Sampling using Adaptive Regenerative Processes

Hector McKimm

Hector McKimm (Imperial College London), Andi Wang (Bristol), Murray Pollock (Newcastle), Christian Robert (Paris Dauphine) and Gareth Roberts (Warwick)

17 February 2023

Outline

- Motivation
- The Restore process
  - Definition
  - Minimal Restore
  - Issues
- The Adaptive Restore process
  - Definition
  - Examples
- Conclusion

Introduction: Monte Carlo

Want to comptue:

$$\mathbb{E}_\pi[f(X)] = \int f(x)\pi(x)dx.$$

Monte Carlo: $X_1, X_2, \ldots, X_n \sim \pi$;

$$\mathbb{E}_\pi[f(X)] \approx \frac{1}{n}\sum_{i=1}^{n} f(X_i).$$

MCMC: when direct simulation from π is not possible, simulate a Markov chain with invariant distribution π.

The Metropolis-Hastings Algorithm

Given x_i, generate the next state (Metropolis et al., 1953; Hastings, 1970):

1. Propose a new state:

$$Y_i \sim q(y|x_i)$$

2. Accept or reject:

$$X_{i+1} = \begin{cases} Y_i & \text{w.p. } \alpha(x_i, Y_i) \\ x_i & \text{w.p. } 1 - \alpha(x_i, Y_i) \end{cases}$$

$$\alpha(x, y) = 1 \wedge \frac{\pi(y)q(x|y)}{\pi(x)q(y|x)}$$

Motivation 1: Compensating Dynamics

The building blocks of MCMC: **Markov transition kernels**.

A π-invariant transition kernel P satisfies:

$$\pi = \pi P$$

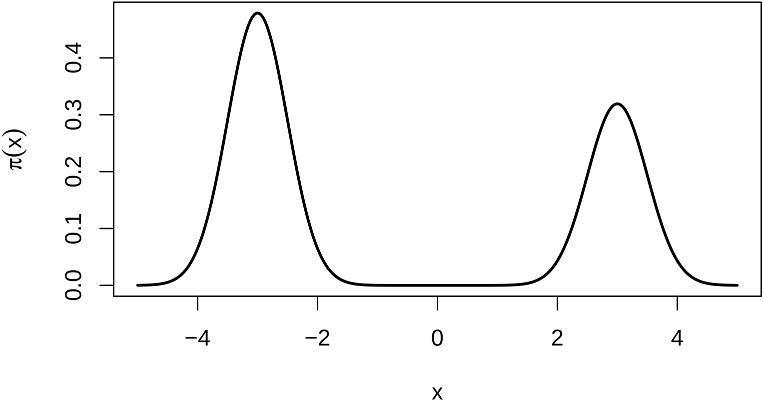
If $P_1, P_2, \ldots, P_n$ are all π-invariant, may be combined:

$$P = P_1 P_2 \cdots P_n$$
$$P = \frac{1}{n}(P_1 + P_2 + \cdots P_n)$$

How may different dynamics, which by themselves are not π-invariant, be combined in such a way that the dynamics **compensate** for each other so that together they are π-invariant?

Motivation 2: Local and Global Dynamics

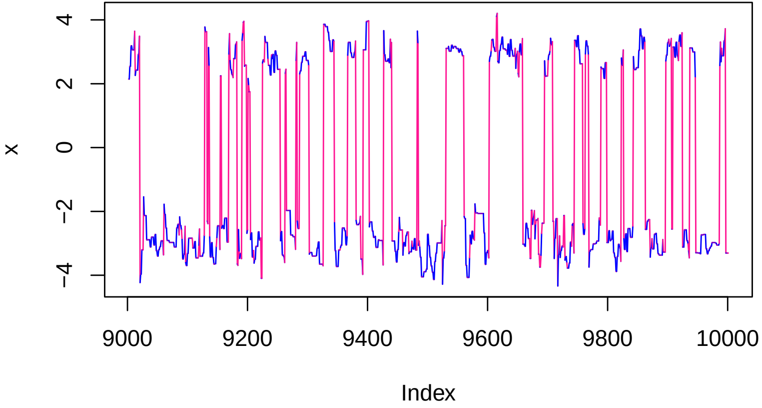
Sampling **multimodal** distributions is difficult!



Motivation 2: Local and Global Dynamics

Proposals

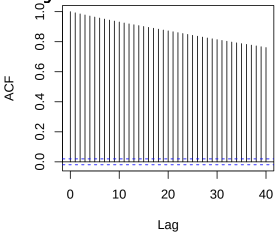
- P_1: $q(y|x) \equiv \mathcal{N}(y; 0, 9)$. Global.
- P_2: $q(y|x) \equiv \mathcal{N}(y; x, 0.5)$. Local.

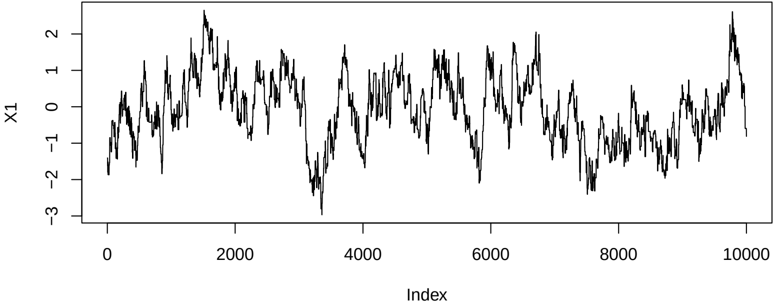


Motivation 2: Local and Global Dynamics

Global moves are desirable for unimodal distributions too!

Example: 100-dimensional Gaussian

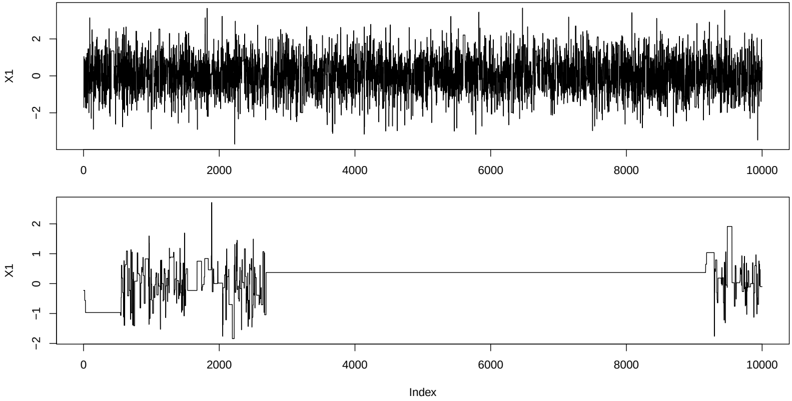




Motivation 2: Local and Global dynamics

Curse of dimensionality: independent moves recede with dimension.

Example: $\pi \equiv \mathcal{N}(0.1\mathbb{1}_d, 0.5I_d + 0.5\mathbb{1}_{d\times d}), q \equiv t_4(0, I)$.



Motivation 3: Regeneration

A **regenerative** stochastic process may be split into i.i.d. cycles, called **tours** (Asmussen, 2003, Chapter 6).

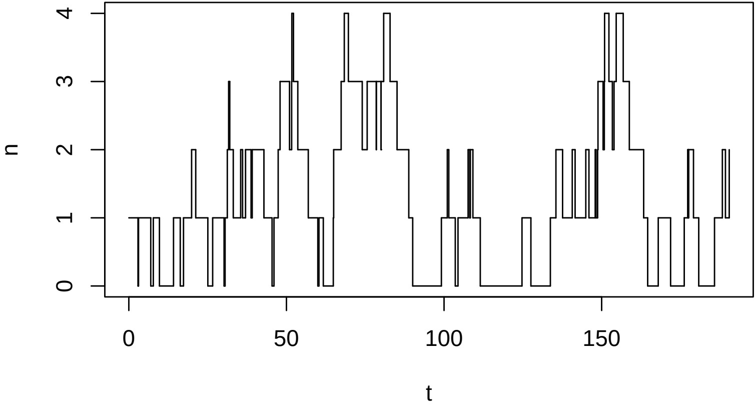


Figure: Example: a queue.

Motivation 3: Regeneration

Nummelin's splitting technique (Nummelin, 1978) may be used to simulate regeneration times in MCMC sampers (Mykland et al., 1995).

Benefits of regenerative simulation:

- Parallel simulation
- Absence of burn-in period
- Get an estimate for the variance of the estimator itself
- Mode jumping

Issue: regenerations receed with dimension.

Motivation 4: Non-reversibility

Reversible Markov chains satisfy the **detailed balance** condition:

$$\pi(dx)P(x, dy) = \pi(dy)P(y, dx), \forall x, y \in \mathbb{R}^d.$$

But there's evidence that non-reversible Markov chains are better (Suwa and Todo, 2010; Turitsyn et al., 2011; Neal, 2004; Chen and Hwang, 2013) in terms of

- asymptotic variance
- speed of convergence

Standard Restore

- $\{Y_t\}_{t\geq 0}$: underlying process (local moves)
- μ: **regeneration distribution** (global moves)
- κ: **regeneration rate**

Restore process $\{X_t\}_{t\geq 0}$ is defined by enriching $\{Y_t\}_{t\geq 0}$ with regenerations from μ at rate κ (Wang et al., 2021).

π-invariance when

$$\kappa = \tilde{\kappa} + C\frac{\mu}{\pi}$$

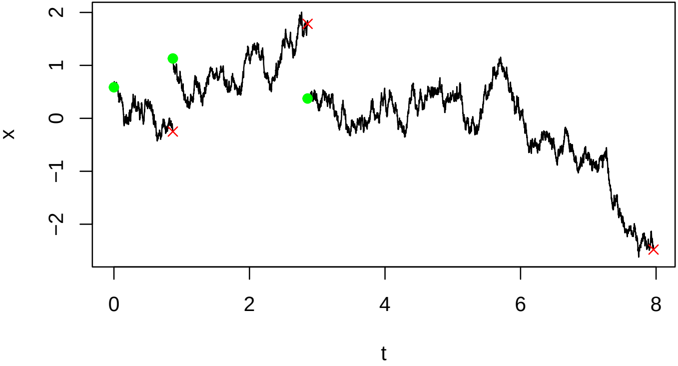
Almost surely:

$$\frac{1}{t}\int_0^t f(X_s)ds \to \mathbb{E}_\pi[f(X)], \quad t \to \infty.$$

Standard Restore: Sample Path

For $\{Y_t\}_{t\geq 0}$ **Brownian Motion** and $U(x) = -\log \pi(x)$:

$$\tilde{\kappa}(x) = \frac{1}{2}\left(||\nabla U(x)||^2 - \Delta U(x)\right).$$



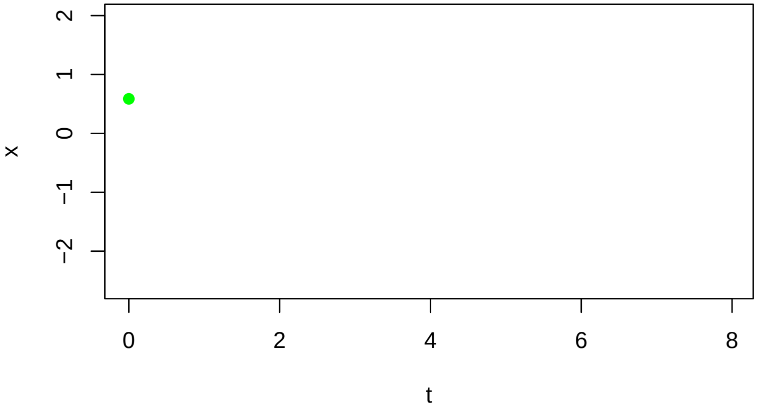
Algorithm Characteristics

- ☑ Compensating Dynamics
- ☑ Local and Global Dynamics
- ☑ Regenerative
- ☑ Non-reversible

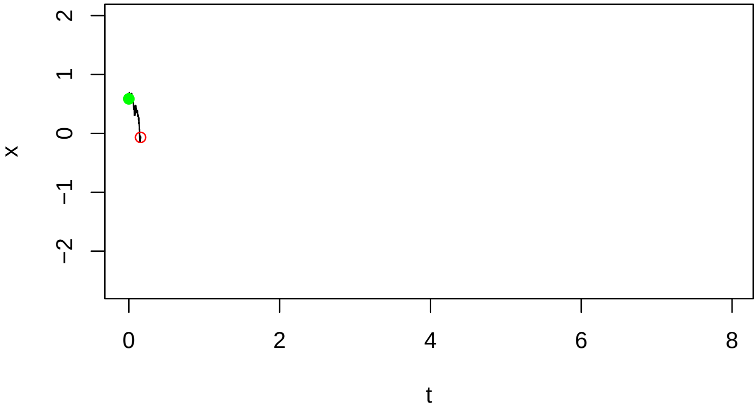
Poisson Thinning

- The regeneration rate itself is a stochastic process
- There's no closed form expression for the regeneration times
- Suppose $\kappa < K$ uniformly, then can use **Poisson Thinning** to simulate regeneration times

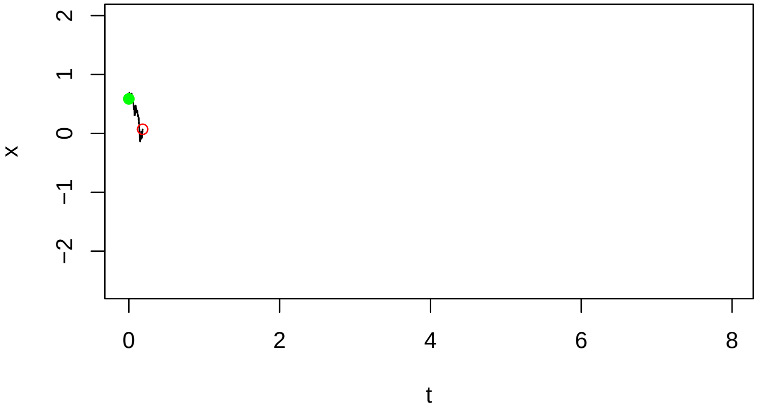
Standard Restore: Simulation via Poisson Thinning



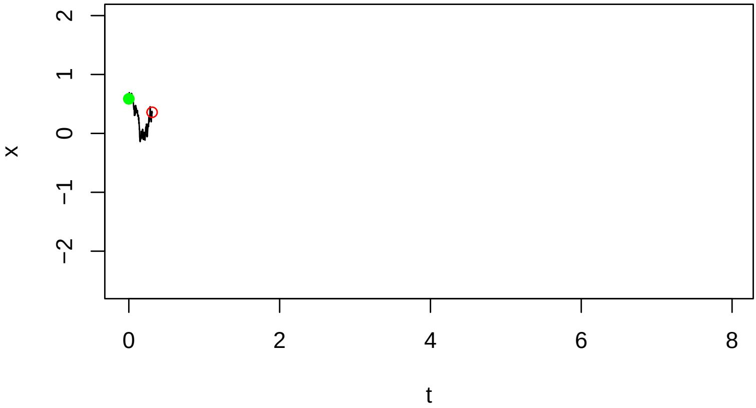
Standard Restore: Simulation via Poisson Thinning



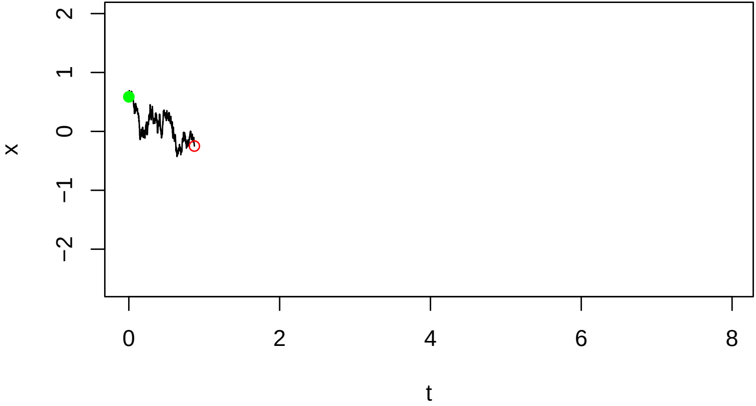
Standard Restore: Simulation via Poisson Thinning



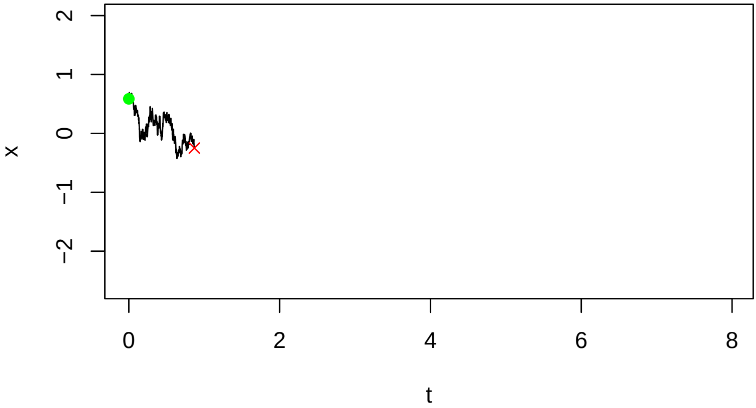
Standard Restore: Simulation via Poisson Thinning



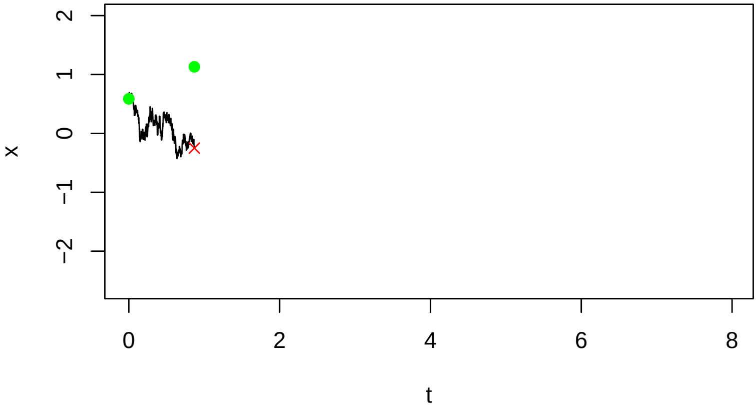
Standard Restore: Simulation via Poisson Thinning



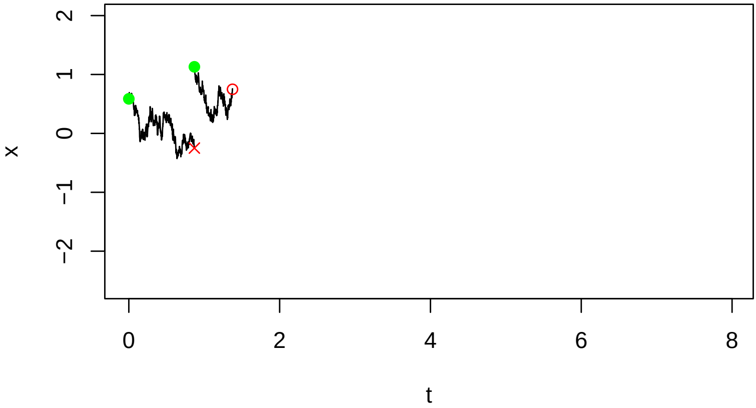
Standard Restore: Simulation via Poisson Thinning



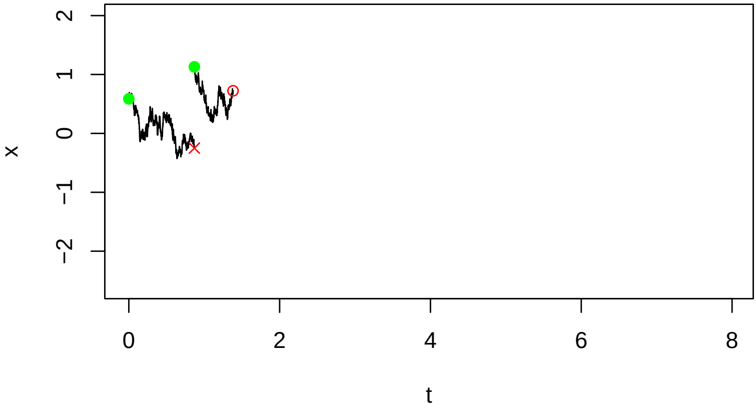
Standard Restore: Simulation via Poisson Thinning



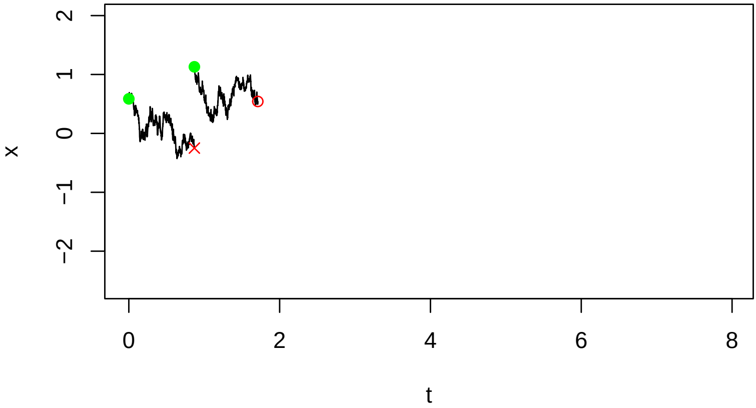
Standard Restore: Simulation via Poisson Thinning



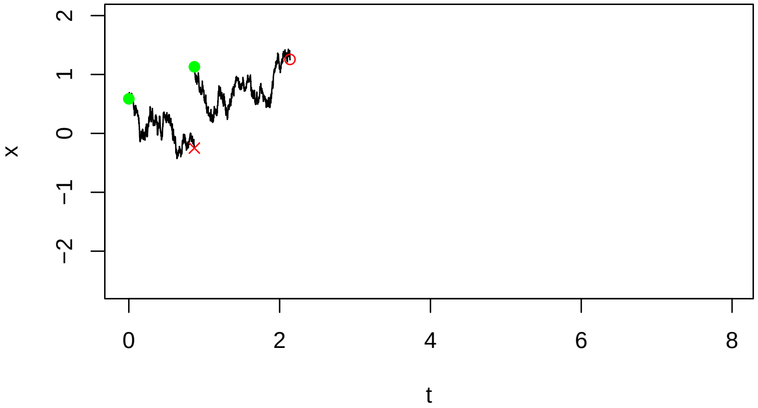
Standard Restore: Simulation via Poisson Thinning



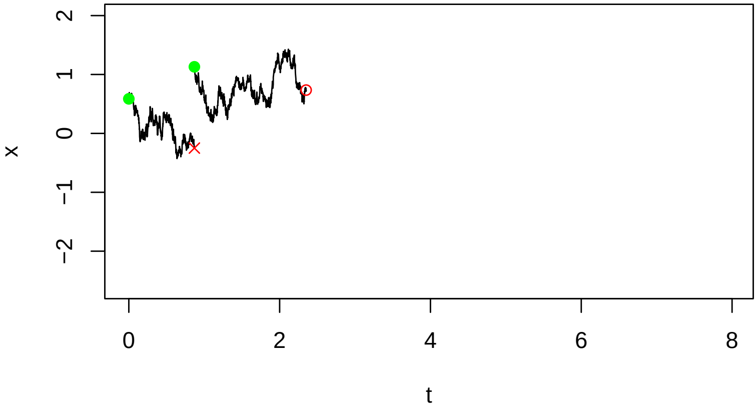
Standard Restore: Simulation via Poisson Thinning



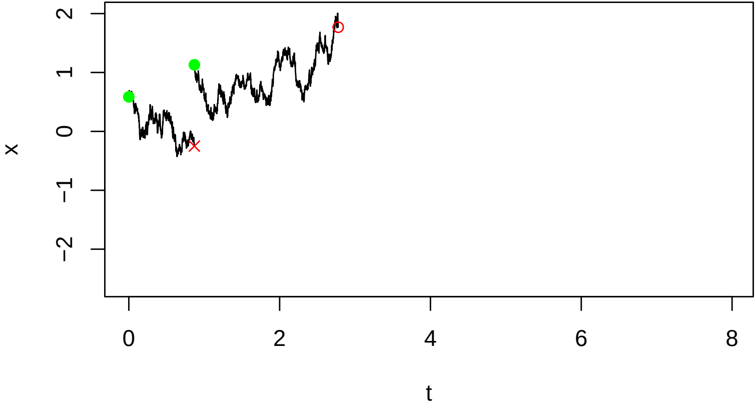
Standard Restore: Simulation via Poisson Thinning



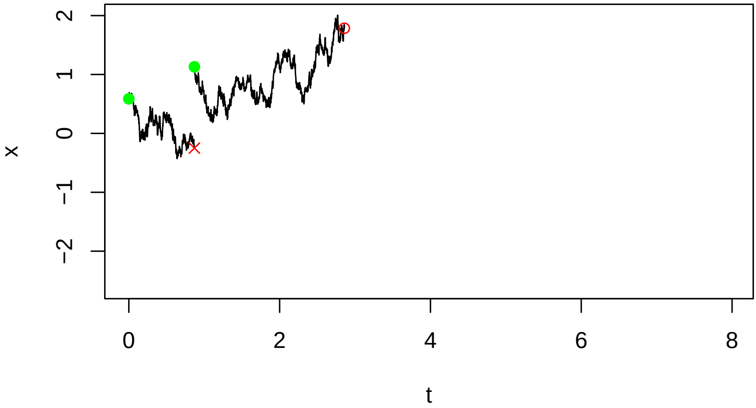
Standard Restore: Simulation via Poisson Thinning



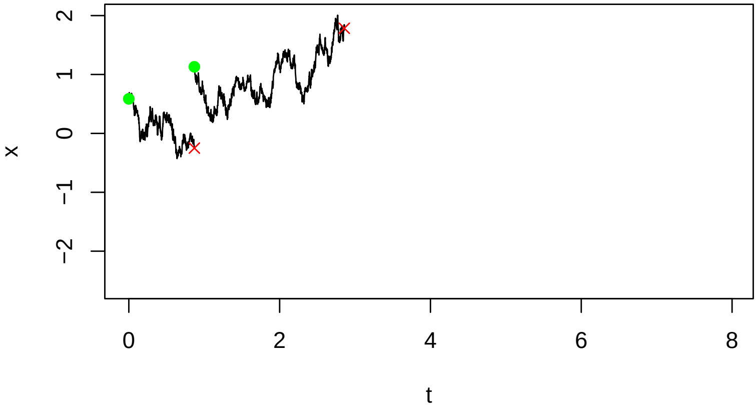
Standard Restore: Simulation via Poisson Thinning



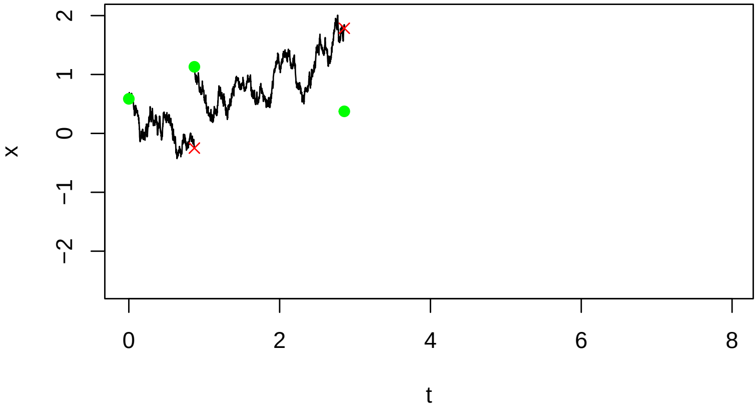
Standard Restore: Simulation via Poisson Thinning



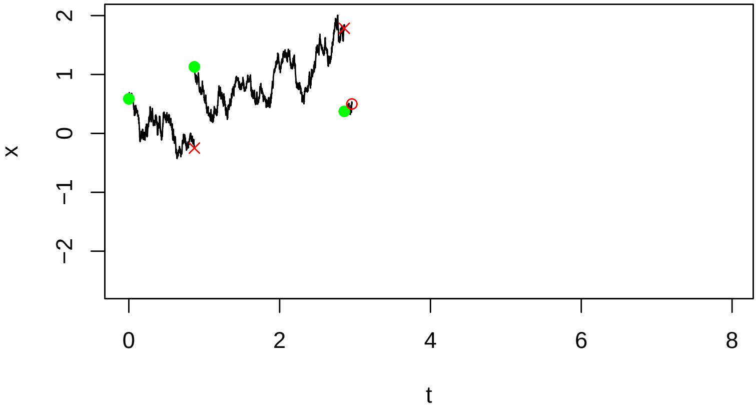
Standard Restore: Simulation via Poisson Thinning



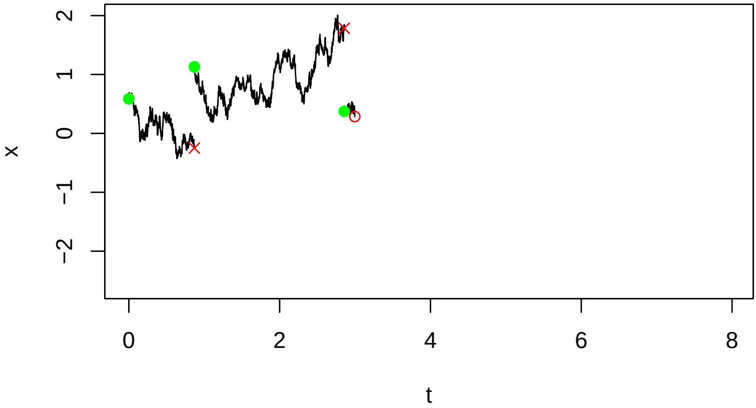
Standard Restore: Simulation via Poisson Thinning



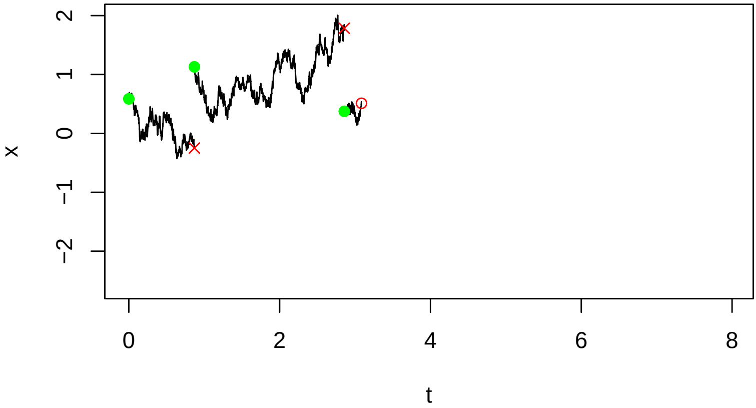
Standard Restore: Simulation via Poisson Thinning



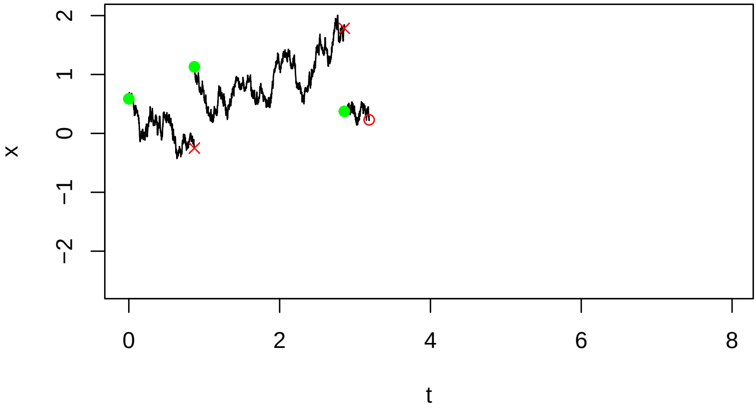
Standard Restore: Simulation via Poisson Thinning



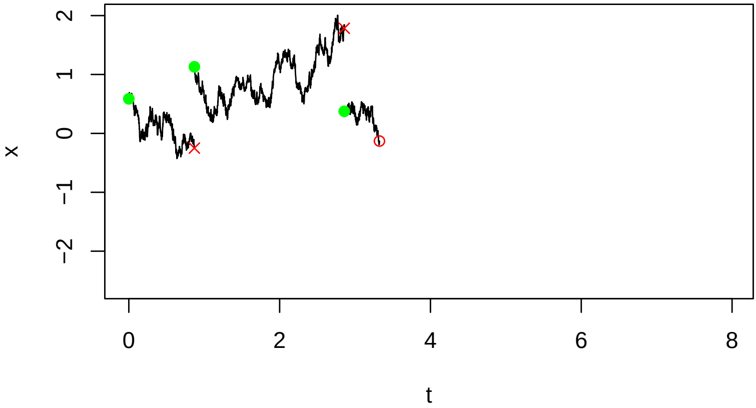
Standard Restore: Simulation via Poisson Thinning



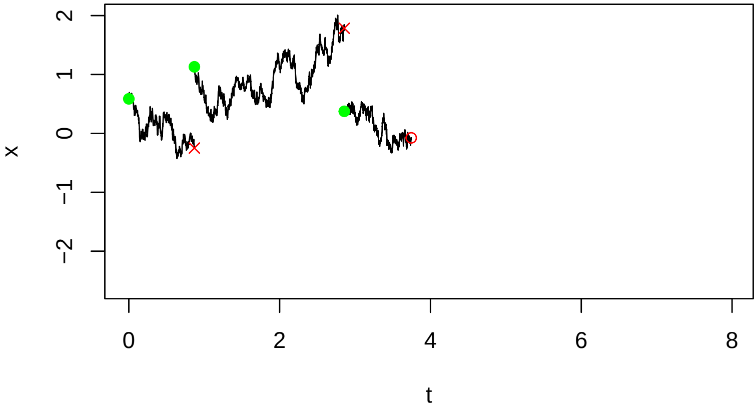
Standard Restore: Simulation via Poisson Thinning



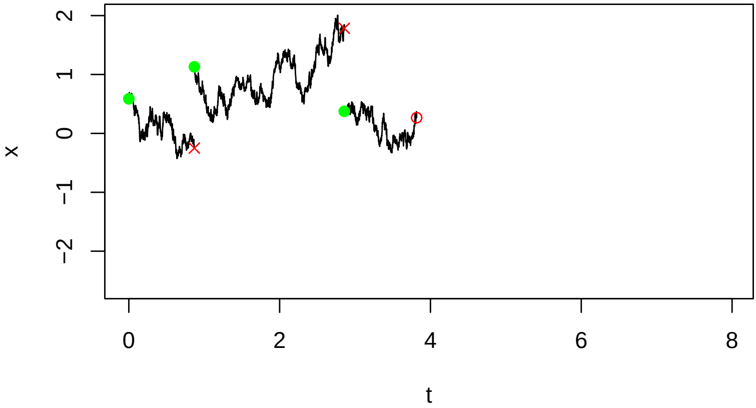
Standard Restore: Simulation via Poisson Thinning



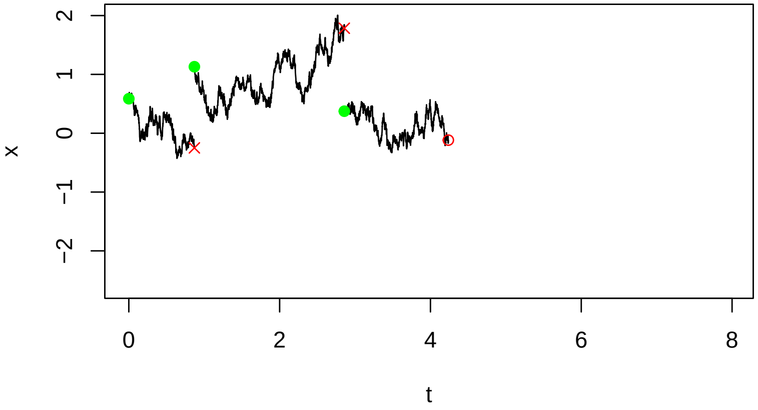
Standard Restore: Simulation via Poisson Thinning



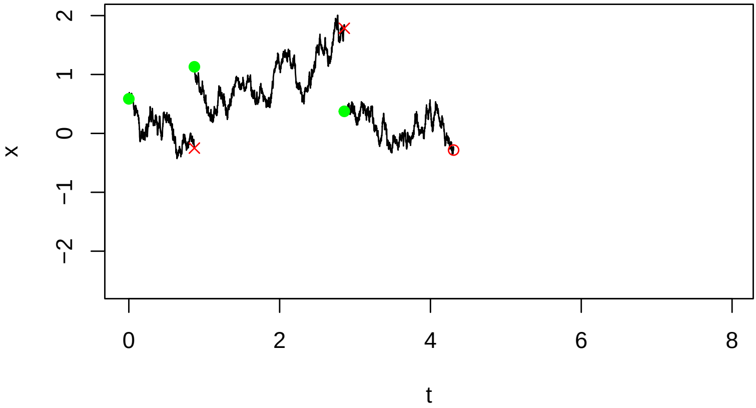
Standard Restore: Simulation via Poisson Thinning



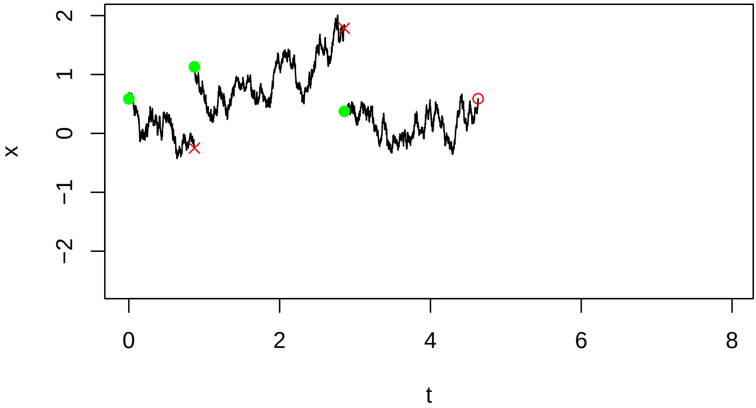
Standard Restore: Simulation via Poisson Thinning



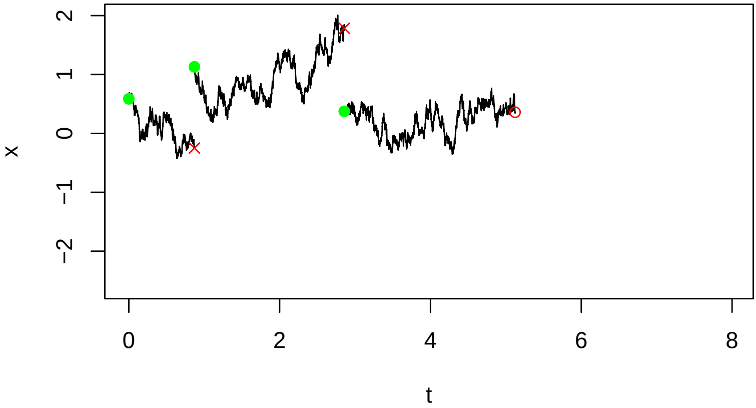
Standard Restore: Simulation via Poisson Thinning



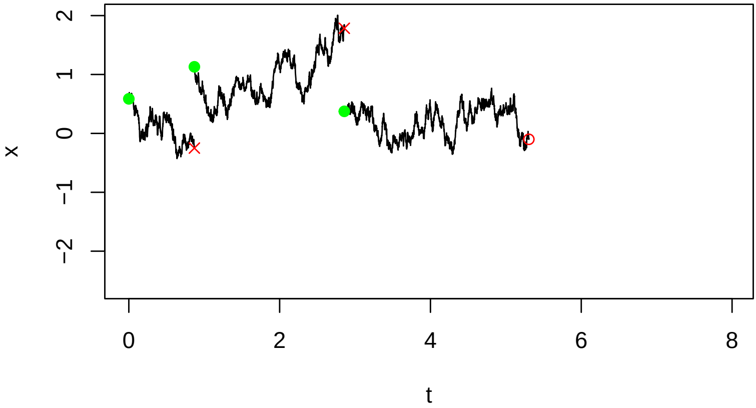
Standard Restore: Simulation via Poisson Thinning



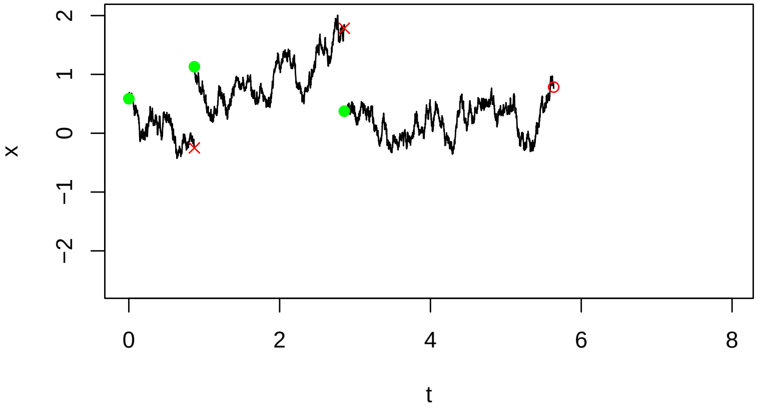
Standard Restore: Simulation via Poisson Thinning



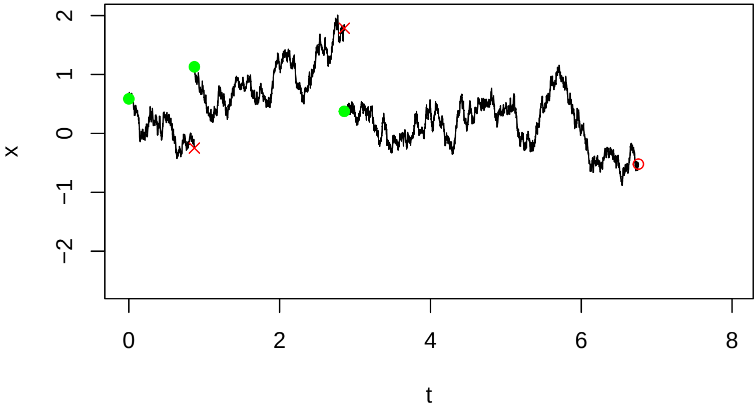
Standard Restore: Simulation via Poisson Thinning



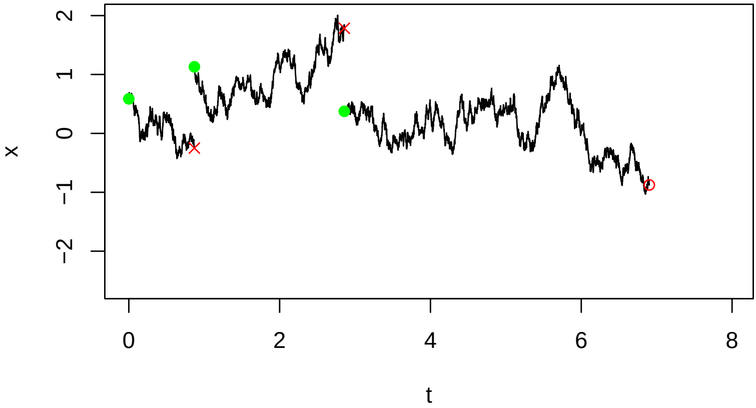
Standard Restore: Simulation via Poisson Thinning



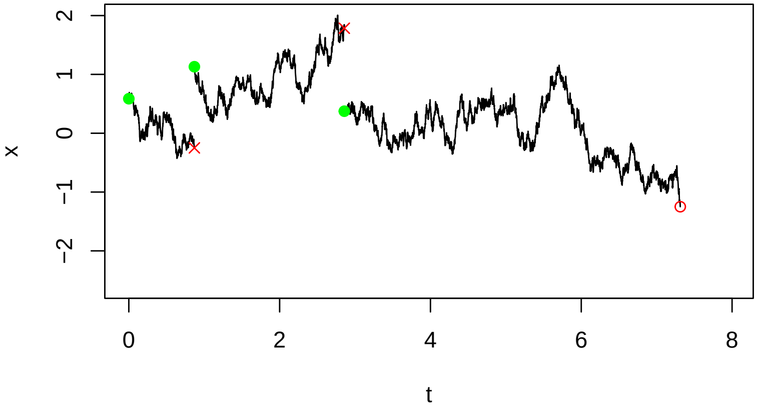
Standard Restore: Simulation via Poisson Thinning



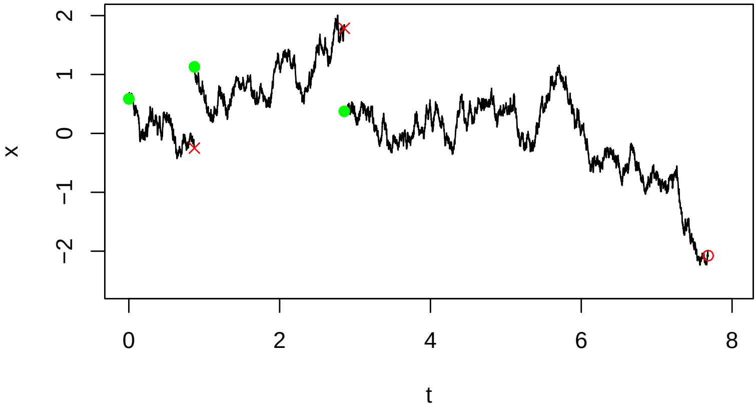
Standard Restore: Simulation via Poisson Thinning



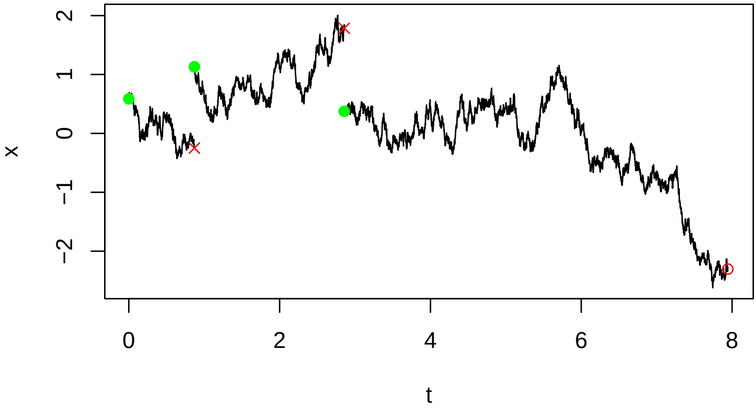
Standard Restore: Simulation via Poisson Thinning



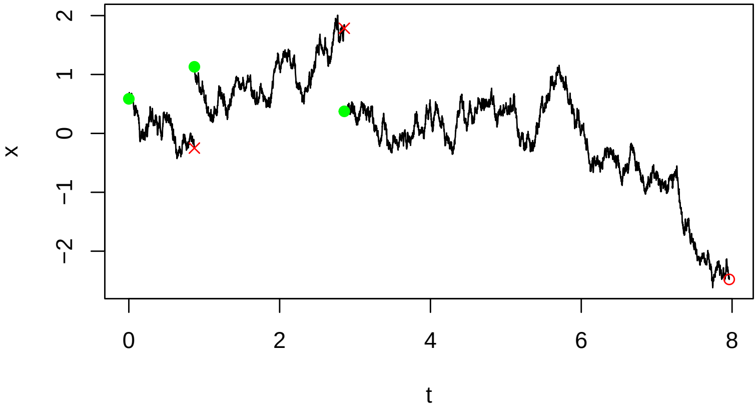
Standard Restore: Simulation via Poisson Thinning



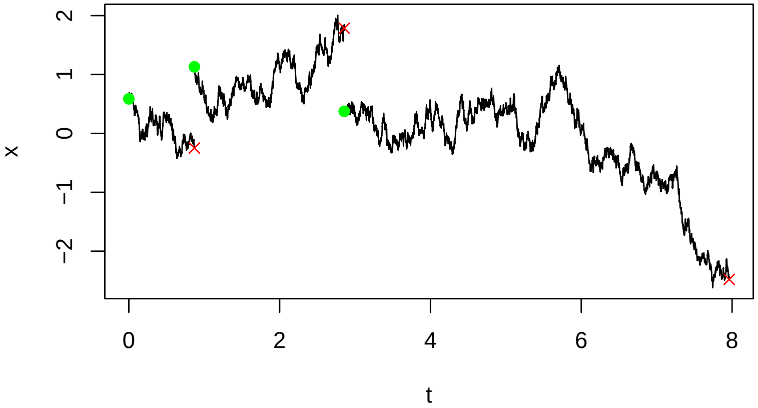
Standard Restore: Simulation via Poisson Thinning



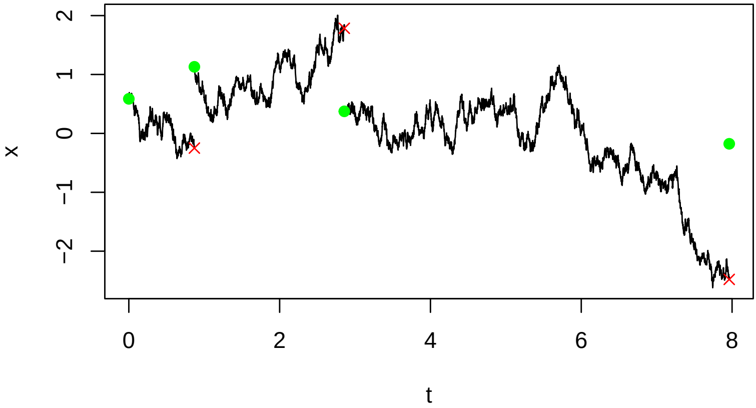
Standard Restore: Simulation via Poisson Thinning



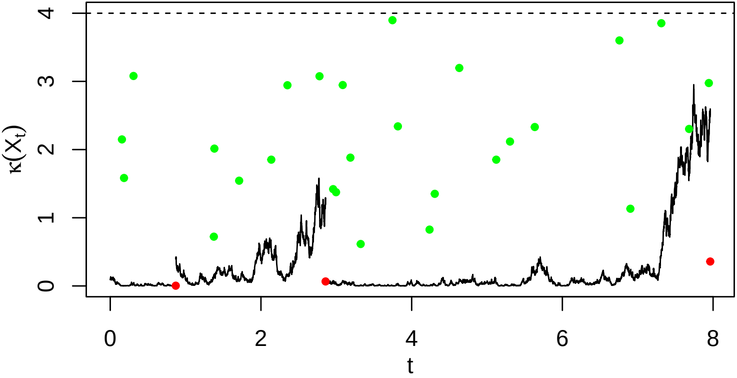
Standard Restore: Simulation via Poisson Thinning



Standard Restore: Simulation via Poisson Thinning



Standard Restore: Simulation via Poisson Thinning



Minimal Restore

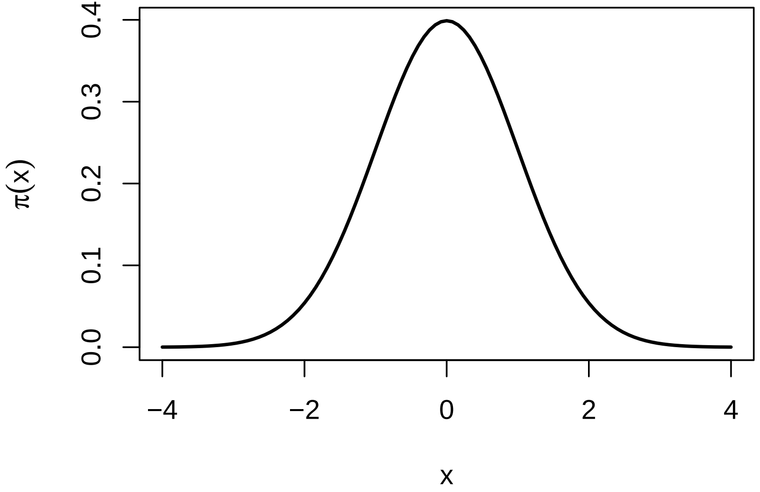
- C^+: minimal regeneration constant
- μ^+ : minimal regeneration distribution
- κ^+ : minimal regeneration rate

$$\kappa^+ := \tilde{\kappa} \vee 0 = \tilde{\kappa} + C^+ \frac{\mu^+}{\pi}.$$

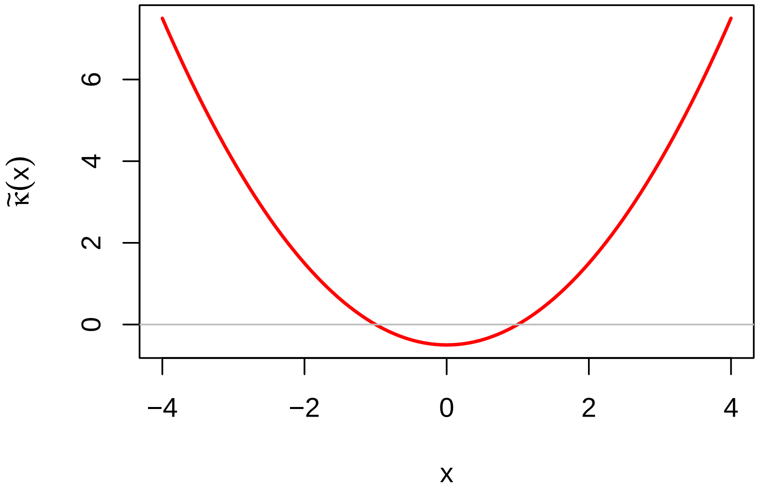
Rearranging:

$$\mu^+ = \frac{1}{C^+}[0 \vee -\tilde{\kappa}]\pi$$

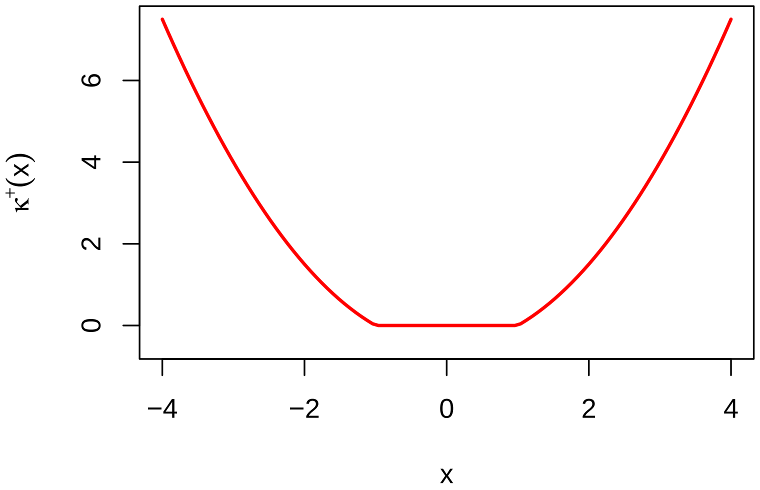
Example: Standard Normal Distribution



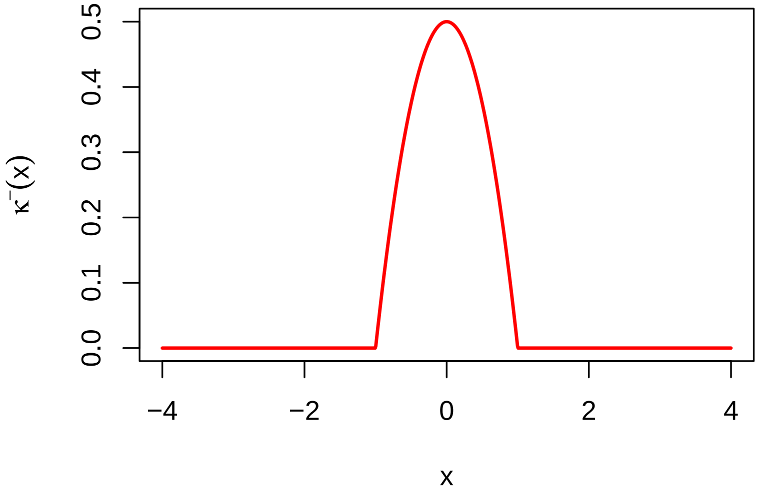
Partial Regeneration Rate



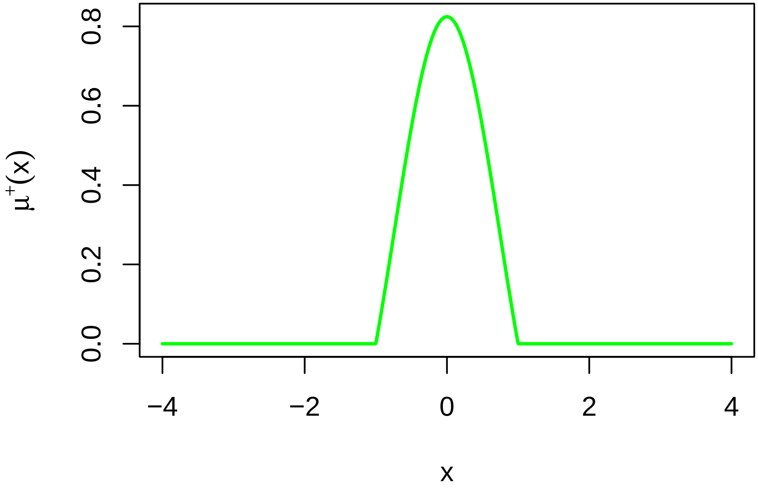
Minimal Regeneration Rate



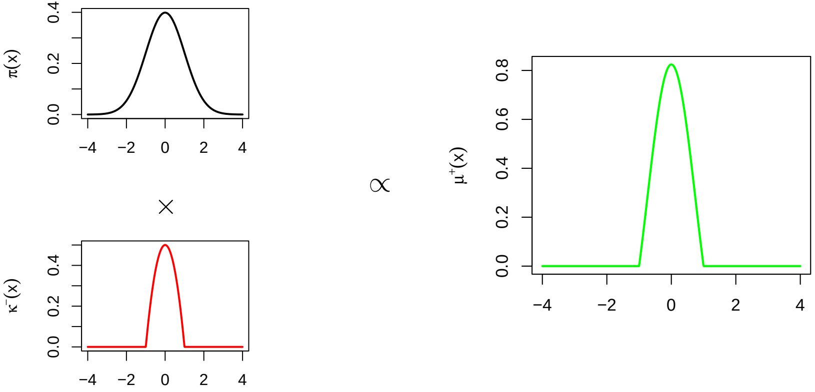
Minimal Regeneration Rate



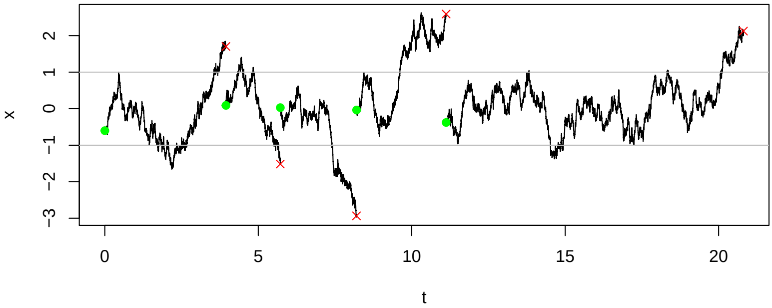
Minimal Regeneration Distribution



Relationship between π, κ^{-} and μ^{+}



Minimal Restore: Sample Path



Very Large Regeneration Rate

- When μ is a bad approximation of π, κ can become very large!
- Example: Logistic Regression model of breast cancer
- Transform π based on its Laplace approximation then choose μ as the standard Gaussian distribution
- $\mathbb{P}(\kappa < 9465) \approx 0.999$.

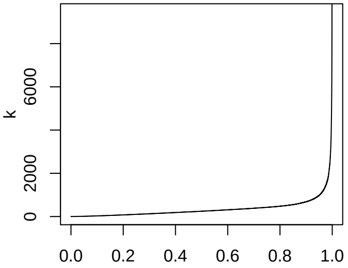


Figure: $\mathbb{P}(\kappa(X) < k), \quad X \sim \pi$

Very large regeneration rate

- When κ is very large, it is due to the ratio μ/π
- $\kappa = \tilde{\kappa} + C\mu/\pi$
- $\mathbb{P}(\tilde{\kappa} < 19.64) \approx 0.999$

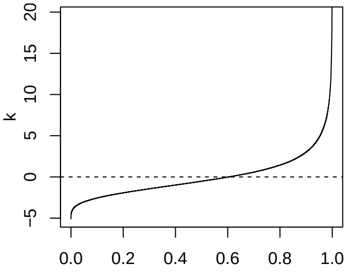


Figure: $\mathbb{P}(\tilde{\kappa}(X) < k), \quad X \sim \pi$

Adaptive Restore

(McKimm et al., 2022)

- Enrich $\{Y_t\}_{t\geq 0}$ with regenerations at rate κ^+ from, at time t, a distribution μ_t
- μ_0: initial regeneration distribution
- The regeneration distribution is updated by adding point masses to it
- π_t: the invariant distribution of the Restore process with fixed regeneration distribution μ_t
- $(\mu_t, \pi_t) \to (\mu^+, \pi)$

Adaptive Restore

$$\mu_t(x) = \begin{cases} \mu_0(x), & N(t) = 0, \\ \frac{t}{a+t} \frac{1}{N(t)} \sum_{i=1}^{N(t)} \delta_{X_{\zeta_i}}(x) + \frac{a}{a+t} \mu_0(x), & N(t) > 0 \end{cases}$$

$\zeta_1, \zeta_2, \ldots, \zeta_{N(t)}$ the arrival times of an inhomogeneous Poisson process $(N(t) : t \geq 0)$ with rate

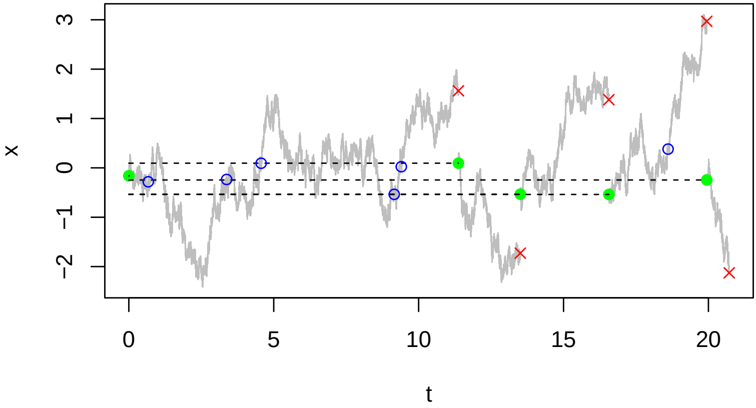
$$t \mapsto \kappa^-(X_t),$$

$$\kappa^-(x) := [0 \vee -\tilde{\kappa}(x)].$$

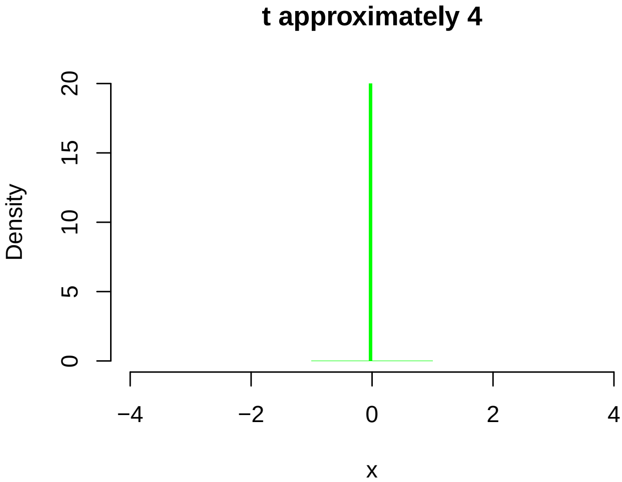
It's assumed $\kappa^- < K^-$ for $K^- > 0$, so the Poisson process may be simulated using Poisson thinning.

Example: Standard Normal

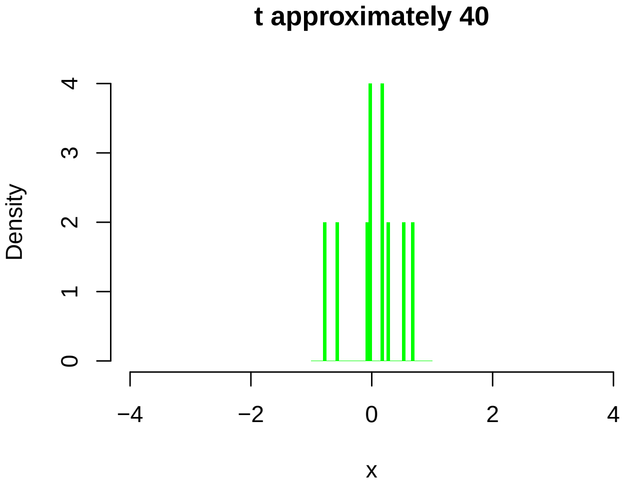
Example: $a = 1, \pi \equiv \mathcal{N}(0, 1), \mu_0 \equiv \mathcal{N}(0, 0.5)$.



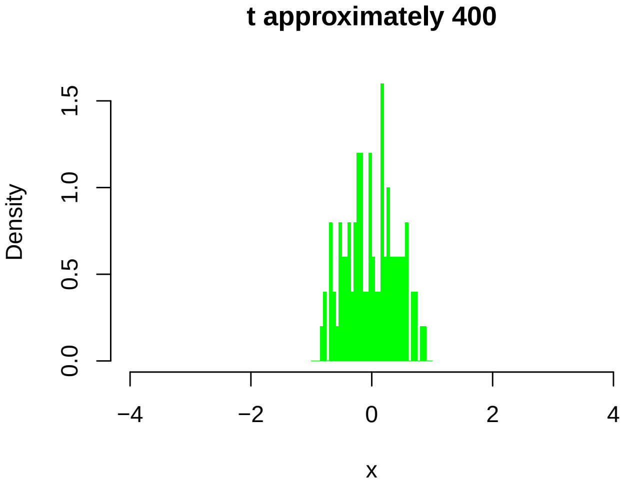
Discrete Component of the Regeneration Distribution



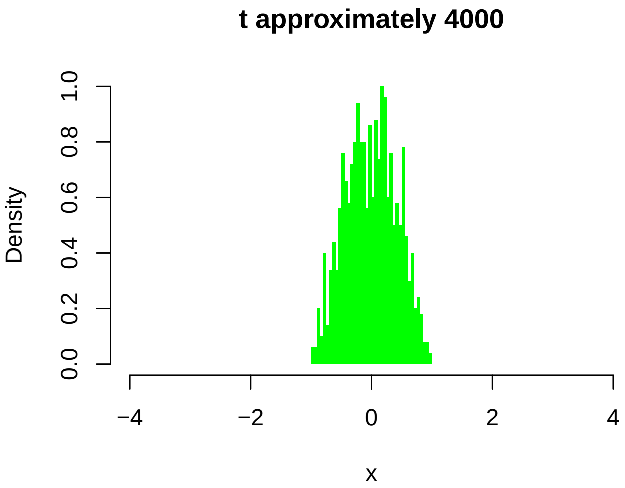
Discrete Component of the Regeneration Distribution



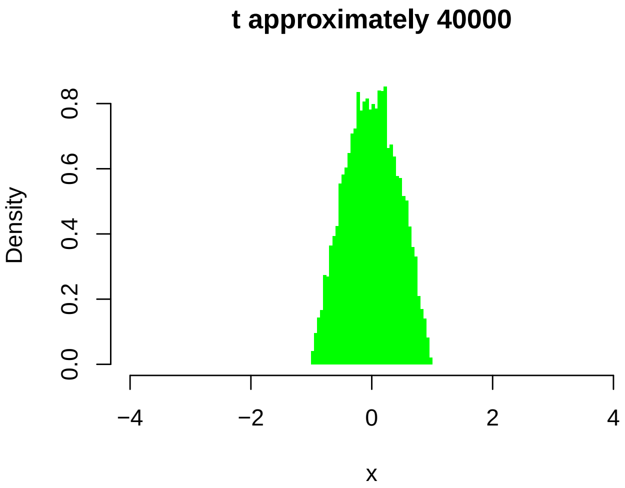
Discrete Component of the Regeneration Distribution



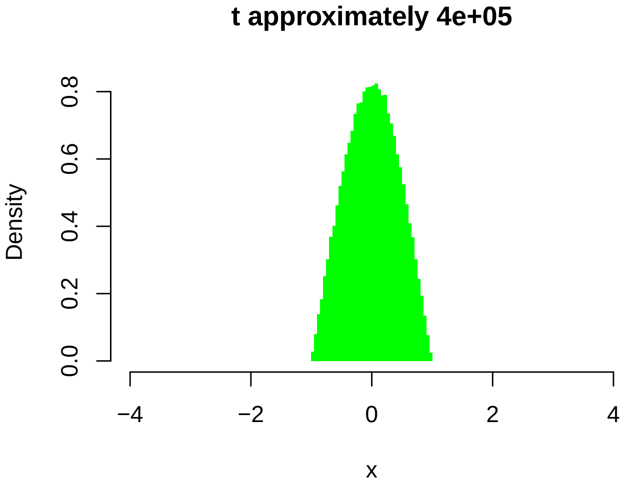
Discrete Component of the Regeneration Distribution



Discrete Component of the Regeneration Distribution

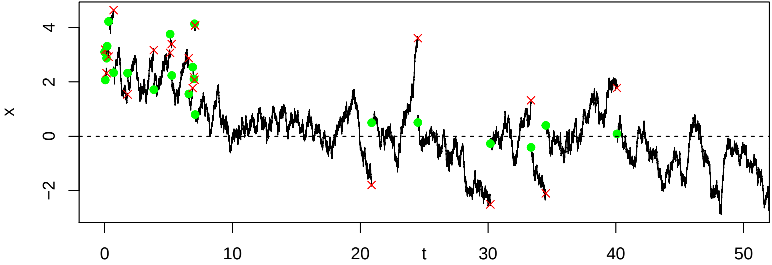


Discrete Component of the Regeneration Distribution



Adaptive Restore: Sample Path

- $\pi = \mathcal{N}(0, 1)$
- $a = 10$
- $\mu_0 = \mathcal{N}(2, 1)$



Algorithm Characteristics

- ☑ Compensating Dynamics
- ☑ Local and Global Dynamics
- ☒ Regenerative
- ☑ Non-reversible

A lot more practical!

Examples

- Logistic Regression Model ($d = 10$)
- Hierarchical Model of Pump Failure ($d = 11$)
- Log-Gaussian Cox Point Proces Model ($d = 25$)
- Multivariate t-distribution ($d = 2$)
- Mixture of Gaussians ($d = 2$)

Conclusion

- Adaptive Restore represents a significant improvement on Standard Restore by making simulation tractable for a wider range of target distributions
- Improvement is greatest for target distributions with skewed tails
- In comparison to simpler algorithms such as Random Walk Metropolis, the process can still be slow to simulate
- Convergence appears to be slow when the target is multimodal
- Novel application of the stochastic approximation technique to establishing convergence of self-reinforcing processes

References I

Asmussen, S. (2003). *Applied Probability and Queues*, volume 51. Springer-Verlag New York.

Chen, T.-L. and Hwang, C.-R. (2013). Accelerating reversible Markov chains. *Statistics & Probability Letters*, 83(9):1956–1962.

Hastings, W. K. (1970). Monte Carlo Sampling Methods Using Markov Chains and Their Applications. *Biometrica*, 57:97–109.

McKimm, H., Wang, A., Pollock, M., Robert, C. P., and Roberts, G. O. (2022). Sampling using Adaptive Regenerative Processes.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). Equation of State Calculations by Fast Computing Machines. *The journal of chemical physics*, 21(6):1087–1092.

Mykland, P., Tierney, L., and Yu, B. (1995). Regeneration in Markov Chain Samplers. *Journal of the American Statistical Association*, 90(429):233–241.

References II

Neal, R. M. (2004). Improving Asymptotic Variance of MCMC Estimators: Non-reversible Chains are Better.

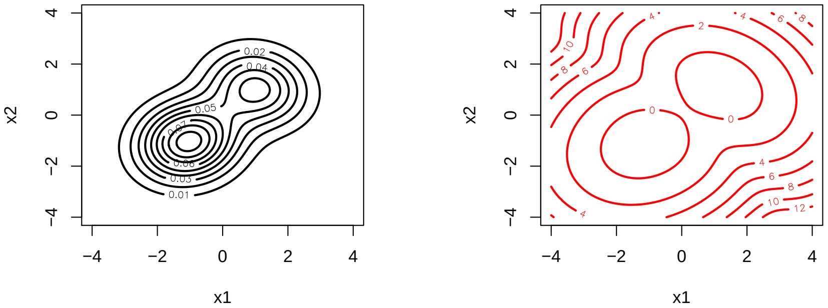
Nummelin, E. (1978). A splitting technique for Harris recurrent Markov chains. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 43(4):309–318.

Suwa, H. and Todo, S. (2010). Markov Chain Monte Carlo Method without Detailed Balance. *Phys. Rev. Lett.*, 105:120603.

Turitsyn, K. S., Chertkov, M., and Vucelja, M. (2011). Irreversible Monte Carlo algorithms for efficient sampling. *Physica D: Nonlinear Phenomena*, 240(4):410–414.

Wang, A. Q., Pollock, M., Roberts, G. O., and Steinsaltz, D. (2021). Regeneration-enriched Markov processes with application to Monte Carlo. *The Annals of Applied Probability*, 31(2):703–735.

Example: Gaussian mixture



Slow convergence due to **urn-like behaviour**: although the chain is guaranteed to converge asymptotically, in finite time the chain is naturally inclined to visit regions it has visited before.