Markov chain Monte Carlo and Perfect Simulation

Lecture at Aristotle University of Thessaloniki

Wilfrid S Kendall

Department of Statistics, University of Warwick

15 May 2024

Warwick Statistics

Introduction

Figure 1: Αριστοτέλης 384–322 BCE

Aristotle:

- "Pleasure in the job puts perfection in the work."
- "The more you know, the more you know you don't know."

Handout available on the web: either use the `QR-code`

or visit https://wilfridskendall.github.io/talks/Thessaloniki-2024/.

Sketch of MCMC (I)

Figure 2: Edward Teller (1908-2003)

The original Markov chain Monte Carlo method (MCMC) was introduced by Metropolis *et al.* (1953). The senior author was Edward Teller ("father of the H-bomb").

[Fermi once said that] Teller was the only monomaniac he knew who had several manias: see Brown & May (2004).

Sketch of MCMC (II)

- Markov chain basics:
  - Transition probabilities $p(a,b)$ (or transition rates in continuous time: unified view using exponential distribution);
  - Equilibrium probabilities $\pi(a)$, balance,

    detailed balance $\pi(a)p(a,b) = \pi(b)p(b,a)$, reversibility;
  - Aperiodicity, recurrence (and uniform and geometric recurrence);
  - under detailed balance we can condition by forbidding transitions;
- We can modify any chain, transition probabilities $p(a,b)$, to leave a specified target distribution $\pi(a)$ invariant, by censoring each possible transition $a \to b$ with probability $\alpha(a,b) \in [0,1]$ such that

  $$\alpha(a,b)\pi(a)p(a,b) = \alpha(b,a)\pi(b)p(b,a);$$
- Common choice: Metropolis-Hastings

  $$\alpha(a,b) = \min\{1, (\pi(b)p(b,a))/(\pi(a)p(a,b))\}.$$
- If result still irreducible aperiodic, then $\pi(a)$ is its long-term equilibrium.
- This is MCMC, now of intense interest to statisticians.
- But, physicists always remind us, physicists got there fifty years earlier!

Sketch of MCMC (III)

Given the $\pi(a)$, how to design a Markov chain to have this as equilibrium?

1. Independence sampler: draw from a fixed probability distribution, apply Metropolis-Hastings;
2. Random walk Metropolis or RWM: propose move using a random walk, apply Metropolis-Hastings;
3. Metropolis-adjusted Langevin or MALA: make a Gaussian jump shifted using gradient of $\log \pi$, apply Metropolis-Hastings.

Can mix-and-match! RWM is often favourite: flexible, not too complicated.

Issues:

- (A) Burn-in: *How long* till approximate equilibrium?
- (B) Scaling: *How big* should be the RWM jump?

Question (B) is about how to get fast mixing. There is a beautiful and useful theory, but that is for another day.

Question (A) is what this lecture is all about.

Sketch of MCMC (IV)

- MCMC practicalities: Burn-in: what to do about it?
  - Theory tends to be much too pessimistic. Example: Zanella (2015a, 2015b) developed statistical methods for Anglo-Saxon history: a *simplified* model appeared to converge approximately in 10^5 steps (about 1 week on compute cluster), *versus* 10^9 steps in theory (around 2 centuries);
  - Is (a) one long run better or (b) many short runs? (Option (b) requires starts of short runs spread "evenly" over the sample space — almost as hard in high dimensions as the original problem!)
  - Diagnostics? (Meta-theorem: for any diagnostic technique there is a chain for which the technique is deceptive!)
  - Conclusion: effective MCMC requires very careful thought about appropriate length of run — think deeply about the problem!
- Can there ever be a better way?

Perfect Simulation

- The idea of exact simulation / Coupling from the Past (CFTP) / perfect simulation Propp & Wilson (1996) (Persi Diaconis: "Like seeing the landscape of Mars for the first time");
- Ideas (of "*classic CFTP*"):
  - extending simulation *backwards* through time,
  - exploit monotonicity by coupling maximal and minimal processes,
  - seek coalescence;
- Details for *random-walk-CFTP*, which can be boosted as above to provide simple image reconstruction of an image using Ising model, Propp & Wilson (1996) show how to vary a clever algorithm to get exact samples for critical Ising model (this is what impressed Diaconis);
- "Perfect simulation" (WSK, 1998): because everyone knows it isn't going to be perfect, whereas people might imagine "exact simulation" would somehow miraculously defeat numerical approximation error :-).

An example and some theory

- An intensely visual example, which helps many people see intuitively what is going on here, is *DeadLeaves-CFTP* (WSK & Thönnes, 1999) (technically, *Occluded CFTP*);
- What about cases where monotonicity fails? or there isn't a sensible "maximal" process? WSK (1998):
  - cross-couple upper and lower envelope processes,
  - dominate by amenable "dominating process" (time-reversible, can draw from equilibrium, can couple target processes below dominating process);
- Theoretical limits: *in principle*
  - *Classical CFTP* equivalent to uniform ergodicity (Foss & Tweedie, 1998).
  - *Dominated CFTP* is achievable under geometric ergodicity (WSK, 2004).
  - It is even possible to carry out Dominated CFTP in some **non**-geometrically ergodicity cases [Connor & WSK (2007); *nb* corrigendum];
- We can use *Dominated CFTP* to carry out perfect simulation for stable point processes (WSK & Møller, 2000);
- Detailed expositions are given by WSK (2005), Huber (2015). WSK (2015) shows how to implement CFTP in R.

Applications to Queues and Epidemics

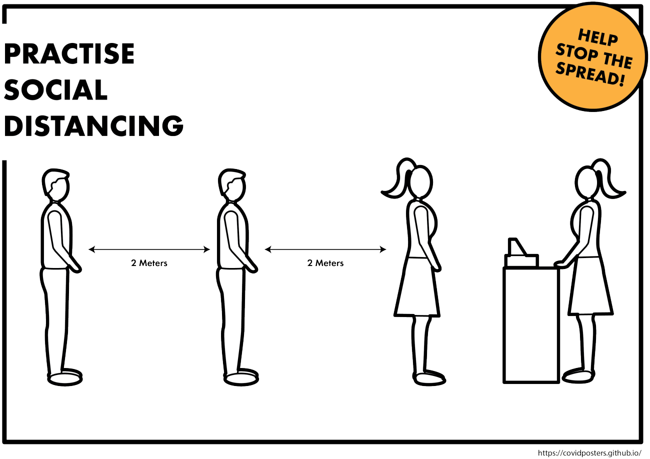


Figure 3: An illustration introducing *both* queues *and* epidemics!

Perfect Queues

The simplest queuing model (Poisson arrivals, exponential service times, single server) can be analyzed very thoroughly;

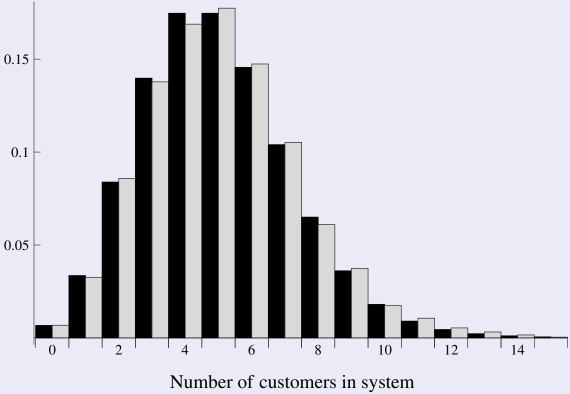
- Poisson arrivals are not unreasonable, but exponential service times are ludicrous. The $M/G/1$ case of general service time *for just one server* can use the "embedded chain" (sample at instants of departure);
- Multi-server case: computation of eg waiting-time distribution is out of reach so use simulation (and insights from Kiefer & Wolfowitz, 1955);
- Sigman (2011) shows how to do CFTP in the "super-stable" case (traffic so low that it could have been handled by just one server), using Dominated CFTP and comparing to a "Processor-sharing" discipline.

- Connor & WSK (2015) show how to extend Sigman (2011), showing how Dominated CFTP can be applied to simulate (sub-critical!) queues perfectly (and this has now been generalized by others to the case of non-Poissonian inter-arrival times). (Technical point: pathwise domination requires service times to be assigned in order of commencement of service!)
  - dominate $M/G/cFCFS$ (FCFS means first come first served) by $M/G/cRA = [M/G/1RA]^c$ (RA means assign to individual servers on arrival);
  - use fact that *workload* of $M/G/1FCFS$ is same as $M/G/1PS$ which can be run backwards in time in equilibrium; (PS means arrivals Share Processor.)
  - so $[M/G/1PS]^c$ can be used to provide Dominated CFTP.
- Connor & WSK (2015) also compare
  (a) CFTP coupling when dominating process empties,
  (b) a faster CFTP coupling using upper and lower processes starting respectively at dominating process and at empty state.

Warwick Statistics

Results (I)

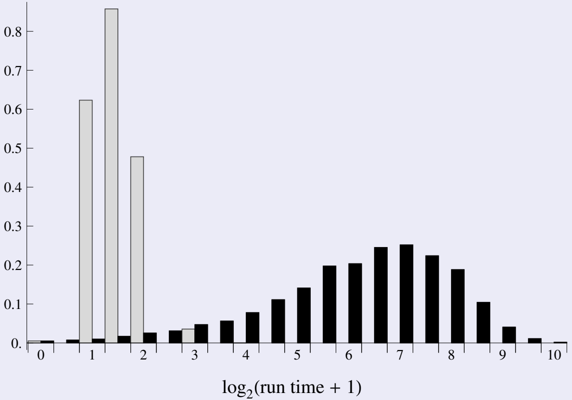
Histogram of customer numbers for $M/M/c$ queue in equilibrium: arrival rate 10, service rate 2, and 10 servers, comparing theory (available for $M/M/c$ queue) with results of Connor & WSK (2015) algorithm.



Results (II)

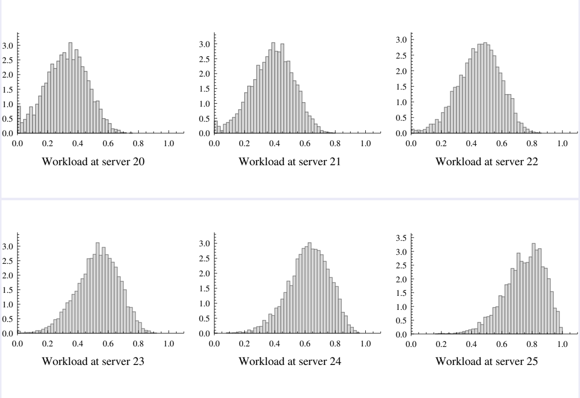
Comparison of log-run times for

- (a) CFTP coupling when dominating process empties (solid bars),
- (b) a faster CFTP coupling using upper and lower processes (grey bars).



Results (III)

Workload distributions of six most heavily loaded servers in an $M/G/c = 25$ queue with Uniform(0, 1) service time distribution and arrival rate 25.



Perfect Epidemics

- Even the simple case of deterministic S-I-R permits only *partial* closed-form solution;
- Suppose **only** the "removal times" are observed? Can the infection times be simulated perfectly? (Given basic parameters.)
- **YES**: work in progress by Connor and Kendall. Here is a GIF illustrating this for a real-life small-pox epidemic;
- (be clear about assumptions!)
- (indicate how the perfect simulation algorithm can be used as a high-dimensional integration device to enable simulation-based Bayesian inference!).

Conclusion

- You don't always have to put up with burn-in issues when doing MCMC;
- CFTP works even for significantly complex and relevant models of real-life phenomena;
- *Of course* really detailed models are still going to resist perfect simulation: but it will always be helpful to compare with a simpler model (using fewer parameters!).
- CFTP is clearly an important tool to be considered by the investigator seeking to do accurate and informative MCMC.

References I

Brown, H. & May, M. (2004) Edward Teller in the Public Arena. *Physics Today*, **57**, 51–53.

Connor, S.B. & WSK (2007) Perfect simulation for a class of positive recurrent Markov chains. *Annals of Applied Probability*, **17**, 781–808.

Connor, S.B. & WSK (2015) Perfect simulation of M/G/c queues. *Advances in Applied Probability*, **47**, 1039–1063.

Foss, S.G. & Tweedie, R.L. (1998) Perfect simulation and backward coupling. *Stochastic Models*, **14**, 187–203.

Huber, M.L. (2015) *Perfect Simulation*. Boca Raton: Chapman; Hall/CRC.

Kiefer, J. & Wolfowitz, J. (1955) On the Theory of Queues With Many Servers. *Transactions of the American Mathematical Society*, **78**, 1.

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., & Teller, E. (1953) Equation of State Calculations by Fast Computing Machines. *The Journal of Chemical Physics*, **21**, 1087.

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

R Development Core Team (2010) R: A Language and Environment for Statistical Computing.

Sigman, K. (2011) Exact simulation of the stationary distribution of the FIFO M/G/c queue. *Journal of Applied Probability*, **48**, 209–213.

References II

WSK (1998) Perfect Simulation for the Area-Interaction Point Process. *Probability towards 2000* (Accardi, L. & Heyde, C.C. eds). Springer-Verlag, pp. 218–234.

WSK (2004) Geometric ergodicity and perfect simulation. *Electronic Communications in Probability*, **9**, 140–151.

WSK (2005) Notes on Perfect Simulation. Singapore: World Scientific, pp. 93–146.

WSK (2015) Introduction to CFTP using R. *Stochastic geometry, spatial statistics and random fields*, *Lecture notes in mathematics*. Springer, pp. 405–439.

WSK & Møller, J. (2000) Perfect simulation using dominating processes on ordered spaces, with application to locally stable point processes. *Advances in Applied Probability*, **32**, 844–865.

WSK & Thönnes, E. (1999) Perfect simulation in stochastic geometry. *Pattern Recognition*, **32**, 1569–1586.

Zanella, G. (2015b) Bayesian Complementary Clustering, MCMC, and Anglo-Saxon Placenames (PhD Thesis).

Zanella, G. (2015a) Random partition models and complementary clustering of Anglo-Saxon place-names. *Annals of Applied Statistics*, **9**, 1792–1822.

Technical information

Image	Attribution	
Aristotle	After Lysippos	*Public domain* via Wikimedia Commons
Edward Teller	Lawrence Livermore National Laboratory restored by w:User:Greg L, Papa Lima Whiskey	*CC BY-SA 3.0* via Wikimedia Commons
Perfect Ising	Result of code written by WSK	
Dead leaves	Result of code written by WSK	
Queues	https://covidposters.github.io/	*Open source*
$M/M/c$ customers	Result of code written by Stephen Connor	
$M/M/c$ runtimes	Result of code written by Stephen Connor	
$M/M/c$ loads	Result of code written by Stephen Connor	
Epidemic	Result of code written by WSK	

These notes were produced from `Thessaloniki-2024.qmd`:

`git` repository information.
Version: `0.0.46 (Wed May 8 16:11:07 2024 +0100)` Author: Wilfrid Kendall <`W.S.Kendall@warwick.ac.uk`> Date: Tue Wed 8 17:55:14 2024 +0100