OxWaSP: Probability & Approximation

Lecture 2: Perfect Simulation and Queues

Stephen B. Connor Wilfrid S. Kendall
stephen.connor@york.ac.uk w.s.kendall@warwick.ac.uk

Supported by EPSRC EP/J009180, EP/K013939.

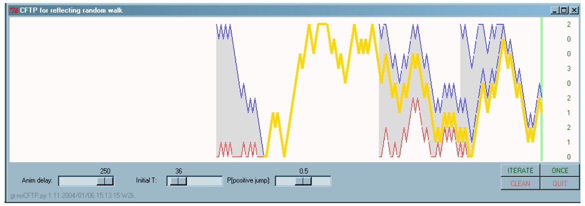
Department of Mathematics, University of York

Department of Statistics, University of Warwick

22 October 2018

Perfect simulation: the basic idea.

Propp and Wilson (1996) Coupling from the Past (*CFTP*).
Very basic case: reflected simple random walk on state-space $\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$.



Ingredients:

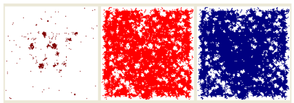
- Past → Present, not Present → Future;
- Monotonicity;
- Capture all starts using upper and lower processes;
- Re-use randomness.

Other examples of perfect simulation

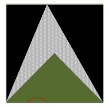
“Backwards from the past” constructions to analyse very general queues were introduced by Loynes 1962.

1. Dead Leaves example (WSK and Thönnes 1999):

2. Simple image analysis (Propp and Wilson 1996):



3. Use of small sets (Murdoch and Green 1998):



WSK (2015): “Introduction to *CFTP* using *R*”.

CFTP for "bounded" queues

1. Classic *CFTP* requires "bounded" state-space.
2. So focus initially on queues with finite capacity ("turn customers away when waiting room is full"). Seek to find exact draws from equilibrium.
3. Murdoch and Takahara (2006) study *CFTP* for: $M/M/1/c$, $M/M/c/c$, $GI/M/c/c$, $GI/G/c/c$. Here c means, if arrivals find c people already in the system then they go away never to return.
4. Integrated Masters Dissertation, Liu (2015): $M/D/1/c$ based on residual total workload. Problem: Need to alter upper process very carefully to allow for failure of monotonicity!

The trouble with M/D/1/c

Monotonicity can fail!

- Suppose Q_1 is at capacity, with current customer almost having completed service;
- and suppose Q_2 has one spare place, so Q_2 is "smaller" than Q_1.
- Someone arrives! There is room in Q_2 but not in Q_1.
- So Q_2 can jump above Q_1. This can break monotonicity.
- Fix:
  use $M/D/1/c+1$ for upper process,
  use $M/D/1/c-1$ for lower process.

  Failure of boundedness would be much trickier.

Classic CFTP and unbounded state-spaces

What to do if the state-space is "unbounded"? Example: the Lindley recursion for $GI/G/1$ provides a representation that *nearly* delivers *CFTP*, but unboundedness means we can't determine when to stop (see Foss and Tweedie 1998).

- "Unbounded" state-spaces can sometimes be small! (Key: is the chain uniformly ergodic? If so then classic *CFTP* is possible *in principle*; Foss and Tweedie 1998. Practical example from point process theory: Häggström, van Lieshout, and Møller 1999)
- One could truncate and hope. (Somewhat misses point of perfect simulation.)
- Murdoch (2000): induce uniform ergodicity by occasional proposals using independence sampler.
- Dominated *CFTP* (*domCFTP*): next section.

Basic idea for dominated *CFTP*

- Classic *CFTP* fails when we can't construct both upper and lower processes.
- So try to replace these by random processes in statistical stationarity ("dominating processes").
- If we can guarantee a coupling which maintains "sandwiching" for the resulting envelope processes, and if they coalesce to run as the target process, then we can rescue the *CFTP* algorithm.
- We'll need to be able to do the following:
  - draw from joint stationary distribution of the dominating processes;
  - simulate them backwards in time;
  - use their paths to simulate envelope processes forwards in time to coalesce as target processes;
  - check for coalescence, and ensure this happens on a reasonable time-scale.

Birth-death chain

Nonlinear birth and death processes: I

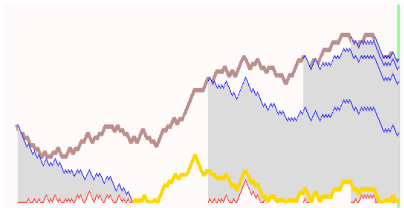
Adapted from WSK (1997).

- X is nonlinear immigration-death process:
  $X \to X - 1$ at rate μX;
  $X \to X + 1 \ldots \quad \alpha_X$ where $\alpha_X \leq \alpha_\infty < \infty$.
  No maximum (not uniformly ergodic), so no classic *CFTP*!
- *Bound* by *linear* immigration-death process Y:
  $Y \to Y - 1$ at rate μY;
  $Y \to Y + 1 \ldots \quad \alpha_\infty$.
- Produce X from Y by censoring births and deaths:
  if $Y \to Y - 1$ then $X \to X - 1$ with c.prob. X/Y;
  if $Y \to Y + 1 \ldots \quad X \to X + 1 \ldots \quad \alpha_X/\alpha_\infty$.

The domCFTP construction

Nonlinear birth and death processes: II

- Given trajectory of Y, build trajectories of X starting at every $0 \leq X_0 \leq Y_0$ and then staying below Y.
- Because Y is *reversible*, with *known equilibrium* (detailed balance!) we can simulate Y backwards, then run forwards with sandwiched X realizations.
- Identify "golden thread" of X started arbitrarily far back in time. (Animation describes analogous random walk.)



Dominated CFTP: a recipe

Basic ingredients:

- dominating process:
  - draw from equilibrium;
  - simulate backwards in time.
- sandwiching:
  $\text{Lower}_1 \precsim \text{Lower}_2 \precsim \ldots \precsim \text{Targets} \precsim \ldots \precsim \text{Upper}_2 \precsim \text{Upper}_1$
- coalescence:
  eventually a Lower and an Upper process must coalesce to produce a Target process.
- computability:
  is the price tag of Perfection too high?

Generalities

- Corcoran and Tweedie (2001) produce a general recipe for *domCFTP* for Harris chains. Roughly speaking:
  - Identify small set to which dominating process returns;
  - Check for regeneration at each visit;
  - Check domination maintained if no regeneration.
- WSK (2004) used this idea to show *domCFTP* is possible in principle for all geometrically ergodic chains:
  - Use Foster-Lyapunov criterion to build set-valued dominating process;
  - Model using a discrete-time $D/M/1$ queue!
  - Ensure ergodicity by sub-sampling;
  - Exploit regeneration (uses small-set theory).
- Connor and WSK (2007b, 2007a) extends this to some non-geometrically ergodic positive-recurrent chains.

Super-stable M/G/c and *domCFTP* (I)

Sigman (2011) pioneered *domCFTP* for multi-server queues.
Key step: find amenable dominating process.

- Restrict to super-stable case (queue remains stable if we remove all but one server).
- Workload of $M/G/1$ queue as stable dominating process.
- Same $M/G/1$ workload if PS not FCFS.
- But $M/G/1[PS]$ is dynamically reversible (so we can reverse time in equilibrium).
- Recover $M/G/1[FCFS]$ from workload $M/G/1[PS]$.
- $M/G/c$: *FCFS* workload smaller than $M/G/1[FCFS]$ (FCFS is optimally efficient: Wolff 1977, 1987).
- In particular, domination is sample-wise if each service time is assigned at initiation of service.
- Coalescence forced when $M/G/1[PS]$ empties. (Finite mean if finite second moment of service time.)

Super-stable M/G/c and *domCFTP* (II)

Natural objections to what has been achieved so far

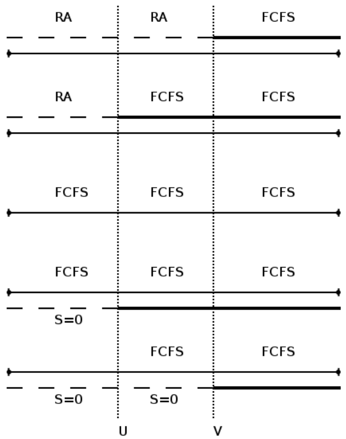
The Sigman (2011) result is pioneering, but only partial.

1. Coalescence is ensured by running backwards in time till dynamically-reversed $M/G/1[PS]$ empties.
   But this will be inefficient if the target $M/G/c$ workload is such that $M/G/1$ is nearly unstable.
2. Worse, the interesting case for $M/G/c$ is exactly when the $M/G/1$ is not stable.
   (When else would you pay for more than one server?!)
3. Sigman (2012) describes an importance-sampling approach when the corresponding $M/G/1$ is not stable.
   Unfortunately mean termination time is infinite.

Stable M/G/c and *domCFTP* (I)

Connor and Kendall (2015): dominate with $M/G/c[RA]$.
RA = "random assignment", so c copies of $M/G/1$, independent fractional arrival processes.

- Evidently stable if and only if $M/G/c$ is stable!
- Domination! (extend Asmussen 2003, Chapter XII)



Stable M/G/c and *domCFTP* (I)

Algorithm 1 (coalescence when $M/G/c[RA]$ empties):
NB: Processes run backwards are crowned with a tilde.

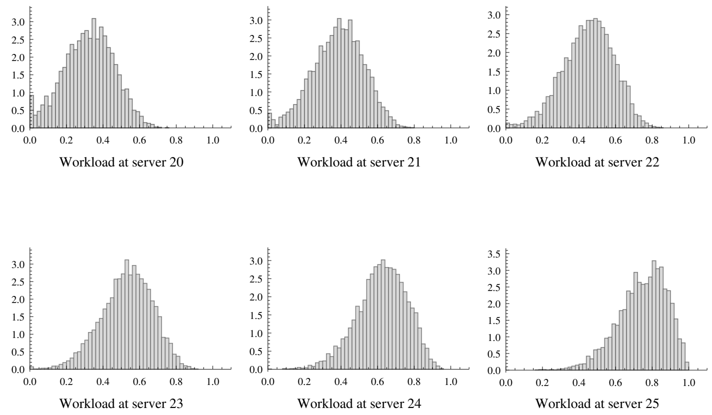
1. *Consider* $(M/G/1[PS])^c$ *process* $\tilde{Y}$, *run backwards in statistical equilibrium. Draw from* $\tilde{Y}(0)$.
2. *Simulate c components of reversed-time* $(\tilde{Y}(\tilde{t}) : \tilde{t} \geq 0)$ *over* $[0, \tilde{\tau}]$, *where* $\tilde{\tau}$ *is smallest reversed time such that all components are empty at* $\tilde{\tau}$.
3. *Use* $(\tilde{Y}(\tilde{t}) : \tilde{t} \in [0, \tilde{\tau}])$ *to construct (dynamic) time reversal, and so build* $(Y(t) : \tau \leq t \leq 0)$, *an* $M/G/c[RA]$ *process (set* $\tau = -\tilde{\tau}$*).*
4. *Use Y to evolve X, an* $M/G/c[FCFS]$ *process, over* $[\tau, 0] = [-\tilde{\tau}, 0]$, *started in the empty state.*
   Now return $X(0)$.

Stable M/G/c and *domCFTP* (II)

Algorithm 2 (use upper and lower processes):

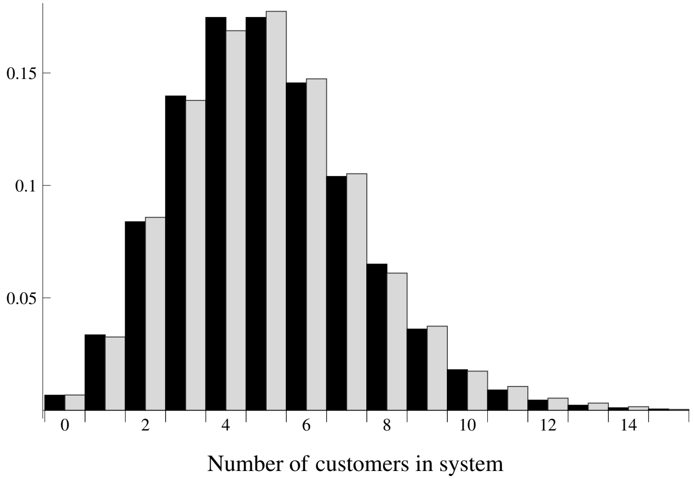
1. *Consider a $(M/G/1[PS])^c$ process $\widetilde{Y}$, run backwards in statistical equilibrium. Draw from $\widetilde{Y}(0)$.*
2. *Fix suitable $\widetilde{T} = -T$. Evolve queue for server j (independently of other servers) until first time $\widetilde{\tau}_j \geq \widetilde{T}$ that this server is empty, for $j = 1, \ldots, c$.*
3. *Construct $M/G/1[FCFS]$ Y_j over corresponding reversed time interval $[-\widetilde{\tau}_j, 0]$, for $j = 1, \ldots, c$.*
4. *Produce lists $\mathcal{L}^*_T$, $\mathcal{L}_T$ of service durations, arrival times.*
5. *Construct upper sandwiching process, $U_{[T,0]}$ over $[T, 0]$.*
6. *Construct lower sandwiching process, $L_{[T,0]}$ over $[T, 0]$.*
7. *Check for coalescence.* Otherwise, extend.

Results of *domCFTP* for $M/G/c$ (I)



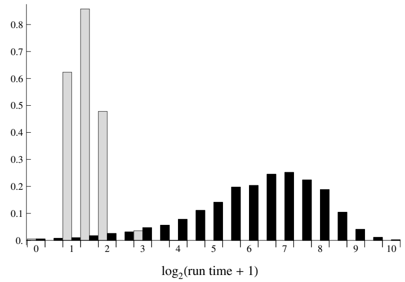
Equilibrium distribution of final 6 coordinates of Kiefer-Wolfowitz workload vector:
arrival rate $\lambda = c = 25$, service durations Uniform($[0, 1]$).
(5000 draws, Algorithm 2.)

Results of *domCFTP* for $M/G/c$ (II)



Number of customers for $M/M/c$ queue in equilibrium when $\lambda = 10$, $\mu = 2$ and $c = 10$. Black bars summarize theoretical number of customers in system; light grey bars summarize 5000 draws, Algorithm 2. χ^2-test: p-value 0.62.

Results of *domCFTP* for $M/G/c$ (III)



Distribution of time taken to detect coalescence under Algorithms 1, 2 applied to $M/M/c$ queue, 5000 runs, $\lambda = 10$, $\mu = 2$ $c = 10$.
Black bars show distribution of $\log_2(\widetilde{\tau} + 1)$ for Algorithm 1 ($\widetilde{\tau}$ is first time at which $\widetilde{Y}$ empties).
Light grey bars show distribution of $\log_2(\widetilde{T} + 1)$ for Algorithm 2, where $\widetilde{T}$ is smallest time needed to detect coalescence using binary back-off.
(Algorithm 2 speed-up also suggested by calculations.)

Conclusion

It is highly feasible to produce exact simulations of stable $M/G/c$ queues (if service distribution has finite second moment) using *domCFTP*.

Further possibilities:

- $M/G/c$ behaviour is not encoded by first two moments of service distribution (Gupta et al. 2009).
- General input processes: Blanchet and Wallwater (2014), Blanchet et al. (2015).
- Omnithermal *domCFTP* for range c (Connor 2016).
- Networks? Sigman (2013), Blanchet and Dong (2013).
- Compartment models? Connor and WSK are working on SIR epidemics.
- "Efficient simulation"?

Asmussen, S. (2003).
Applied probability and queues.
Springer.

Blanchet, J. and J. Dong (2013, dec).
Perfect sampling for infinite server and loss systems.

Blanchet, J., Y. Pei, and K. Sigman (2015).
Exact sampling for some multi-dimensional queueing models with renewal input.
arXiv 1512.07284, 23.

Blanchet, J. H. and A. Wallwater (2014).
Exact Sampling of Stationary and Time-Reversed Queues.
arXiv 1403.8117, 1–30.

Cai, Y. and WSK (2002, jul).
Perfect simulation for correlated Poisson random variables conditioned to be positive.
Statistics and Computing 12, 229–243.

Connor, S. B. (2016).
Omnithermal Perfect Simulation for Multi-server Queues.
arXiv 1610.06012(October), 1–13.

Connor, S. B. and W. S. Kendall (2015, feb).
Perfect Simulation of M/G/c Queues.
Advances in Applied Probability 47(4), 1039–1063.

Connor, S. B. and WSK (2007a).
Correction: Perfect Simulation for a Class of Positive Recurrent Markov Chains.
The Annals of Applied Probability 17(5/6), 1808–1810.

Connor, S. B. and WSK (2007b, jan).
Perfect simulation for a class of positive recurrent Markov chains.
The Annals of Applied Probability 17(3), 781–808.

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

Foss, S. G. and R. L. Tweedie (1998).
Perfect simulation and backward coupling.
Stochastic Models 14(1), 187–203.

Gupta, V., M. Harchol-Balter, J. G. Dai, and B. Zwart (2009, aug).
On the inapproximability of M/G/K: why two moments of job size distribution are not enough.
Queueing Systems 64(1), 5–48.

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.

Huber, M. (2015, nov).
Perfect Simulation.
Boca Raton: Chapman and Hall/CRC.

Kiefer, J. and J. Wolfowitz (1955).
On the theory of queues with many servers.
Transactions of the American Mathematical Society 18(1), 1–18.

Kolmogorov, A. N. (1936).
Zur Theorie der Markoffschen Ketten.
Math. Annalen 112, 155–160.

Liu, I.-W. (2015).
Simulating Queues Perfectly.
Msc, University of Warwick.

Loynes, R. (1962).
The stability of a queue with non-independent inter-arrival and service times.
Math. Proc. Camb. Phil. Soc. 58(3), 497–520.

Lubetzky, E. and A. Sly (2014, jul).
Universality of cutoff for the Ising model.
arXiv 1407.1761, 24–27.

Murdoch, D. J. (2000).
Exact sampling for Bayesian inference: unbounded state spaces.
In *Monte Carlo methods (Toronto, ON, 1998)*, Volume 26 of *Fields Inst. Commun.*, pp. 111–121. Providence, RI: Amer. Math. Soc.

Murdoch, D. J. and P. J. Green (1998).
Exact Sampling from a Continuous State Space.
Scandinavian Journal of Statistics Theory and Applications 25, 483–502.

Murdoch, D. J. and G. Takahara (2006, jan).
Perfect Sampling for Queues and Network Models.
ACM Transactions on Modeling and Computer Simulation 16(1), 76–92.

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.

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

Sigman, K. (2012, nov).
Exact simulation of the stationary distribution of the FIFO M/G/c queue: the general case for $\rho<c$.
Queueing Systems: Theory and Applications 70(1), 37–43.

Sigman, K. (2013).
Using the M/G/1 queue under processor sharing for exact simulation of queues.
Annals of Operations Research.

Wolff, R. W. (1977).
An upper bound for multi-channel queues.
Journal of Applied Probability 14(4), 884–888.

Wolff, R. W. (1987).
Upper bounds on work in system for multichannel queues.
Journal of Applied Probability 24(2), 547–551.

WSK (1997).
On some weighted Boolean models.
In D. Jeulin (Ed.), *Advances in Theory and Applications of Random Sets*, Singapore, pp. 105–120. World Scientific.

Warwick Statistics

WSK (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. University of Warwick Department of Statistics: Springer-Verlag.

WSK (2004, oct).
Geometric Ergodicity and Perfect Simulation.
Electronic Communications in Probability 9(Paper 15), 140–151.

WSK (2015).
Introduction to CFTP using R.
In V. Schmidt (Ed.), *Stochastic Geometry, Spatial Statistics and Random Fields*, Lecture Notes in Mathematics, pp. 405–439. Springer.

Warwick Statistics

WSK and J. Møller (2000, sep).
Perfect simulation using dominating processes on ordered state spaces, with application to locally stable point processes.
Advances in Applied Probability 32(3), 844–865.

WSK and E. Thönnes (1999).
Perfect Simulation in Stochastic Geometry.
Pattern Recognition 32(9), 1569–1586.

Warwick Statistics