4.3 Markov chain Monte Carlo

→ Simulation method based on Markov chains

4.3.1 Simulation of rand. variables

Simulation means to imitate the outcomes of a r.v. X

One way is e.g. to use random numbers of the uniform distribution on $[0,1]$ (i.e. $U(0,1)$)

The random numbers of the unif. distr. can be e.g. imitated by means of the congruential random number generator, which generates "rand. numbers" $R_0, R_1, R_2, \dots$ according to the formula

$$R_n = \frac{Y_n}{m}, \qquad (4.3.1.1)$$

where

$$Y_{n+1} = (aY_n + b) \bmod m, \quad Y_0 \in \{0,1,\dots,m-1\}$$

for "nice" $a, b, m \in \mathbb{N}_0$

(e.g. $a = 31415821, b = 1, m = 108$)

! $20 \bmod 4 = 0$, $21 \bmod 4 = 1$, $22 \bmod 4 = 2$, ...

Consider e.g. discrete r.v. $X \in \mathcal{I} = \{x_1, x_2, x_3, \dots\}$

set $p_j = P(X = x_j)$

Def. $F_0 = 0$, $F_K = \sum_{j=1}^{K} p_j$

Then one shows that X has the same distr. as the r.v.

$$X^* = x_K \text{ if } F_{K-1} < U \leq F_K \qquad (4.3.1.2)$$

where $U \sim U(0,1)$ ($F_{K,K\geq 1}$ partition of $[0,1]$!)

→ simulation procedure:

Choose successively $U = R_0, U = R_1, U = R_2, \dots$ in (4.3.1.1)

(4.3.1.2) → the output $X^* = X^*(R_0), X^* = X^*(R_1), X^* = X^*(R_2), \dots$ gives a simulation of X

4.3.2 Simulation of Markov chains

→ procedure to imitate the outcomes of the r.v.'s $X_0, X_1, X_2, \dots$ of a Markov chain with initial distr $\lambda = (\lambda_i)_{i\in\mathcal{I}}$ and trans. matrix $P = (p_{ij})_{i,j\in\mathcal{I}}$ which can be described as follows:

(84) Denote by $A_i, i \in \mathbb{J}$ disjoint subintervals of $[0,1]$ with lengths

$$|A_i| = \lambda_i$$

s.t.

$$\bigcup_{i \in \mathbb{J}} A_i = [0,1]$$

Similarly, let for each $i \in \mathbb{J}$ $A_{ij}, j \in \mathbb{J}$ be a partition of $[0,1]$ s.t.

$$|A_{ij}| = p_{ij}$$

Define the functions

$$G_0 : [0,1] \to \mathbb{J} \text{ and } G : \mathbb{J} \times [0,1] \to \mathbb{J} \quad (4.3.2.1)$$

by $G_0(u) = i$ if $u \in A_i$

and $G(i,u) = j$ if $u \in A_{ij}$

Consider $U_i, i \geq 0$ i.i.d r.v.'s with common law $U \sim \mathcal{U}(0,1)$

Define

$$\begin{cases} X_0 = G_0(U_0) \\ X_{n+1} = G(X_n, U_{n+1}), \ n \geq 0 \end{cases} \quad (4.3.2.2)$$

$$\Rightarrow P(X_0 = i) = P(U_0 \in A_i) = P(U \in A_i) = \lambda_i$$

and $P(X_{n+1} = i_{n+1} \mid X_0 = i_0, \dots, X_n = i_n) =$

$P(G(i_n, U_{n+1}) = i_{n+1} \mid X_0 = i_0, \dots, X_n = i_n)$

$\overset{\text{indep.}}{=} P(G(i_n, U_{n+1}) = i_{n+1}) = P(U_{n+1} \in A_{i_n i_{n+1}})$

$= P(U \in A_{i_n i_{n+1}}) = p_{i_n i_{n+1}}$ ($\overset{\text{similarly}}{=} P(X_{n+1} = i_{n+1} \mid X_n = i_n)$).

$\Rightarrow X_n, n \geq 0$ Markov (λ, P)

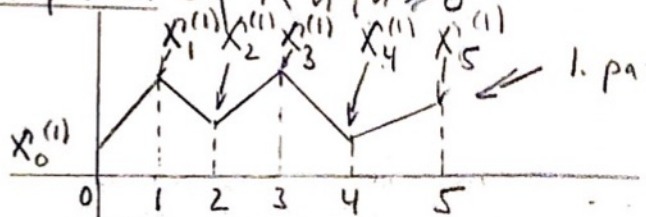
($\longrightarrow$ gives a proof of the existence of Markov chains!)

$\longrightarrow$ simulation procedure:

1. Draw successively numbers $U_0^{(1)}, U_1^{(1)}, U_2^{(1)} \dots$ from the rand. number generator for $U \sim \mathcal{U}(0,1)$ (see 4.3.1)

(4.3.2.2) $\longrightarrow$ output $X_0^{(1)} = G_0(U_0^{(1)}), X_1^{(1)} = G(X_0^{(1)}, U_1^{(1)}), X_2^{(1)} = G(X_1^{(1)}, U_2^{(1)}), \dots$

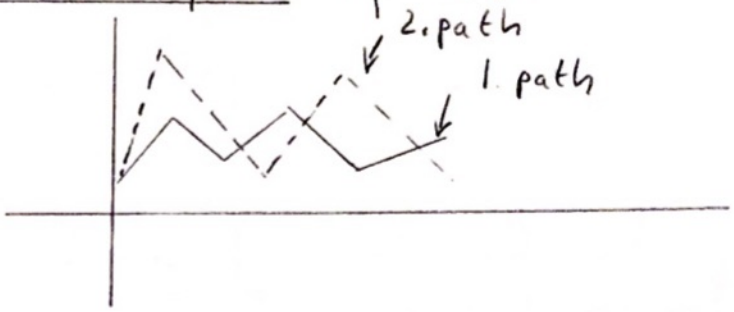
gives the 1. simulation path of $X_n, n \geq 0$



2. Draw successively numbers $U_0^{(2)}, U_1^{(2)}, U_2^{(2)}, \dots$ from the rand. numb. gen. for $U \sim \mathcal{U}(0,1)$

(4.3.2.2) $\longrightarrow$ output $X_0^{(2)} = G_0(U_0^{(2)}), X_1^{(2)} = G(X_0^{(2)}, U_1^{(2)}), \dots$

gives the 2. simulation path of $X_{n,n\geq 0}$



⋮

n-th simulation path

⋮

4.3.3 Hastings-Metropolis algorithm

Suppose we are interested to calculate

$$\theta = E[h(X)] = \sum_{j\geq 1} h(x_j)\cdot P(X=x_j), \quad (4.3.3.1)$$

where

$X \in \mathcal{I} = \{x_1, x_2, \dots\}$ is a discrete raud. vector and

h function on $\mathcal{I}$

It is sometimes computationally difficult to evaluate θ in (4.3.3.1)

→ possible solution: Monte Carlo simulation, that is consider i.i.d r.v.'s $X_i, i\geq 1$ with common law X

$$\xrightarrow{\text{strong law of large numbers}} \frac{1}{n}\sum_{i=1}^{n} h(X_i) \xrightarrow[n\to\infty]{} E[h(X)] = \theta$$

→ Monte Carlo simulation procedure:
Choose N sufficiently large. Then draw successively numbers $Y_1, Y_2, \dots, Y_N$ from the raud. numb. gen. for X (see (4.3.1.2))

$$\rightarrow \quad \frac{1}{N}\sum_{i=1} h(Y_i) \approx \theta$$

→ disadvantage of this approach: In general it is not easy to simulate raud. vectors X (e.g. if X is a vector of dependent r.v.'s). On the other hand the distr. of X is sometimes given by

$$P(X=x_j) = C\cdot b_j, \; j\geq 1, \quad (4.3.3.2)$$

where b_j is specified and C is not computationally feasible

Here "not computationally feasible" means the following:

If $P(X=x_j) = C \cdot b_j$ then

$$1 = \sum_{j \in J} P(X=x_j) = C \cdot \sum_{j \in J} b_j$$

So $C = 1 / \sum_{j \in J} b_j$

However, if $|J|$ = number of elements in J is large (e.g. 10^{30}) then the computation of $\sum_{j \in J} b_j$ is in general not feasible

→ possible solution to (4.3.3.2):

Hastings-Metropolis algorithm

→ Idea: Construction of a (reversible) Markov chain $X_n, n \geq 0$ with invariant distr.

$$\pi_j = P(X=x_j) = C \cdot b_j, \; j \geq 1 \qquad (4.3.3.3)$$

and application of the ergodic theorem

Let $Y_n, n \geq 0$ be a Markov chain with transition prob. $q_{ij}, i,j \geq 1$

Further define the probabilities

$$\alpha_{ij} = \min\left(\frac{\pi_j \cdot q_{ji}}{\pi_i \, q_{ij}}, 1\right), i,j \geq 1 \qquad (4.3.3.4)$$

Set

$$p_{ij} = \begin{cases} q_{ij}\alpha_{ij}, & \text{if } i \neq j \\ q_{ii} + \sum_{k \neq i} q_{ik}(1-\alpha_{ik}), & \text{if } i=j \end{cases} \qquad (4.3.3.5)$$

$$\Rightarrow \sum_{k \geq 1} p_{ik} = \sum_{k \neq i} p_{ik} + p_{ii} \overset{(4.3.3.5)}{=}$$

$$\sum_{k \neq i} q_{ik} \cdot \alpha_{ik} + q_{ii} + \sum_{k \neq i} q_{ik}(1-\alpha_{ik})$$

$$= q_{ii} + \sum_{k \neq i} q_{ik} = \sum_{k \geq 1} q_{ik} = 1$$

$\Rightarrow$ $p_{ij}, i,j \geq 1$ trans. matrix

Denote by $X_n, n \geq 0$ a Markov chain with trans. matrix (4.3.3.5)

Now choose $q_{ij}, i,j \geq 1$ s.t. $X_n, n \geq 0$ is irreducible (e.g. $q_{ij} > 0, i,j \geq 1$)

Check that π and P are in detailed balance, i.e

$$\pi_i p_{ij} = \pi_j p_{ji}, \quad i \neq j$$

(4.3.5.5) $\Longleftrightarrow$

$$\pi_i q_{ij} \alpha_{ij} = \pi_j q_{ji} \alpha_{ji}, \quad i \neq j \qquad (*)$$

1. case: $\alpha_{ij} = \dfrac{\pi_j q_{ji}}{\pi_i q_{ij}} \quad (\leq 1)$

$\Longrightarrow \pi_j q_{ji} \leq \pi_i q_{ij} \Longrightarrow 1 \leq \dfrac{\pi_i q_{ij}}{\pi_j q_{ji}}$

(4.3.3.4) $\Longrightarrow \alpha_{ji} = 1$

$\Longrightarrow$ (*) holds

2. case: $\alpha_{ij} = 1$

(4.3.3.4) $\Longrightarrow$

$$\frac{\pi_j q_{ji}}{\pi_i q_{ij}} \geq 1 \Longrightarrow \pi_j q_{ji} \geq \pi_i q_{ij} \Longrightarrow$$

$$1 \geq \frac{\pi_i q_{ij}}{\pi_j q_{ji}} \Longrightarrow \alpha_{ji} = \frac{\pi_i q_{ij}}{\pi_j q_{ji}}$$

$\Longrightarrow$ (*) holds

Th. 3.7.5 $\Longrightarrow$ $X_n, n \geq 0$ reversible and has an invariant distribution given by $\pi_j = P(X = x_j) = C \cdot b_j$

Th. 3.5.7 $\Longrightarrow$ $X_n, n \geq 0$ positive recurrent (i.e. $m_i < \infty, i \geq 1$)

Let $f(j) \overset{\text{def}}{=} h(x_j)$ in (4.3.3.1) be bounded

Ergodic theorem (Th. 3.8.2) $\Longrightarrow$

$$\frac{1}{n} \sum_{i=0}^{n-1} f(X_i) \xrightarrow[n \to \infty]{} \overline{f},$$

where

$$\overline{f} = \sum_i \pi_i \cdot \underbrace{f(i)}_{=h(x_i)}$$

Since $\pi_i = P(X = x_i), i \geq 1$ we get

$$\overline{f} = \sum_{i \geq 1} P(X = x_i) \cdot h(x_i) = E[h(X)] = \Theta$$

Choose N large. Then simulate a path $X_0, X_1, X_2, \ldots$ according to Sect. 4.3.2

$$\Longrightarrow \frac{1}{N} \sum_{i=0}^{N-1} f(X_i) \approx \Theta$$

Note: $\alpha_{ij} \overset{\text{def}}{=} \min\left(\dfrac{\pi_j q_{ji}}{\pi_i q_{ij}}, 1\right) \overset{\pi_j = C \cdot b_j}{=} \min\left(\dfrac{b_j q_{ji}}{b_i q_{ij}}, 1\right)$ ← not dependent on C!

$\Longrightarrow p_{ij}, i, j \geq 1$ explicitly known

5. Continuous-time Markov chains

$\longrightarrow$ stoch. proc. $X_t, t \in J$ without memory, where J is a time-interval (e.g. $J = [0, \infty)$)

5.1 Continuous-time stochastic processes

In this section we consider continuous-time stoch. proc.

$$X_t, 0 \leq t < \infty$$

with values in countable state space I.

In what follows we want to study right-continuous proc. $X_t, 0 \leq t < \infty$ in the following sense:

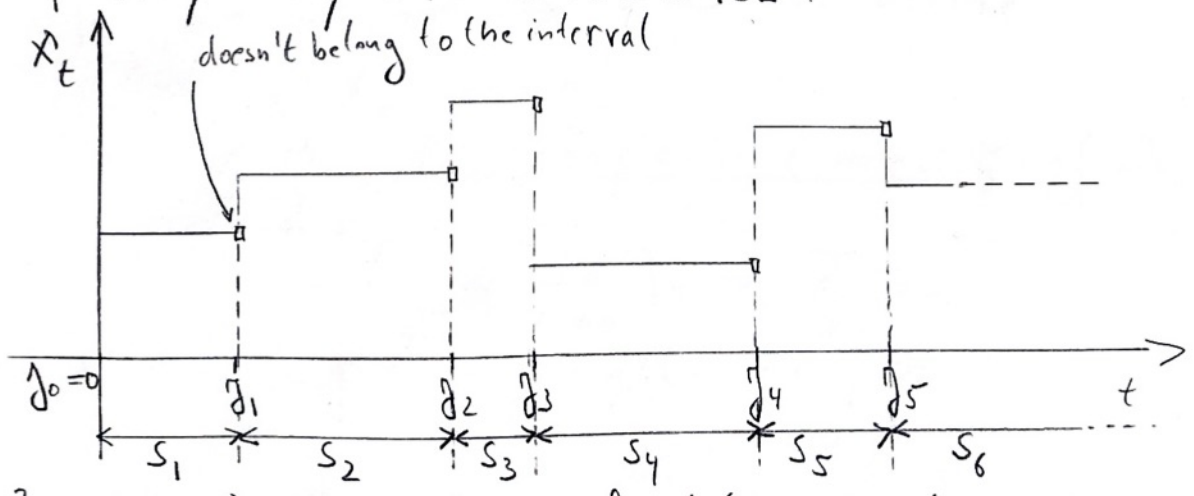
For all $t \geq 0$ and outcomes of X_t there exists a $\varepsilon > 0$ such that

$$X_s = X_t \qquad (5.1.1)$$

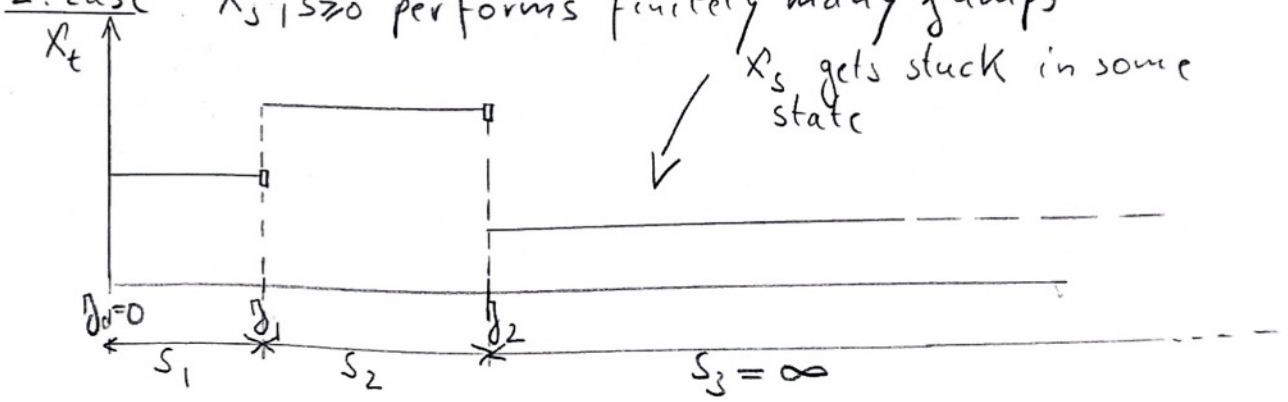
for $t \leq s \leq t + \varepsilon$.

$\longrightarrow$ 3 possible types of paths ($t \mapsto X_t$)

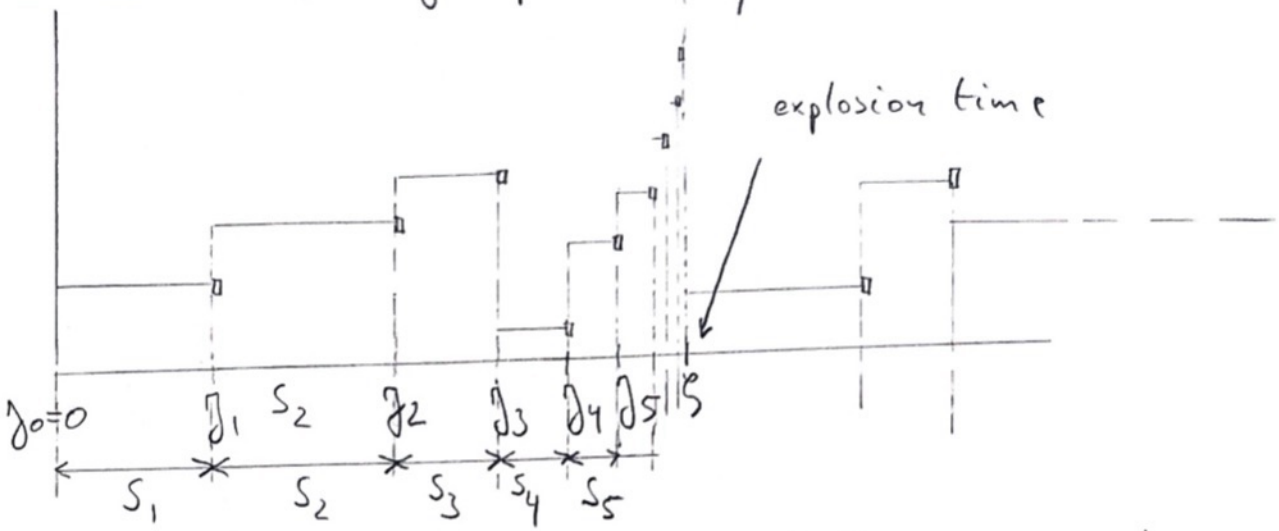
1. case: $X_s, s \geq 0$ has infinitely many jumps, but only finitely many on intervals $[0, t]$:



2. case: $X_s, s \geq 0$ performs finitely many jumps



3. case: $X_s, s \geq 0$ jumps infinitely often on an interval $[0,t]$



$\longrightarrow$ $\tau_0, \tau_1, \tau_2, \ldots$ jump times of $X_s, s \geq 0$ and

$S_1, S_2, S_3, \ldots$ holding times

or more precisely: Def. stopping times τ_n by

$$\tau_{n+1} \stackrel{\text{def.}}{=} \inf\{t \geq \tau_n : X_t \neq X_{\tau_n}\}, \quad \tau_0 = 0 \qquad (5.1.2)$$

for $n \geq 0$, where $\inf \emptyset \stackrel{\text{def}}{=} \infty$

and random times by S_n by

$$S_n = \begin{cases} \tau_n - \tau_{n-1}, & \text{if } \tau_{n-1} < \infty \\ \infty & \text{else} \end{cases} \qquad (5.1.3)$$

If $\tau_n < \infty$, but $\tau_{n+1} = \infty$ we define

$$X_\infty = X_{\tau_n} \qquad (5.1.4)$$

Further, the (first) explosion time ζ of $X_s, s \geq 0$ is defined by

$$\zeta = \sup_{n \geq 0} \tau_n = \sum_{n \geq 1} S_n \qquad (5.1.5)$$

In the sequel we call the proc.

$$Y_n \stackrel{\text{def.}}{=} X_{\tau_n}, \quad n \geq 0 \qquad (5.1.6)$$

the jump chain of $X_s, s \geq 0$, if $Y_n, n \geq 0$ is a discrete-time Markov chain