Perfect simulation without looking back. A case study using a Gibbs sampler.

Laird Breyer
Dept of Mathematics,
Lancaster University.

joint work with Gareth Roberts,
Lancaster University

Lancaster, 11 May 2000

mailto:l.breyer@lancaster.ac.uk

Introduction - Markov chains

- A Markov chain, written $X_1, X_2, X_3, \dots \in E$ is the stochastic analogue of a discrete time dynamical system:

$$X_{t+1} = F(X_t).$$

- If $F(x)$ is fixed (deterministic) we have a dynamical system
- If $F(x)$ is chosen randomly at each iteration, but with some fixed transition density

$$\mathbb{P}\big(F(x) = y\big) = p(x, y),$$

then we have a Markov chain.

- Often only $p(x, y)$ is specified - then there are many possible choices of F (Borovkov + Foss, 1992)

Introduction - ergodicity

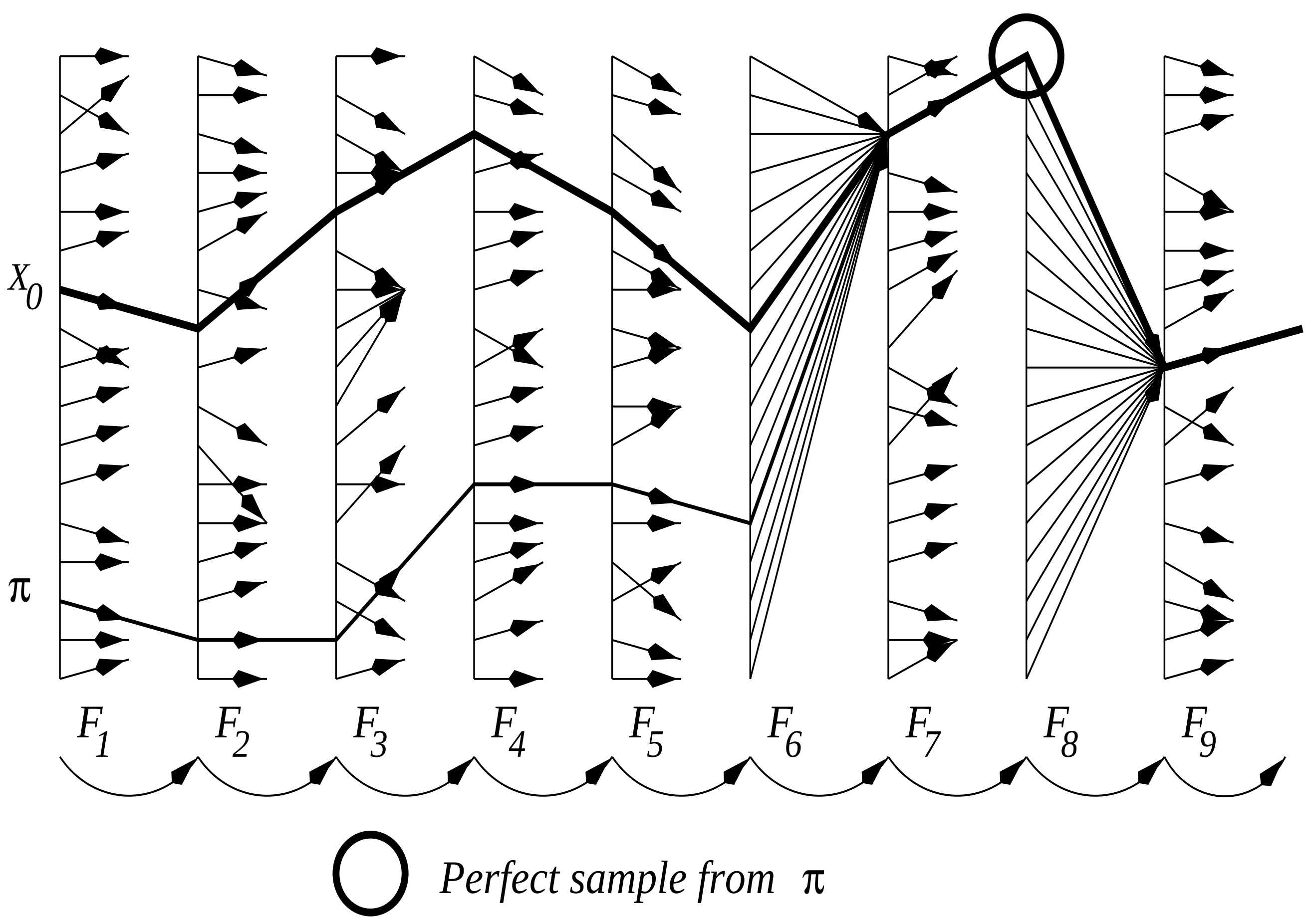
- A Markov chain which visits all parts of its state space E sufficiently frequently is called positive recurrent
- In this case, if it is not periodic, the chain settle down over time to some equilibrium distribution π on E:

$$\lim_{t \to \infty} \mathbb{P}\big(X_t \in A\big) = \pi(A)$$

$$\lim_{t \to \infty} \frac{1}{t} \sum_{s=1}^{t} f(X_s) = \int f d\pi \text{ with probability 1.}$$

- Both these results are interesting for Statistics (and Physics, and ...) as a way of doing approximate integration over a given probability distribution π in a complicated space.

Read Once CFTP



Read Once CFTP: a theorem

- $P(x, dy) = \mathbb{P}\big(F(x) \in dy\big)$.
- Assume $\mathbb{P}\big(F \text{ is coalescent}\big) = \epsilon > 0$, and write $\mu(dy) = \mathbb{P}\big(F(x_0) \in dy \,|\, F \text{ is coalescent}\big)$
- This gives $P(x, dy) = (1-\epsilon)Q(x, dy) + \epsilon\mu(dy)$, and $\pi P = \pi$.
- **Theorem**. We have

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

Proof of Theorem

- **Theorem**. We have

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

- **Proof** Using stationarity $\pi P = \pi$, we have

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

As identities between positive kernels, these are true when applied to any bounded test function $f : E \to \mathbb{R}$. Changing variables $k - s \to s$ gives

$$(1-\epsilon)^k \langle \pi Q^k, f \rangle = \langle \pi, f \rangle - \epsilon \sum_{s=0}^{k-1} (1-\epsilon)^s \langle \mu Q^s, f \rangle.$$

Since also $\left| \langle \pi Q^k, f \rangle \right| \le \|f\|$, it is now clear that we obtain (1) by letting $k \to \infty$.

Ways of coupling Markov chains I

- Very easy when state space is finite: $E = \{0, \dots, n\}$.
-

$$\begin{cases} \mathbb{P}(X_{t+1} = X_t \pm 1) = 1/2 & \text{if } 0 < X_t < n \\ \mathbb{P}(X_{t+1} = n-1) = 1 & \text{if } X_t = n \\ \mathbb{P}(X_{t+1} = 1) = 1 & \text{if } X_t = 0. \end{cases}$$

- Here π is shte uniform distribution on $\{0, \dots, n\}$.
- Take X_t, X'_t to be independent copies of this chain.
- Can show that $\mathbb{P}(T < \infty) = 1$, so we have convergence!

Ways of coupling Markov chains II

- For general state spaces E, use the *Splitting technique*. (Nummelin, Athreya + Ney, Meyn + Tweedie, Rosenthal)
- Let X_t and X'_t be two independent chains with transition density $p(x, y)$. Assume there exists a set $C \subset E$ (called a small set!) such that

$$\min_{x \in C} p(x, y) \geq \epsilon \mu(y), \qquad \int \mu(y) dy = 1.$$

- Wait until both chains are simultaneously in C at some time τ say.
- With probability ϵ, choose $Y \sim \mu$ and make both chains jump to Y, i.e. $X_{\tau+1} = X'_{\tau+1} = Y$.
- With probability $1 - \epsilon$, set

$$\begin{cases} X_{\tau+1} \sim Q(X_\tau, \cdot) & = (1-\epsilon)^{-1}(p(X_\tau, \cdot) - \epsilon\mu(\cdot)) \\ X'_{\tau+1} \sim Q(X'_\tau, \cdot) & = (1-\epsilon)^{-1}(p(X'_\tau, \cdot) - \epsilon\mu(\cdot)) \end{cases}$$

Coupling without Analysis

- 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.

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

Theorem I

- Suppose the state space E can be entirely covered by a finite union of 1-small sets $C_1, C_2, \ldots, C_N$ say, and suppose we can choose a probability density q in such a way that

$$\epsilon_i \mu_i(\cdot) \le \inf_{z \in C_i} p(z, \cdot) \le \sup_{z \in C_i} p(z, \cdot) \le \gamma_i q(\cdot), \quad i = 1, \ldots, N,$$

  where the γ_i are constants. Then if we make the proposals $Y_1, Y_2, \ldots,$ we have

$$\mathbb{P}\Big(\mathcal{C}_{Y_t} \circ \cdots \circ \mathcal{C}_{Y_1}(F) \text{ has finite range eventually}\Big) = 1.$$

- The convergence is geometrically fast!

Theorem II

- We are never sure that we have made enough proposals $Y_1, Y_2, \ldots$ to guarantee that the new map $\mathcal{C}_{Y_n} \circ \cdots \circ \mathcal{C}_{Y_1}(F)$ has finite range. Fortunately, there exists a different procedure which allows us to generate a finitely coupled map with certainty.
- (Theorem) By using the "Dead Leaves Method" we can generate in a finite number of iterations a finitely coupled map with certainty!

Theorem III

- Suppose we can couple the transition maps $F_t(x)$ into new maps $\widetilde{F}_t(x)$ which have finite range with probability 1. This only works if the transition density $p(x, y)$ is uniformly ergodic. Then any two Markov chains $\widetilde{X}_t$ and $\widetilde{X}'_t$ defined by

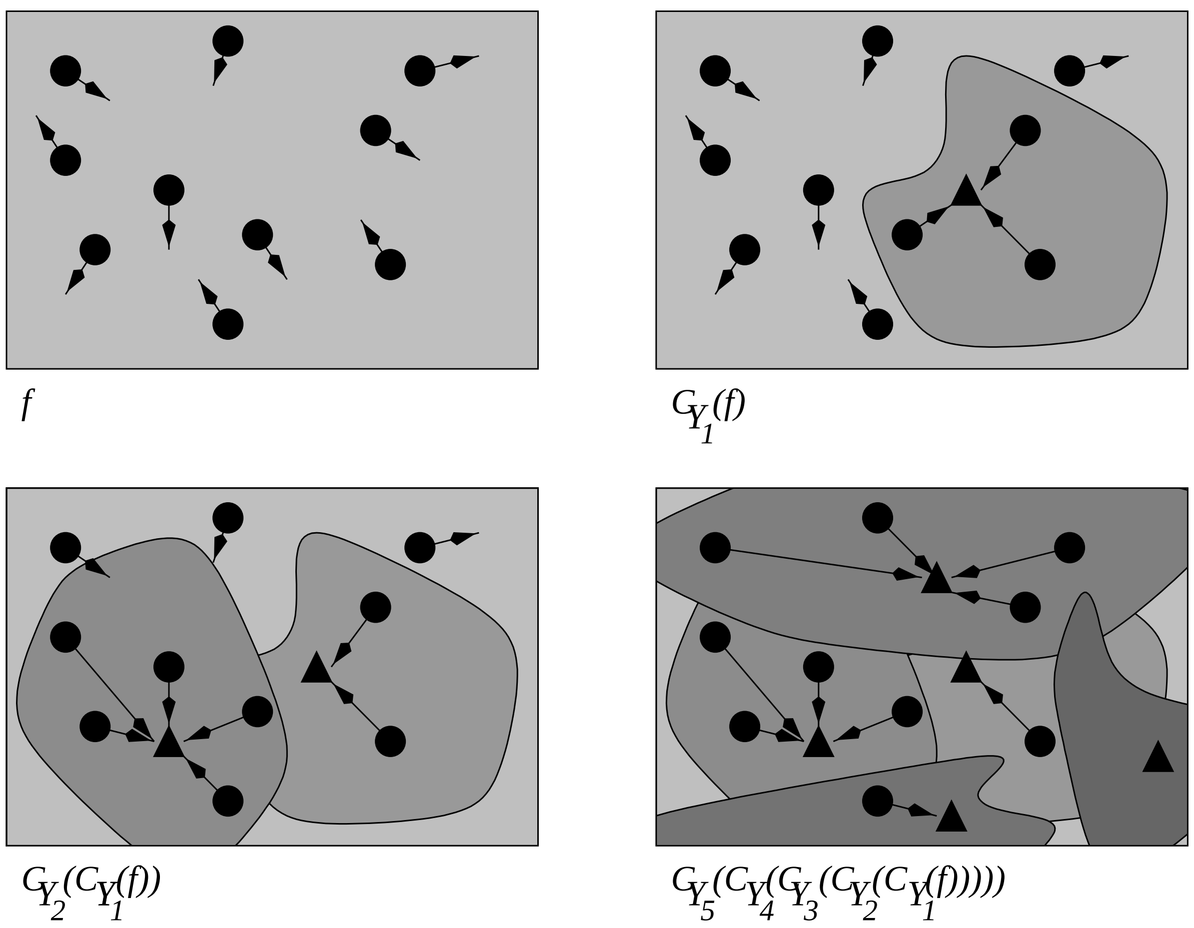
  $$\widetilde{X}_{t+1} = \widetilde{F}_t(\widetilde{X}_t), \quad \widetilde{X}'_{t+1} = \widetilde{F}_t(\widetilde{X}'_t), \quad \widetilde{X}_0, \widetilde{X}'_0 \text{ arbitrary}$$

  will couple successfully in a finite time. Moreover, *all* Markov chains of the above type, where $\widetilde{X}_0$ ranges over all points in the state space, must coalesce in a finite time!

Proof of Theorem I

Consider the basins of attraction

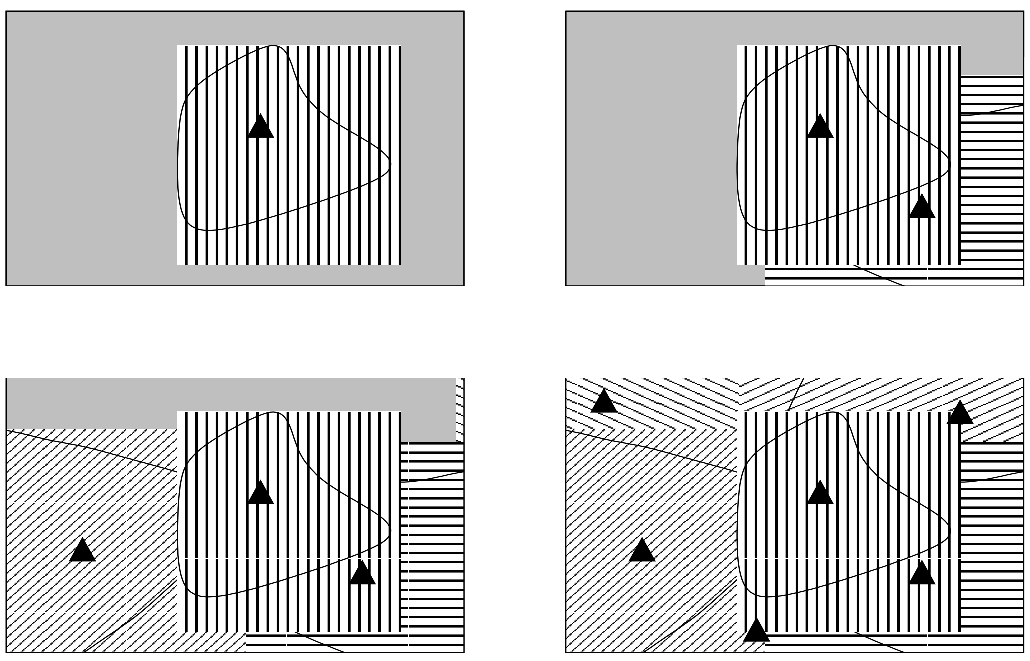
$$\text{Basin}(Y, F, \xi) = \Big\{ x : p(x, Y)q(F(x)) > \xi p(x, F(x))q(Y) \Big\}.$$



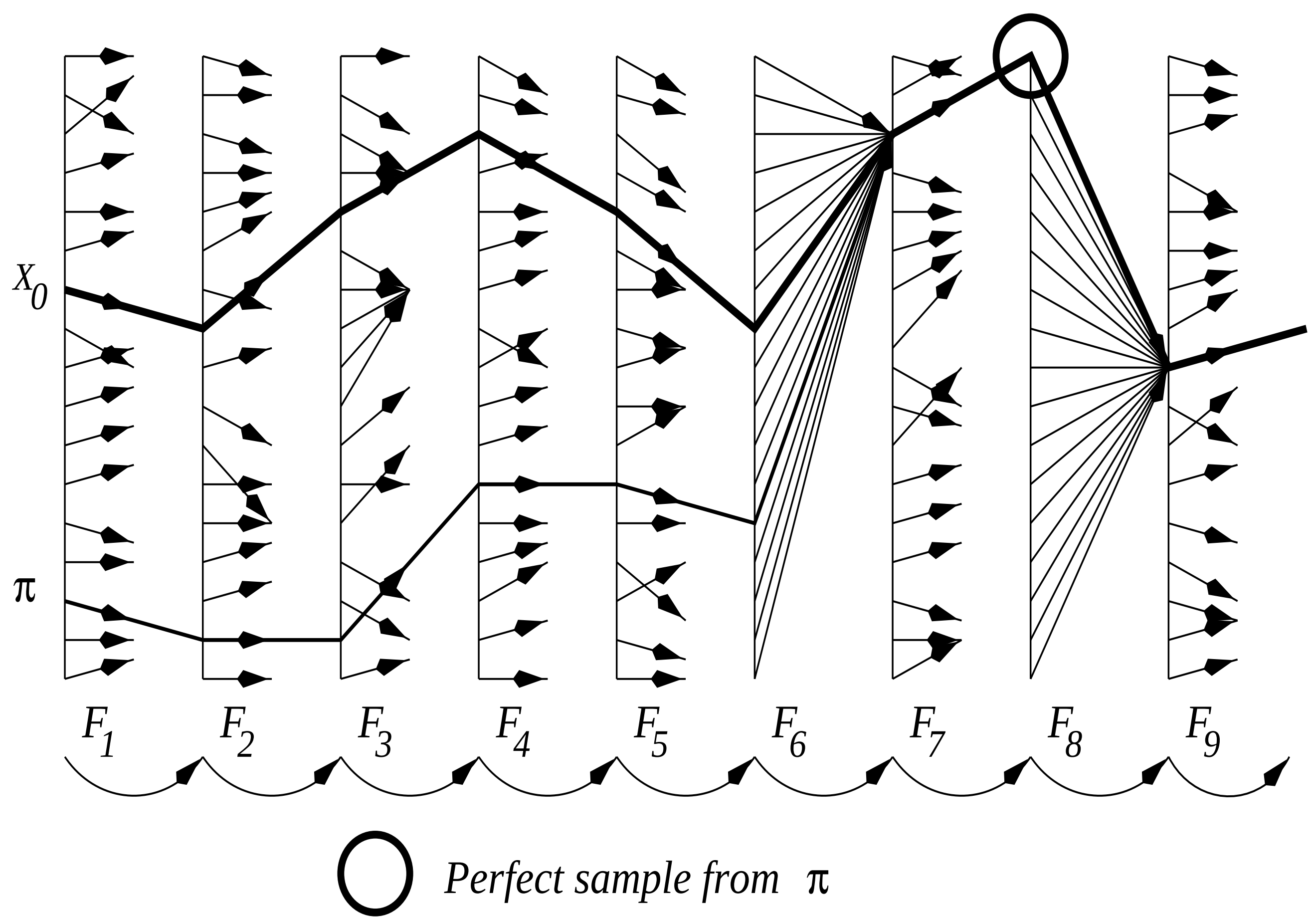
Proof of Theorem II

Consider the core regions in each basin

$$\text{Core}(Y,\xi) = \left\{ x : x \in \text{Basin}(Y,F,\xi) \text{ for all } F \right\}$$



Read Once CFTP



Assumptions for Read-Once CFTP

Assumption I A source of independent random update functions $f(x)$ exists, each satisfying

$$\int \pi(x)\mathbb{P}\Big(f(x) \in dy\Big)dx = \pi(y)dy,$$

and the simulation of π should use updates of this type. The target density $\pi(x)$ is known up to a normalization constant.

Assumption II The update functions f have a common probability density $p(x, y)$, which is known up to a normalizing constant:

$$\mathbb{P}\Big(f(x) \in dy\Big) = p(x, y)dy, \quad x \in E.$$

Building the maps F_t

(a) Resetting the state Let $b(x)$ be a proposal density, $\pi(x)/b(x) \to 0$ as $x \to \infty$

$$R_B(x) = \begin{cases} B & \text{if } \pi(B)b(x) > \psi\pi(x)b(B) \\ x & \text{otherwise,} \end{cases}$$

where $\psi \sim U[0,1]$.

(b) Coupling updates $f_1, f_2, \ldots$ Given $x \mapsto f(x)$ (Assumption I), write

$$\boldsymbol{\mathcal{C}}_Y(f)(x) = \begin{cases} Y & \text{if } p(x,Y)q(f(x)) > \xi p(x,f(x))q(Y) \\ f(x) & \text{otherwise,} \end{cases}$$

where $\xi \sim U[0,1]$ independently, and $p(x,y)$ comes from Assumption II. Choose a finite IID sequence $\mathcal{Y} = (\mathcal{Y}_\infty, \mathcal{Y}_\in, \ldots, \mathcal{Y}_\tau)$ from q.

$$\boldsymbol{\mathcal{C}}_{\mathcal{Y}}(f) := \boldsymbol{\mathcal{C}}_{Y_1,Y_2,\ldots,Y_\tau}(f) = \boldsymbol{\mathcal{C}}_{Y_\tau} \circ \boldsymbol{\mathcal{C}}_{Y_{\tau-1}} \circ \cdots \circ \boldsymbol{\mathcal{C}}_{Y_2} \circ \boldsymbol{\mathcal{C}}_{Y_1}(f).$$

(c) Definition of maps F_t Choose $f_1, \ldots, f_m$, and put

$$F(x) = \boldsymbol{\mathcal{C}}_{\mathcal{Y}_m}(f_m) \circ \cdots \circ \boldsymbol{\mathcal{C}}_{\mathcal{Y}_1}(f_1) \circ R_B(x),$$

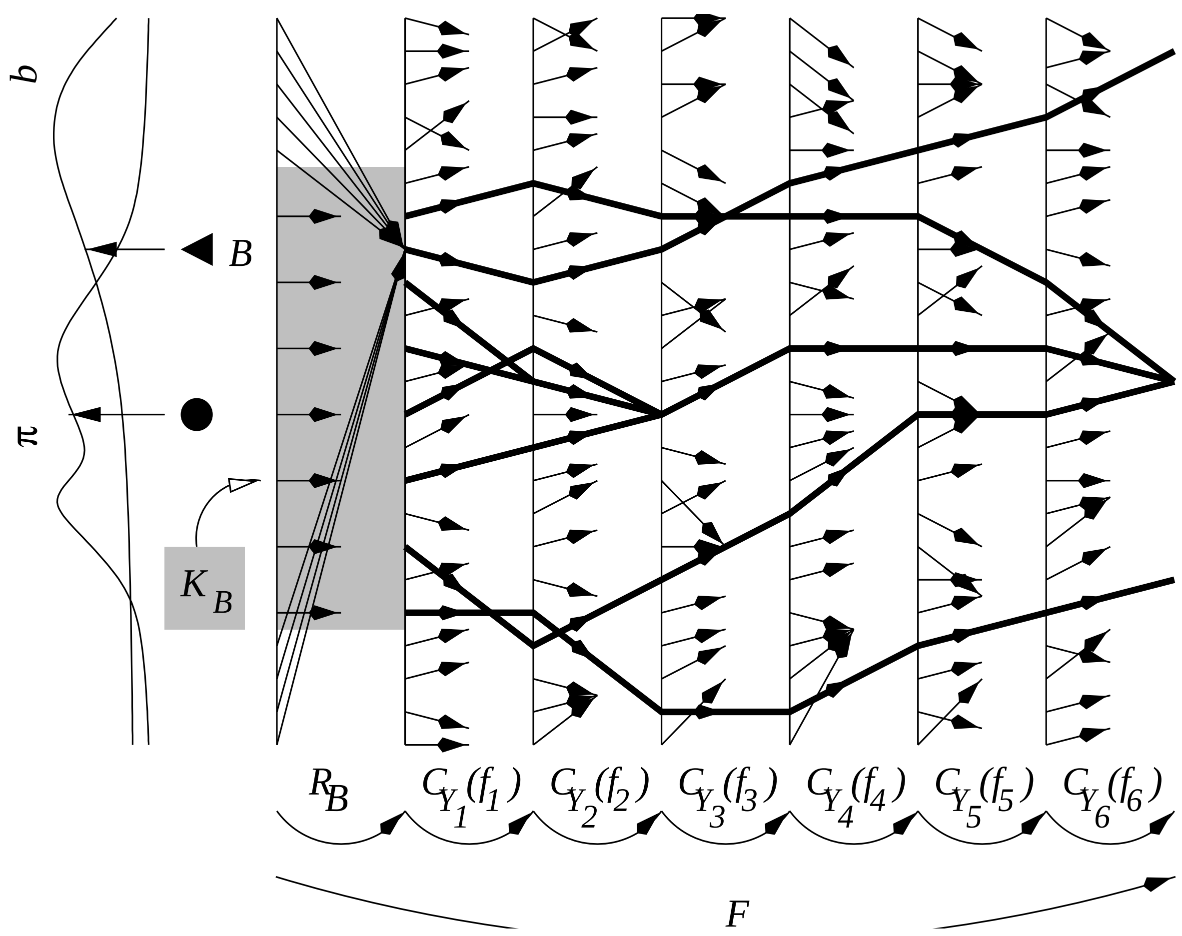
Effect of Independence Sampler

Which is more likely? For each pair (x, B), the map

$$R_B(x) = \begin{cases} B & \text{if } \pi(B)b(x) > \psi\pi(x)b(B) \\ x & \text{otherwise,} \end{cases}$$

chooses the most likely configuration

Overview of compound map F



Pump Example (Autogamma)

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

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

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

$$\beta' \sim \pi_0(\cdot \,|\, \lambda_1, \dots, \lambda_{10}) = \Gamma(\gamma + 10\alpha, \delta + \sum_{k=1}^{10} \lambda_k),$$
$$\lambda'_k \sim \pi_k(\cdot \,|\, \beta') = \Gamma(\alpha + s_k, \beta' + t_k), \quad k = 1, \dots, 10.$$

- Transition density is

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

Pump Example: Constructing $R_B(x)$

- Take $B = (B_0, \dots, B_{10})$ such that $B_0 \sim \Gamma(\gamma, \delta)$, $B_k \sim \Gamma(\alpha, B_0)$ for $k \geq 1$. Then

$$b(x)/\pi(x) = \Gamma(\gamma)^{-1}\Gamma(\alpha)^{-10}\delta^{\gamma} \exp\Big(-\sum_{k=1}^{10}(s_k \log \lambda_k - t_k\lambda_k)\Big).$$

- Consequently,

$$K_B = \Big\{x : \sum_{k=1}^{10}(s_k \log \lambda_k - t_k\lambda_k) \geq -\log \psi + \sum_{k=1}^{10}(s_k \log B_k - t_k B_k)\Big\}.$$

- Simpler to take $|\lambda| = \lambda_1 + \cdots + \lambda_{10}$,

$$K_B \subset \Big\{(\beta, \lambda_1, \dots, \lambda_{10}) : 0 \leq |\lambda| \leq \Big(\log \psi - \sum_{k=1}^{10}(s_k \log B_k - t_k B_k)\Big)/\max_j t_j\Big\}.$$

Pump Example: Constructing $\mathcal{C}_Y(f)$

- Take $Y \sim q$ where

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

- 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

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

- After simplification,

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

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