s+1})}\right)_+,$$

Pump Example (Autogamma)

- Let $x = (\beta, \lambda_1, \ldots, \lambda_{10})$, simulate

$$\pi(x) = \exp\left\{ (10\alpha + \gamma - 1)\log\beta - \delta\beta + \sum_{k=1}^{10}\Big((s_k + \alpha - 1)\log\lambda_k - (\beta + t_k)\lambda_k \Big) \right\},$$

Pump Example (Autogamma)

- Let $x = (\beta, \lambda_1, \ldots, \lambda_{10})$, simulate

$$\pi(x) = \exp\left\{ (10\alpha + \gamma - 1) \log \beta - \delta\beta + \sum_{k=1}^{10} \Big((s_k + \alpha - 1) \log \lambda_k - (\beta + t_k)\lambda_k \Big) \right\},$$

- **Gibbs sampler:** One sweep is $f : (\beta, \lambda_1, \ldots, \lambda_{10}) \mapsto (\beta', \lambda'_1, \ldots, \lambda'_{10})$, where

$$\beta' \sim \pi_0(\cdot \mid \lambda_1, \ldots, \lambda_{10}) = \Gamma(\gamma + 10\alpha, \delta + \sum_{k=1}^{10} \lambda_k),$$

$$\lambda'_k \sim \pi_k(\cdot \mid \beta') = \Gamma(\alpha + s_k, \beta' + t_k), \quad k = 1, \ldots, 10.$$

Pump Example: coupling

- Transition density is

$$p(\beta, \lambda_1, \ldots, \lambda_{10}; b, l_1, \ldots, l_{10}) = \pi_0(b \,|\, \lambda_1, \ldots, \lambda_{10}) \prod_{k=1}^{10} \pi_k(l_k \,|\, b).$$

Pump Example: coupling

- Transition density is

$$p(\beta, \lambda_1, \ldots, \lambda_{10}; b, l_1, \ldots, l_{10}) = \pi_0(b \,|\, \lambda_1, \ldots, \lambda_{10}) \prod_{k=1}^{10} \pi_k(l_k \,|\, b).$$

- Take $Y \sim q$ where

$$q(y_0, y_1, \ldots, y_{10}) = \pi_0(y_0 \,|\, \lambda_1^*, \ldots, \lambda_{10}^*) \prod_{k=1}^{10} \pi_k(y_k \,|\, y_0),$$

Pump Example: coupling

- After simplification,

$$\boldsymbol{c}_Y(f)(x) = \begin{cases} Y & \text{if } \exp\Big((\beta' - Y_0)(|\lambda| - |\lambda^*|)\Big) > \xi, \\ (\beta', \lambda'_1, \ldots, \lambda'_{10}) & \text{otherwise.} \end{cases}$$

Pump Example: coupling

- After simplification,

$$\mathcal{C}_Y(f)(x) = \begin{cases} Y & \text{if } \exp\Big((\beta' - Y_0)(|\lambda| - |\lambda^*|)\Big) > \xi, \\ (\beta', \lambda'_1, \ldots, \lambda'_{10}) & \text{otherwise.} \end{cases}$$

- Thus

$$\text{Basin}(Y, f, \xi) = \Big\{x : \Big(\psi_0 - (\delta + |\lambda|)Y_0\Big)\Big(|\lambda| - |\lambda^*|\Big) > \log \xi\Big\},$$

Pump Example: coupling

- After simplification,

$$\text{Basin}(Y, f, \xi) = \left\{ x : |\lambda| \in \left[\frac{-b - \sqrt{b^2 - 4ac}}{2a}, \frac{-b + \sqrt{b^2 - 4ac}}{2a} \right] \right\},$$

where $a = Y_0$, $b = \log \xi - Y_0(|\lambda^*| - \delta)$ and $c = \delta(\log \xi - Y_0 |\lambda^*|) - \psi_0$.

Conclusions

- If we can recognize the basins, we can also recognize when the maps G_u coalesce.

Conclusions

- If we can recognize the basins, we can also recognize when the maps G_u coalesce.
- If we can recognize map coalescence, we can also do Perfect Simulation.

Conclusions

- If we can recognize the basins, we can also recognize when the maps G_u coalesce.
- If we can recognize map coalescence, we can also do Perfect Simulation.
- You can find a simulation by clicking here.

Learning more

Preprints and tutorials may be found at the following two sites:

- `http://www.dimacs.rutgers.edu/~dbwilson/exact.html`
- `http://www.lbreyer.com`