Coupling II: applications to simulation

Wilfrid S. Kendall
Department of Statistics, University of Warwick

27 July 2003

Abstract

The constructive nature of probabilistic coupling ("to build Y using the randomness of X") makes it close in spirit to the task of constructing good stochastic simulations. Recently the link between coupling and simulation has been strengthened in striking ways, resulting in so-called "exact" or "perfect simulation". This talk will introduce these developments.

1 Introduction

Häggström (2002) includes discussion of some of these ideas at the level of a student monograph.

Other impending monographs which will discuss *CFTP*: Aldous and Fill (200x); Møller and Waagepetersen (2003); Roberts *et al.*

- http://www.warwick.ac.uk/statsdept/staff/WSK/talks/durham-lms2.pdf
  http://research.microsoft.com/~dbwilson/exact/

2 More on mixing

Recall (continuous-time!) random walk on n-dimensional Boolean hypercube.

Case $n = 1$: State space $\{0, 1\}$. Let X, Y start at 0, 1. How to couple them?

- We want $0 \to 1$ transitions at rate $1/\alpha$, and $1 \to 0$ transitions at rate $1/\alpha$.
- We want X and Y to meet (at time $T_{0,1}$ in notation of previous lecture).

> **Construction:** Supply (a) Poisson process (rate $1/\alpha$) of $0 \to 1$ transitions, (b) *ditto* of $1 \to 0$ transitions. Apply them where applicable to X, Y. **Clearly** X, Y have desired distributions.

Coupling happens at first instant of combined Poisson process, so $T_{0,1}$ is Exponential of mean $\alpha/2$.

Case $n > 1$: Couple each coordinate independently. Replace α by n to play fair. So (everything) coupled at time which is maximum of n independent Exponentials each of mean $n/2$.

> **Deduce** from coupling inequality that mixing occurs before time $n \log(n)/2$:-).

Questions:

Does it make sense to return first coupled value? (Yes here, no in "nearly every" other case.)

> But suppose I run algorithm from $[-T, 0]$ for increasing T, instead of from $[0, T]$ for increasing T? (This *will* work always: CFTP.)

Olle Häggström points out, mixing actually occurs around time $n \log(n)/4$. Can you find a coupling argument to show this?

3 *CFTP*

Theorem 1 *If coalescence is almost sure then* CFTP *delivers a sample from the equilibrium distribution of the Markov chain X corresponding to the random input-output maps $F_{(-u,v]}$.*

Proof: For each $[-n, \infty)$ use input-output maps $F_{(-n,t]}$

$$X_t^{-n} \quad = \quad F_{(-n,t]}(0) \qquad \text{for } -n \leq t\,.$$

Assume finite coalescence time $-T$ for F. Then (3 lines!)

$$\begin{aligned} X_0^{-n} &= X_0^{-T} \qquad \text{whenever } -n \leq -T\,; \\ \mathcal{L}(X_0^{-n}) &= \mathcal{L}(X_n^0) \end{aligned}$$

If X converges to an equilibrium π then

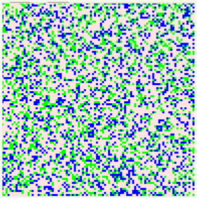
$$\operatorname{dist}_{tv}(\mathcal{L}(X_0^{-T}), \pi) = \lim_n \operatorname{dist}_{tv}(\mathcal{L}(X_0^{-n}), \pi) = \lim_n \operatorname{dist}_{tv}(\mathcal{L}(X_n^0), \pi) = 0$$

(dist_{tv} is total variation) hence giving the required result. □

3.1 Classic *CFTP*

The original Propp and Wilson (1996) idea showed how to make exact draws from the critical Ising model. A rather simpler application uses the heat-bath bath sampler to make exact draws from the sub-critical Ising model.

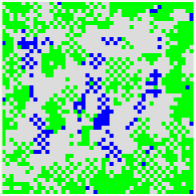
Classic *CFTP* for the Ising model (simple, sub-critical case). Heat-bath dynamics run from past; compare results from maximal and minimal starting conditions.



Green denotes both spin up; *blue* denotes both spin down; *misty-rose* denotes disagreement between maximal and minimal.

Approaches based on Swendsen-Wang ideas work for critical case. Huber (2003).
Under conditioning by noisy data, the difficulties caused by phase-transition phenomena disappear.

Classic *CFTP* for the Ising model conditioned by noisy data. Without influence from data ("external magnetic field") this Ising model would be supercritical.



Green denotes both spin up; *blue* denotes both spin down; *grey* denotes disagreement between maximal and minimal. In the summary image, *orange* denotes coalesced spin up, image spin down while *pink* denotes coalesced spin down, image spin up.

3.2 Falling leaves

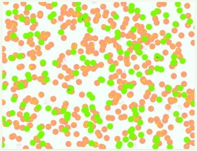
Kendall and Thönnes (1999) describe a visual and geometric application of *CFTP* in mathematical geology: this particular example being well-known to workers in the field previous to the introduction of *CFTP* itself.

Occlusion *CFTP* for the falling leaves of Fontainbleau.
(Why "occlusion"? the *CFTP* builds up the result directly.)

3.3 Dominated *CFTP* for area-interaction point processes

Dominated CFTP replaces the deterministic maximum by a known random process run backwards in time, providing starts for upper- and lower-envelope processes guaranteed to sandwich a valid simulation. It works, for example, on both attractive and repulsive area-interaction point processes (Kendall 1998; Kendall and Møller 2000).

Application of Dominated *CFTP* for attractive area-interaction point process with geometric marking using Poisson processes in disks (Kendall 1997).

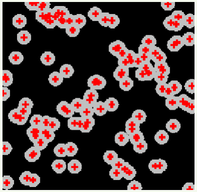


See also Huber (1999)'s notion of a "swap move". If birth proposal is blocked by just one point, then replace old point by new in a *swap*, with swap probability p_{swap} which we are free to choose. Hence "bounding chain", "sure/not sure" dichotomy.

3.4 Fast attractive area-interaction *CFTP*

Häggström, van Lieshout, and Møller (1999) describe fast *CFTP* for attractive area-interaction point processes using special features.

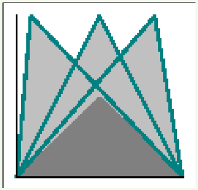
Gibbs' sampler *CFTP* for the attractive area-interaction point process as a marginal of a two-type soft-core repulsion point process.



3.5 Small-set *CFTP*

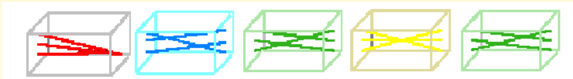
Green and Murdoch (1999) showed how to use small sets to carry out *CFTP* when the state-space is continuous with no helpful ordering. Their prescription includes the use of a partition by several small sets, to speed up coalescence.

Small set *CFTP* in nearly the simplest possible case: a triangular kernel over $[0, 1]$.



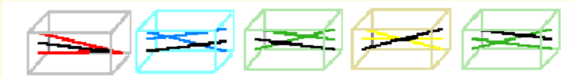
4 Fill's method (FMMR)

The alternative to *CFTP* is *Fill's algorithm* (Fill 1998; Thönnes 1999), at first sight quite different, based on the notion of a *strong uniform time* T. Fill et al. (2000) establish a profound link. We explain using "blocks" as input-output maps for a chain.



First recall that *CFTP* can be viewed in a curiously redundant fashion as follows:

- **Draw from equilibrium** $X(-T)$ and run forwards;
- continue to increase T until $X(0)$ is coalesced;
- return $X(0)$.



Key observation: By construction, $X(-T)$ is independent of $X(0)$ and T so . . .

- Condition on a convenient $X(0)$;
- Run X backwards to a fixed time $-T$;
- Draw blocks conditioned on the X transitions;
- **If** coalescence **then** **`return`** $X(-T)$ **else** **`repeat`**.

> "It's a kind of magic . . ."
> Queen

Is there a *dominated* version of Fill's method?

5 Price of perfection

Coupling of couplings:...

$$|p_t(x_1,y) - p_t(x_2,y)| \quad \leq$$
$$\leq \quad |\mathbb{P}[X_1(t)=y|X_1(0)=x_1] - \mathbb{P}[X_2(t)=y|X_2(0)=x_2]| \leq$$
$$|\mathbb{P}[X_1(t)=y|\tau>t, X_1(0)=x_1] - \mathbb{P}[X_2(t)=y|\tau>t, X_2(0)=x_2]|$$
$$\times\, \mathbb{P}[\tau>t|X(0)=(x_1,x_2)]$$

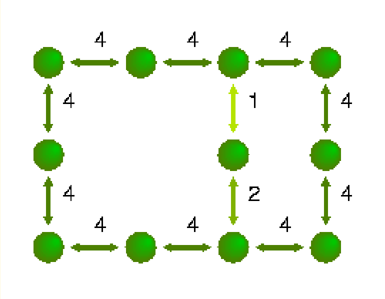
Suppose $|p_t(x_1,y) - p_t(x_2,y)| \approx c\exp(-\mu_2 t)$
while $\mathbb{P}[\tau>t|X(0)=(x_1,x_2)] \approx c\exp(-\mu t)$
Let X^* be a coupled copy of X but begun at (x_2,x_1):

$$|\,\mathbb{P}[X_1(t)=y|\tau>t, X_1(0)=x_1] - \mathbb{P}[X_2(t)=y|\tau>t, X_2(0)=x_2]\,|$$
$$= \quad |\,\mathbb{P}[X_1(t)=y|\tau>t, X(0)=(x_1,x_2)] -$$
$$\mathbb{P}[X_1^*(t)=y|\tau>t, X^*(0)=(x_2,x_1)]\,|$$

So let σ be time when X, X^* couple:

$$\leq \mathbb{P}[\sigma>t|\tau>t, X(0)=(x_1,x_2)] \quad (\approx c\exp(-\mu' t))$$

Thus $\mu_2 \geq \mu' + \mu$.



Burdzy and Kendall (2000), Mountford and Cranston (2000), Burdzy and Chen (2002), also Kumar and Ramesh (2001).

5.1 Domination and small sets

Corcoran and Tweedie (2001) describe how to mix dominated *CFTP* and small set *CFTP*. The upper envelope process must be formulated carefully ....
The idea is close to Foster-Liapunov techniques for assessing geometric ergodicity *etc* for Markov chains. Foster-Liapunov uses a Liapunov function V to deliver a controlled supermartingale off a small set:

$$\mathbb{E}\left[V(X_{n+1})|X_n\right] \quad \leq \quad \lambda V(X_n) + \beta\, \mathbb{I}\left[X_n \in C\right] .$$

Temptation: define dominating process using V. There is an interesting link – Rosenthal (2002) draws it even closer – but:

Existence of Liapunov function
doesn't ensure dominated *CFTP*

There are perverse examples satisfying the supermartingale inequality, but failing the stochastic dominance required of $V(X)$ by dominated *CFTP* ... :-(.[1]

[1] *Later:* I have discovered how to fix this using sub-sampling.

6 Combinations and variations

(Other things I'd have liked to talk about . . .)

6.1 Layered Multishift *CFTP*

(Wilson (2000b) and further work by Corcoran and Schneider (2002)) Issue: how to draw simultaneously from Uniform$(x, x+1)$ for all $x \in \mathbb{R}$, and to try to couple the draws? Answer: draw a uniformly random *unit span integer lattice*, ... Now think about more general distributions!

6.2 Read-once randomness

Wilson (2000a) shows how to avoid a conventional requirement of *CFTP*, to re-use randomness used in each cycle.

6.3 Perfect simulation for Peirls' contours *etc*

(Ferrari et al. 2002). The Ising model can be reformulated in an important way by looking only at the *contours* (lines separating ± 1 values). In fact these form a "non-interacting hard-core gas", permitting (theoretically) Ferrari et al. (2002) to apply their variant of perfect simulation (*Backwards-Forwards Algorithm*).

6.4 Randomness Recycler

Fill and Huber (2000) introduce a quite different form of perfect simulation! Overleaf is how they apply their *Randomness Recycler* algorithm to the problem of drawing a random independent subset X of a graph G, weighted exponentially by number of points in X.

Start: $V = \emptyset$, $x \equiv 0$. End: $V = G$, x indicates X membership.

```
while V ≠ G:
    v ← choice(G − V)    # Choose v from G \ V
    V.add(v)
    if uniform(0, 1) ≤ 1/(1 + alpha):
        x[v] ← 0    # Skip v with prob 1/(1 + α)
    else:
        x[v] ← 1    # or tentatively include it ...
        nbd ← [ ]    # ... iterate thro' neighbours
        for w ∈ neighbourhood(v):    # Valid?
            nbd.append(w)
            if x[w] = 1:    # If not valid ...
                x[w] ← 0    # ... remove all
                x[v] ← 0    # "contaminated" vertices
                V ← V − [v] − nbd
                break    # and move on
```

References

Aldous, D. J. and J. A. Fill [200x]. Reversible Markov chains and random walks on graphs. Book in preparation .

Burdzy, K. and Z.-Q. Chen [2002]. Coalescence of synchronous couplings. *Probability Theory and Related Fields 123*(4), 553–578.

Burdzy, K. and W. S. Kendall [2000, May]. Efficient Markovian couplings: examples and counterexamples. *The Annals of Applied Probability 10*(2), 362–409. Also University of Warwick Department of Statistics Research Report 331.

Corcoran, J. and U. Schneider [2002]. Shift and scale coupling methods for perfect simulation. Preprint, University of Colorado.

Corcoran, J. N. and R. L. Tweedie [2001]. Perfect sampling of ergodic Harris chains. *The Annals of Applied Probability 11*(2), 438–451.

Ferrari, P. A., R. Fernández, and N. L. Garcia [2002]. Perfect simulation for interacting point processes, loss networks and Ising models. *Stochastic Process. Appl. 102*(1), 63–88. doi>

Fill, J. A. [1998]. An interruptible algorithm for exact sampling via Markov Chains. *The Annals of Applied Probability 8*, 131–162.

Fill, J. A. and M. Huber [2000]. The randomness recycler: a new technique for perfect sampling. In *41st*

This is a rich hypertext bibliography. Journals are linked to their homepages, and icons which link to preprints are inserted where available. Stable URL links (as provided for example by JSTOR or Project Euclid) have been added where known. Access to such URLs is not universal: in case of difficulty you should check whether you are registered (directly or indirectly) with the relevant provider.

Annual Symposium on Foundations of Computer Science (Redondo Beach, CA, 2000), pp. 503–511. Los Alamitos, CA: IEEE Comput. Soc. Press.

Fill, J. A., M. Machida, D. J. Murdoch, and J. S. Rosenthal [2000]. Extension of Fill's perfect rejection sampling algorithm to general chains. *Random Structures and Algorithms* *17*(3-4), 290–316.

Green, P. J. and D. J. Murdoch [1999]. Exact sampling for Bayesian inference: towards general purpose algorithms (with discussion). In J. Bernardo, J. Berger, A. Dawid, and A. Smith (Eds.), *Bayesian Statistics 6*, pp. 301–321. The Clarendon Press Oxford University Press. Presented as an invited paper at the 6th Valencia International Meeting on Bayesian Statistics, Alcossebre, Spain, June 1998.

Häggström, O. [2002]. *Finite Markov chains and algorithmic applications*, Volume 52 of *London Mathematical Society Student Texts*. Cambridge: Cambridge University Press.

Häggström, O., M. N. M. van Lieshout, and J. Møller [1999]. Characterisation results and Markov chain Monte Carlo algorithms including exact simulation for some spatial point processes. *Bernoulli* *5*(5), 641–658. Was: Aalborg Mathematics Department Research Report R-96-2040.

Huber, M. [1999]. The swap move: a tool for building better Markov chains. In *10th Annual Symposium on Discrete Algorithms*.

Huber, M. [2003]. A bounding chain for Swendsen-Wang. *Random Structures and Algorithms* *22*(1), 43–59. doi>

Kendall, W. S. [1997]. On some weighted Boolean models. In D. Jeulin (Ed.), *Advances in Theory and Applications of Random Sets*, Singapore, pp. 105–120. World Scientific. Also: University of Warwick Department of Statistics Research Report 295.

Kendall, W. S. [1998]. Perfect simulation for the area-interaction point process. In L. Accardi and C. C. Heyde (Eds.), *Probability Towards 2000*, New York, pp. 218–234. Springer-Verlag. Also: University of Warwick Department of Statistics Research Report 292.

Kendall, W. S. and J. Møller [2000, September]. Perfect simulation using dominating processes on ordered state spaces, with application to locally stable point processes. *Advances in Applied Probability 32*(3), 844–865. Also University of Warwick Department of Statistics Research Report 347.

Kendall, W. S. and E. Thönnes [1999]. Perfect simulation in stochastic geometry. *Pattern Recognition 32*(9), 1569–1586. Also: University of Warwick Department of Statistics Research Report 323.

Kumar, V. S. A. and H. Ramesh [2001]. Coupling vs. conductance for the Jerrum-Sinclair chain. *Random Structures and Algorithms 18*(1), 1–17. doi>

Møller, J. and R. Waagepetersen [2003]. *Statistics inference and simulation for spatial point processes*. Monographs on Statistics and Applied Probability. Boca Raton: Chapman and Hall / CRC.

Mountford, T. S. and M. Cranston [2000]. Efficient coupling on the circle. In *Game theory, optimal stopping, probability and statistics*, Volume 35 of *IMS Lecture Notes Monogr. Ser.*, pp. 191–203. Beachwood, OH: Institute of Mathematical Statistics.

Propp, J. G. and D. B. Wilson [1996]. Exact sampling with coupled Markov chains and applications to statistical mechanics. *Random Structures and Algorithms 9*, 223–252. doi>

Rosenthal, J. S. [2002]. Quantitative

convergence rates of Markov chains: A simple account. *Electronic Communications in Probability* *7*, no. 13, 123–128 (electronic).

Thönnes, E. [1999]. Perfect simulation of some point processes for the impatient user. *Advances in Applied Probability* *31*(1), 69–87. Also University of Warwick Statistics Research Report 317.

Wilson, D. B. [2000a]. How to couple from the past using a read-once source of randomness. *Random Structures and Algorithms* *16*(1), 85–113.

Wilson, D. B. [2000b]. Layered Multishift Coupling for use in Perfect Sampling Algorithms (with a primer on CFTP). In N. Madras (Ed.), *Monte Carlo Methods*, Volume 26 of *Fields Institute Communications*, pp. 143–179. American Mathematical Society.
