End-Point Conditioned Markov Process Simulation: Variance Reduction via Bisection

Søren Asmussen & Asger Hobolth

Aarhus University, Denmark

http://home.imf.au.dk/asmus

MCQMC 2010

Warszawa, August 20, 2010

Outline

$X \in \{1, \ldots, N\}$ **finite Markov jump process on** $[0, T]$

Aim:

Efficient simulation of X **conditioned on** $X(0) = a$, $X(T) = b$
(Markovian bridge)

Applications:

statistical inference for
Markov processes
finance
human genetics
genomics

Notation

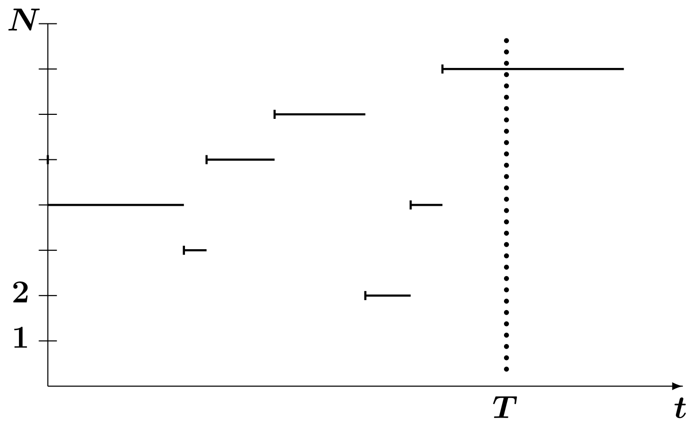
State space $\{1, \ldots, N\}$
Rate (intensity) matrix

$Q = (q_{ij})_{i,j=1,\ldots,N}$

jump $i \longrightarrow j$ at rate q_{ij}, $i \neq j$

exponential(q_i) holding time in i, $q_i = -q_{ii} = \sum_{j \neq i} q_{ij}$

new state $j \neq i$ w.p. q_{ij}/q_i

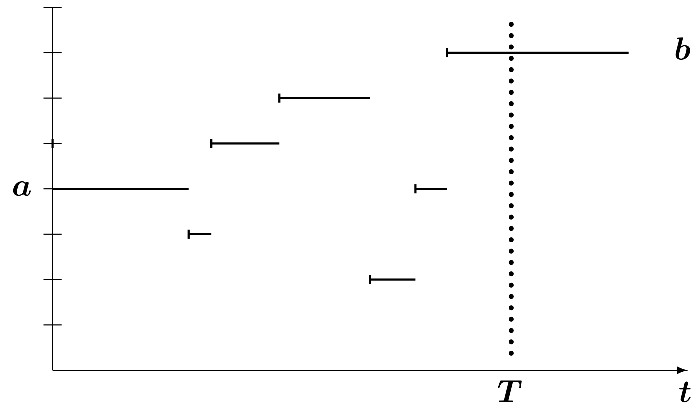


Transition matrix $t \longrightarrow s + t$:

$$P(s) = \mathrm{e}^{Qs} = \sum_{n=0}^{\infty} \frac{(Qs)^n}{n!}$$

(other ways to compute)

Naive unconditioned simulation



$X(0) = a$

first holding time $\exp(q_a)$

new state j w.p. q_{aj}/q_a

next holding time $\exp(q_j)$

new state k w.p. q_{jk}/q_j

and so on until T passed

Conditioned simulation?

Why not rejection sampling:

accept first path s.t. $X(T) = b$?

MLE for Markov processes

X finite Markov

Holding time rates q_i

Jump probabilities $\theta_{ij} = q_{ij}/q_i$

Complete observations in $[0, T]$

T_i = time in state i

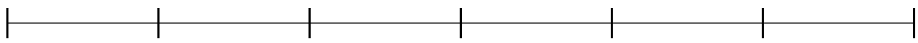
N_{ij} = # jumps $i \to j$

$N_i = \sum_{j \neq i} N_{ij}$

$$\widehat{q}_i = \frac{N_i}{T_i}, \quad \widehat{\theta}_{ij} = \frac{N_{ij}}{N_i}$$

Incomplete observations

$X(0), X(h), \ldots, X(nh) = X(T)$



Incomplete observations

Stochastic EM algorithm:

Iterative, $q_i^{(n)}$, $\theta_{ij}^{(n)}$

Update to $n+1$: in

$$\widehat{q}_i = \frac{N_i}{T_i}, \quad \widehat{\theta}_{ij} = \frac{N_{ij}}{N_i}$$

replace T_i etc. by $\mathbb{E}_{q_i^{(n)}, \theta_{ij}^{(n)}} T_i$

$\mathbb{E}_{q_i^{(n)}, \theta_{ij}^{(n)}}$: by simulation

Parameter estimates ergodic MC (not a.s. convergent)

Variant:

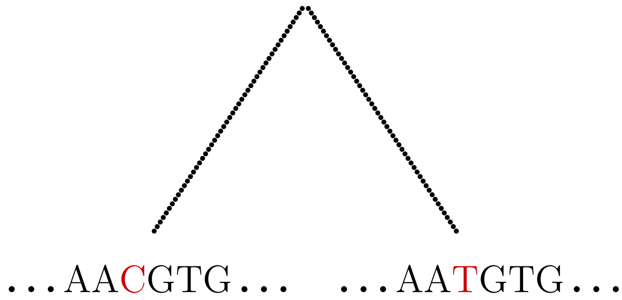
Stochastic approximation EM algorithm

Robbins-Monro type

Same computations

convergent

Coalescence



Estimate time
to common ancestor

At each site

Markovian mutations

at known rates q_A, q_C, q_G, q_T

at known trans. prob's

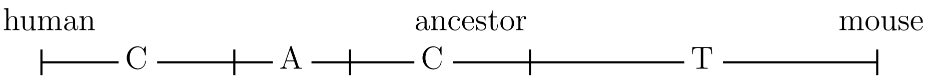
(q_{AG}/q_A for A⟶G etc.)

model as time reversible

from human to mouse:

Markov process,

known endpoints



Example of Q: Kimura 1980

	A	G	C	T
A	$-\alpha-2\beta$	α	β	β
G	α	$-\alpha-2\beta$	β	β
C	β	β	$-\alpha-2\beta$	α
T	β	β	α	$-\alpha-2\beta$

Detailed balance: $\pi_G q_{GT} = \pi_T q_{TG}$

$\Rightarrow$ reversible

Transition matrices $P(s) = \mathrm{e}^{Qs}$
tractable

Diagonalization

$$Q = U(\lambda_1 \dots \lambda_N)_{\mathrm{diag}} U^{-1}$$

$\lambda_1 \dots \lambda_N$ real

$$\mathrm{e}^{Qs} = U(\mathrm{e}^{\lambda_1 s} \dots \mathrm{e}^{\lambda_N s})_{\mathrm{diag}} U^{-1}$$

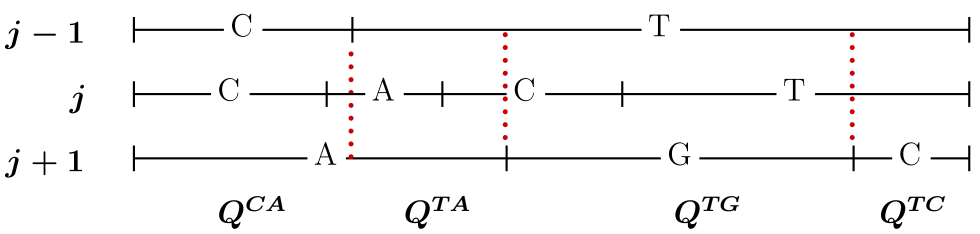
Model too simple:

only one site

dependence on

neighboring sites

e.g. Q at j only depends on states at $j-1$ and $j+1$



Gibbs sampling

Simulate X at change points

(inhom. conditioned MC)

end-p. cond. between

Aim:

simulate X conditioned on

$X(0) = a$, $X(T) = b$

Naive rejection algorithm:

Start from $X(0) = a$

First jump time exponential(q_a)

new state a' chosen w.p. $q_{aa'}/q_a$

($q_a = -q_{aa}$)

Repeat until T passed

Reject path if $X(T) \neq b$.

Inefficient if $p_{ab}(T)$ small

Otherwise hard to beat

(no analytical calculations)

Other algorithms:

Uniformization

Direct simulation

Uniformization

Unconditioned implementation:

Choose $\lambda \geq \max_i q_i$

Simulate Poisson(λ) process
on $[0, T]$

$N(T)$ Poisson(λT)

Uniform epochs

State changes at epochs
according to $p_{ij} = q_{ij}/\lambda$

Dummy transitions

Conditioned implementation:

$$\mathbb{P}(N(T) = n) \propto e^{-\lambda T} \frac{(\lambda T)^n}{n!} p_{ab}^n$$

Sample n, uniform epochs

State change at epoch k, $i \to j$
according to $p_{ij} p_{jb}^{n-k-1} / p_{ib}^{n-k}$

Inefficient if q_i variable

Otherwise often good

Origin: bioinformatics literature

Direct simulation

(Hobolth, 2008)

Start from $X(0) = a$

First jump to a'

with conditioned probability

First jump time from $f_{aa'b}(t)$

(conditioned density)

Repeat until T passed

Messy formulas and r.v. generation

First transition time density $f_{aa'b}(t)$:

$$\frac{q_{aa'}}{p_{ab}(T)} \sum_j U_{a'j} U^{-1}_{jb} \mathrm{e}^{T\lambda_j} \mathrm{e}^{-t(\lambda_j + q_a)}$$

Inversion

Still often good when reversible

Comparisons:

Hobolth & Stone,

Ann. Appl. Statist. (2009)

Blackwell 03
Bladt & Sørensen 05, 09
Hobolth & Jensen 05, 10
Fearnhead & Sherlock 06
Siepel, Pollard & Haussler 06

Bisection for BM in $[0, 1]$

$B(0) = 0$, $B(1) \sim N(0, 1)$

$B(1/2)$ from conditional normal given $B(0), B(1)$

$B(1/4)$ given $B(0), B(1/2)$

$B(3/4)$ given $B(1/2), B(1)$

Go on until 2^{-K},
desired resolution

Brownian bridge given $B(1)$

Linear interpolation: wavelet

Easy way to prove continuity
(Steele's book)

Bisection for Gaussian processes

$X(0) \sim N(\mu_0, \sigma_0^2)$

$X(1) \mid X(0) \;\sim\; N(\mu_{1|0}(X(0)), \sigma_{1|0}^2)$

$X(1/2)$ from conditional normal
given $X(0), X(1)$

$X(1/4), X(3/4)$
given $X(0), X(1/2), X(1)$

Step k: 2^k Gaussians given $2^k{+}1$

SA-Hobolth:

Bisection algorithm for
end-point conditioned CTMP

Discrete events problem

Resolution 0

Key idea:

3 types of intervals:

0,1 or $\geq$2 jumps

0 or 1: easy, can be terminated

Initialization for the case

$$X(0) = X(T) = a$$

No jumps w.p. $\mathrm{e}^{-q_a T}$

W.p. $\mathrm{e}^{-q_a T}/p_{aa}(T)$,

take $X(t) \equiv a$

Otherwise at least two jumps,

go on to recursion

Initialization for the case $X(0) = a \neq b = X(T)$

One jump w.p.

$$r_{ab}(T) = \int_0^T q_a e^{-q_a t} \frac{q_{ab}}{q_a} e^{-q_b(T-t)} \, dt$$

W.p. $r_{ab}(T)/p_{ab}(T)$,
take $X(0:T)$ with one jump.

Jump time density $\propto$ integrand

Otherwise at least two jumps,
go on to recursion

Notes:

r_{ab} explicit, =

$$\begin{cases} q_{ab} e^{-q_b T} T & q_a = q_b \\ q_{ab} e^{-q_b T} \frac{1}{q_b - q_a} (e^{(q_b - q_a)T} - 1) & q_a \neq q_b \end{cases}$$

Simulation of jump time:
truncated exponential, $q_a > q_b$,
minus trunc. exp., $q_a < q_b$
uniform, $q_a = q_b$

Recursion, $a = b$

Know $[0, T]$ has ≥ 2 jumps
$X(T/2) = c$?
And # jumps in $[0, T/2]$, $[T/2, T]$?
$e_a = \mathrm{e}^{-q_a T/2}$, $r_{ab} = r_{ab}(T/2)$,
$p_{ab} = p_{ab}(T/2)$

case	number of jumps in $[0, T/2]$	number of jumps in $[T/2, T]$	(unconditional) probability
1	0	0	$e_a e_a$
2	0	≥ 2	$e_a(p_{aa} - e_a)$
3	≥ 2	0	$(p_{aa} - e_a)e_a$
4	≥ 2	≥ 2	$(p_{aa} - e_a)(p_{aa} - e_a)$
5	1	1	$r_{ac} r_{ca}$
6	1	≥ 2	$r_{ac}(p_{ca} - r_{ca})$
7	≥ 2	1	$(p_{ac} - r_{ac})r_{ca}$
8	≥ 2	≥ 2	$(p_{ac} - r_{ac})(p_{ca} - r_{ca})$

Sample among cases 2-8
(total appr. $4N$)
Finish intervals with 0 or 1 jumps
Go on with ≥ 2

$a \neq b$: very similar.

Algorithmically:

Initialization

In step k:

Bookkeeping:

List of recorded jump times and new values
$(t_1, y_1), (t_2, y_2), \ldots, (t_{n_k}, y_{n_k})$

List of left endpoints of intervals with ≥ 2 jumps
$a_1 2^{-k}T, a_2 2^{-k}T, \ldots, a_{m_k} 2^{-k}T$

Computations

$P(2^{-k-1}T)$

$e_i = \exp\{-q_i 2^{-k-1}T\}$

$$r_{ij}(2^{-k-1}) = \int_0^{2^{-k-1}} q_i \mathrm{e}^{-q_i t} \frac{q_{ij}}{q_i} \mathrm{e}^{-q_j (T-t)} \, \mathrm{d}t$$

Sample case $(1{:}4N)$ m_k times

Find jump times and new values in intervals with 1 jump

Update lists

Theorem *All 4 algorithms have*

$$c(T) \;=\; \max_{a,b} E_{ab}(\text{cost for } T) \;=\; O(T)$$

in T.

Proof for Bisection

Choose k_T with $T2^{-k_T} \approx 1$,
i.e. $k_T \approx \log_2 T$.

$$\begin{aligned}
c(T) &\le d + 2c(T/2) \\
&\le d + 2d + 4c(T/4) \\
&\le d + 2d + 4d + 8c(T/8) \\
&\le d \sum_{1}^{k_T} 2^k \;+\; 2^{k_T} c(1) \\
&\approx d_1 2^{k_T} \;+\; 2^{k_T} c(1) \\
&\approx (d_1 + c(1))T\,.
\end{aligned}$$

Proof of $c(1) < \infty$: Poisson bound.

$$\begin{aligned}
c_i(T) &\le k_i T,\; i = 1, 2, 3, 4. \\
c_i(T) &\le \ell_i q T,\; q = \max_i q_i.
\end{aligned}$$

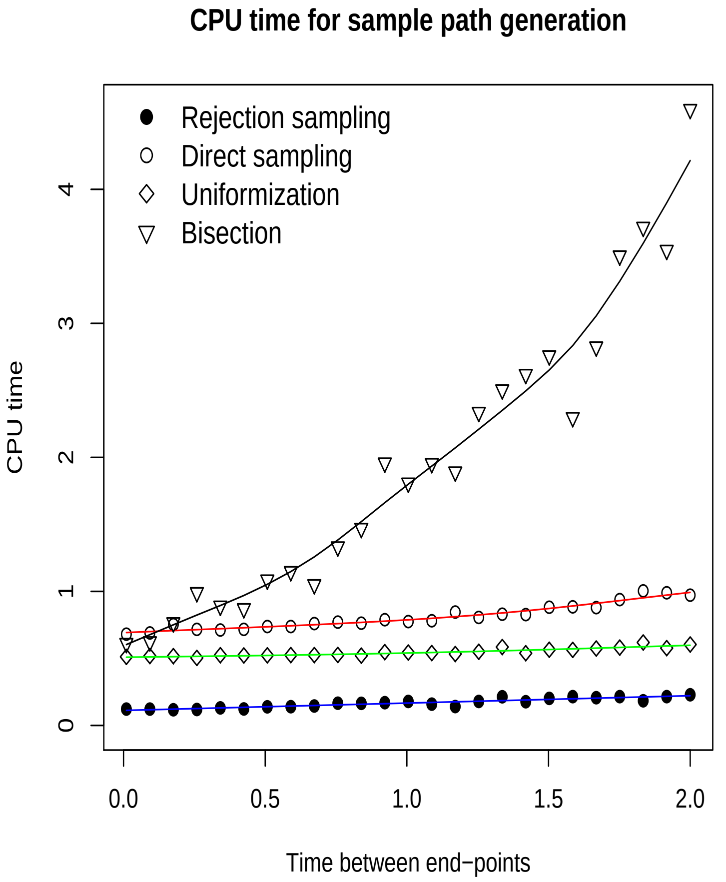
Comparisons:

depend on situation

Bad example

$$Q = (1/10)\begin{bmatrix} -11 & 6 & 3 & 2 \\ 4 & -9 & 3 & 3 \\ 2 & 3 & -9 & 4 \\ 2 & 3 & 6 & -11 \end{bmatrix}$$

and $X(0) = 1$ **and** $X(T) = 2$.



CPU time for sample path generation
Rejection sampling
Direct sampling
Uniformization
Bisection
CPU time
0
1
2
3
4
0.0
0.5
1.0
1.5
2.0
Time between end−points

Figure 1:

Variance reduction

Return to BM

Estimate price

$$\mathbb{E}e^{-rT}\Big[\int_0^T \exp\{rt+\sigma B(t)\}\,\mathrm{d}t \;-\; K\Big]^+$$

of Asian option

Precision improvement:
reduce variability of B by
QMC or stratification
on "most important" variables

1st MI: $B(T)$

2nd MI: $B(T/2)$

3rd MI: $B(T/4)$, $B(3T/4)$

Simulate B by bisection

Importance sampling example for Gaussian processes

Estimate $\mathbb{P}(\tau < \infty)$,

$\tau = \inf\{t : X(t) \geq x\}$

E.g. FBM: $X(t) = B^H(t) - ct$

Most likely time $t(x)$

maximize t**-density**

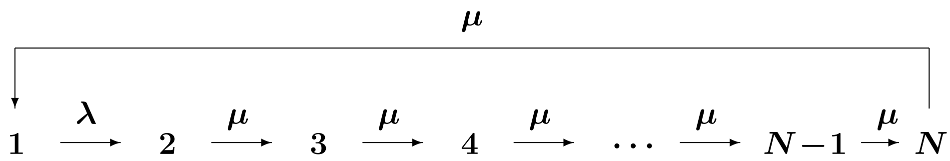
$$\frac{1}{\sqrt{2\pi}\sigma_t} \exp\{-(x - \mu_t)^2/2\sigma_t^2\}$$

Simulate $X\big(t(x)\big)$ **as** $\mathcal{N}(x, \widetilde{\sigma}^2_{t,x})$,

instead of $\mathcal{N}(\mu_t, \sigma_t)$

Paths in $[0, t(x)]$ **and** $[t(x), T]$**: bisection**

CTMC example:
cyclic process



Estimate $\mathbb{P}_{11}$(one cycle in $[0, 1]$)

Attempt variance reduction
via the r.v. determining
the $4N$ choices in first step
(midpoint state, 0,1,≥ 2 jumps
in $[0, 1/2]$ and $[1/2, 1]$)

Often $\mu = \lambda$; then
$\mathbb{E}_1$(return time to 1) $= N/\lambda$
$\lambda >> N$: many cycles typical
$\lambda \approx N$: one cycle typical
$\lambda << N$: zero cycles typical

Stratification
QMC
Importance sampling

Stratification

Stratify midpoint state

(not # jumps)

Proportional allocation

$R = R_1 + \cdots + R_N$ replications

$R_s \approx \mathbb{P}\big(X(1/2) = s\big) \cdot R$

In R_s, $X(1/2) = s$

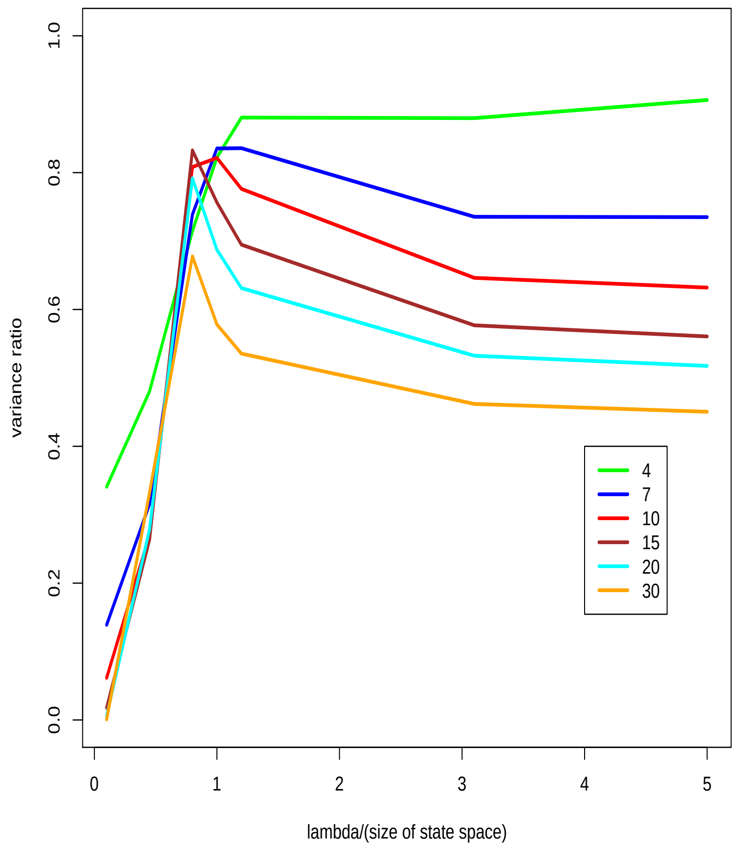

variance ratio
0.0
0.2
0.4
0.6
0.8
1.0
0
1
2
3
4
5
lambda/(size of state space)
4
7
10
15
20
30

Figure 2:

QMC

One-dimensional Halton sequence (van der Corput, base 2)

r.v. determining $4N - 1$ choices

$\boldsymbol{\mu = \lambda = N}$, $\boldsymbol{N = 5}$

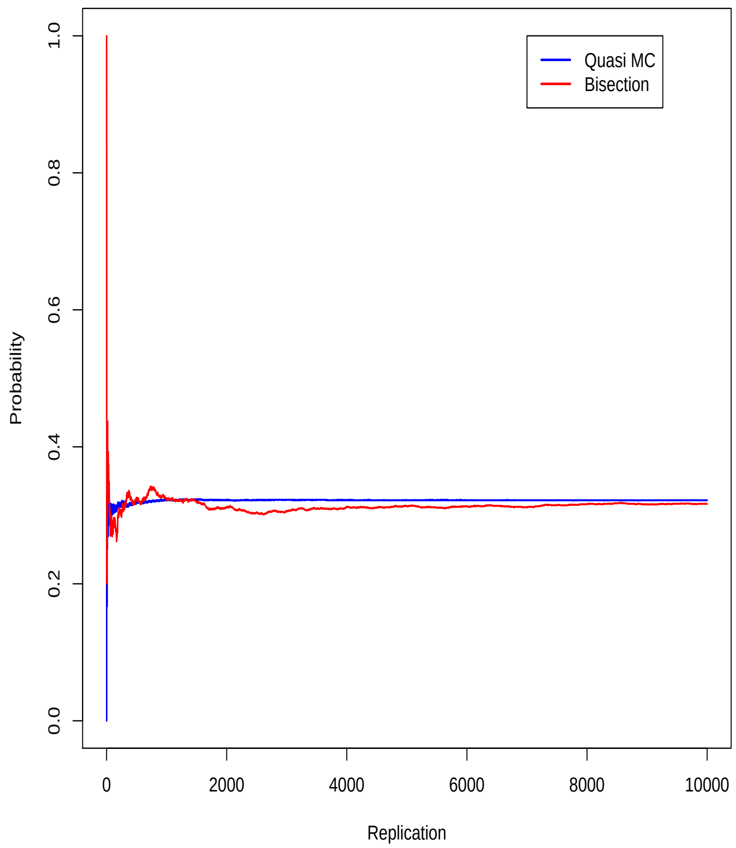

Quasi MC
Bisection
Probability
Replication
0.0
0.2
0.4
0.6
0.8
1.0
0
2000
4000
6000
8000
10000

Figure 3:

Importance sampling

$4N - 1$ choices, probabilities p_i

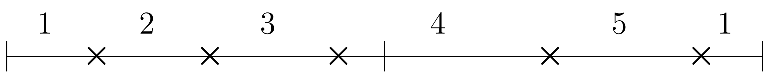
Simulate with $\widetilde{p}_i$

Optimal choice of $\widetilde{p}_i$:

cond. prob given one cycle

Explicit in example

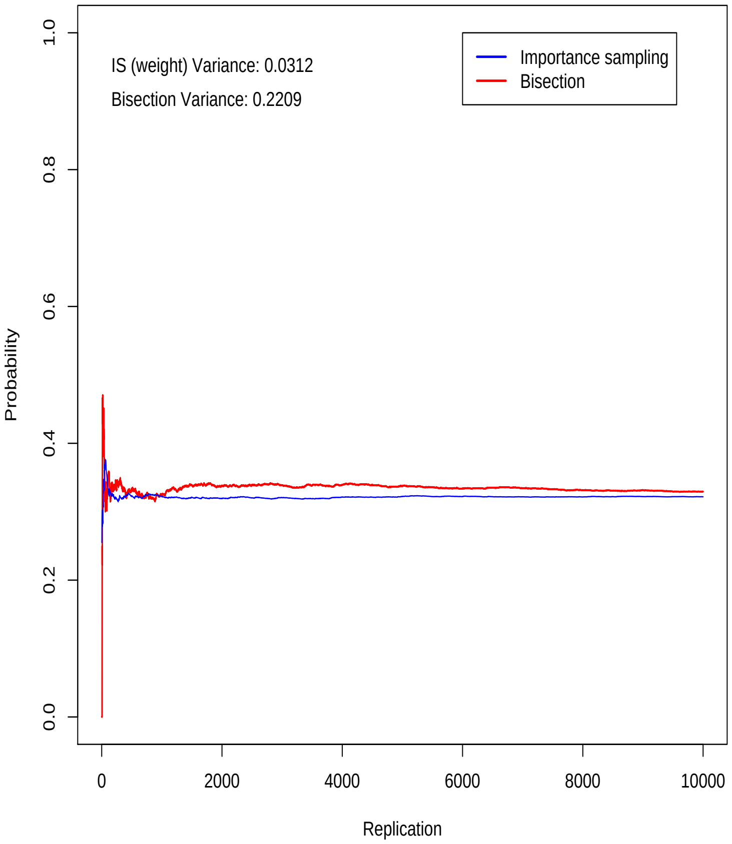
Need N Poisson events in $[0, 1]$



Given N Poisson events in $[0, 1]$:

$M = \#\text{events in } [0, 1/2] \sim \text{bin}(N, 1/2)$

$X(1/2) = M + 1$



IS (weight) Variance: 0.0312
Bisection Variance: 0.2209
Importance sampling
Bisection
Probability
0.0
0.2
0.4
0.6
0.8
1.0
0
2000
4000
6000
8000
10000
Replication

Figure 4:

Summary

- Bridge constructions relevant in many areas. E.g. diffusion bridges, Voss, Makarov, . . . (here), Roberts, Sørensen, . . .
- Implementation for finite CTMP bridges
- Comparisons with other methods
- Potential for var. red.

Still more to do!

Variance reduction
via hybrid algorithms
Only bisection in first steps
e.g. $T/2$, $T/4$, $3T/4$
e.g. rejection sampling
in the 4 intervals

Involve more points than $T/2$