7. Variance of the Committor Function

A natural question arising the context of committor function computation on the basis of time series is of how the discrete committor function q does depend on uncertainties in the underlying data. In this chapter we will present an approach which allows to estimate these uncertainties element-wise. We will restrict ourselves to the case of Markov chains, i.e. discrete in time and *in space*. The idea behind that approach is to estimate the uncertainties in terms of the element-wise variances of a discrete committor function ensemble resulting from an ensemble of transition matrices distributed according to the discrete likelihood function $\mathcal{L}_d$. We will devise an appropriate Monte Carlo Markov Chain (MCMC) sampling procedure and will illustrate the approach on examples. The extension to Markov jump processes will not be discussed here and will be subject to further investigations.

7.1. The Discrete Committor Function

Let $P \in \mathbb{R}^{d\times d}$ be the transition matrix of a Markov chain on the state space $S \cong \{1,\dots,d\}$. As shown for example in [10] via *first step analysis*, the discrete committor function $q : S \mapsto [0,1]$ with respect to two disjoint, non-empty sets $A, B \subset S$ satisfies the discrete committor equation:

$$\begin{cases} \sum_{j=1}^{d}(p_{ij}-\delta_{ij})q_j = 0 & \forall\ i \in S\setminus(A\cup B), \\ q_i = 0 & \forall\ i \in A, \\ q_i = 1 & \forall\ i \in B, \end{cases} \tag{7.1}$$

where δ_{ij} is the Kronecker symbol. When only a finite observation $Y = \{y_0 = X(t_0), \dots, y_N = X(t_N)\}$ of the Markov chain is available, the transition matrix of the underlying Markov chain is not accessible and has to be estimated from the data. Unlike to the case of Markov jump processes, the inverse modeling of a Markov chain, i.e. reconstructing a Markov chain on the basis of a finite observation Y, is easy. It is defined as the Markov chain which most likely explains the observed data, i.e. which maximizes its discrete likelihood. Recall that the discrete likelihood function of an observation Y is given by (cf. Sect. 5.2)

$$\mathcal{L}_d(Y;P) = \prod_{k=0}^{N-1} p_{y_k,y_{k+1}} = \prod_{i,j\in S} p_{ij}^{c_{ij}}, \tag{7.2}$$

where $P = (p_{ij})_{i,j\in S}$ is the transition matrix of the underlying Markov chain and the frequency matrix $C = (c_{ij})_{i,j\in S}$ provides the number of consecutively observed transitions between states.

Remark 7.1.1. *We want to point out that here and in the following we assume that the prior probability over the transition matrices* ***before*** *observing any data is simply a uniform distribution. In particular, that assumption implies*

$$\mathcal{L}_d(Y;P) = \mathcal{L}_d(P;Y).$$

Henceforth, we will denote for a fixed observation Y the discrete likelihood function in (7.2) by $\mathcal{L}_d(P)$.

The maximum likelihood estimator (MLE) $\hat{P} = (\hat{p}_{ij})_{i,j\in S}$, i.e. the transition matrix which maximize the discrete likelihood function (7.2) given the observation Y, is unique and its entries $\hat{p}_{ij}$ can be expressed in terms of the frequency matrix,

$$\hat{p}_{ij} = \frac{c_{ij}}{c_i}, \tag{7.3}$$

where $c_i = \sum_{k\in S} c_{ik}$ is the total number of observed transitions leaving the state i.

Due to the finite number of observations, the transition probabilities in the MLE $\hat{P}$ are afflicted with uncertainty. The question is how do these uncertainties affect the committor function $\hat{q}$ computed via

$$\begin{cases} \sum_{j=1}^{d}(\hat{p}_{ij} - \delta_{ij})\hat{q}_j = 0 & \forall\ i \in S \setminus (A \cup B), \\ \hat{q}_i = 0 & \forall\ i \in A, \\ \hat{q}_i = 1 & \forall\ i \in B. \end{cases} \tag{7.4}$$

In other words, we are interested in the error $\|q - \hat{q}\|$ given an observation Y but, unfortunately, that error cannot directly be measured since the "true" committor function $q = (q_i),\ i \in S$ is unknown and the MLE $\hat{P}$ does not indicate the involved uncertainties. However, following standard reasonings, the error $\|q - \hat{q}\|$ can be estimated via the *variance* of the committor function given an observation Y.

7.2. Metropolis Markov Chain Monte Carlo

One way to estimate the variance of the committor is to draw an ensemble of transition matrices $\{P_1, \ldots, P_k\}$ from the *conditional probability distribution* of all possible transition matrices given the observation Y. Then the variance of the committor $\hat{q}$ is approximately given by the variance of the resulting ensemble of committor functions $\{q_1, \ldots, q_k\}$ computed via (7.1), respectively. One option to generated such an ensemble of transition matrices can be found in [85]. They follow a Bayesian approach to derive a *conditional probability distribution* of all possible transition matrices given the observation Y by assuming that the prior probability over the transition matrices before observing any data is given by Dirichlet distributions. Moreover, they derive efficient methods to sample from the resulting posterior distribution. However, the resulting ensembles do depend *explicitly* on the choice of parameters for the prior Dirichlet distributions and, therefore, the Bayesian approach is from our point of view inappropriate for the computation of the variance of the committor function $\hat{q}$.

We follow an alternative approach via Markov Chain Monte Carlo (MCMC) simulation. For notational convenience we denote the set of all transition matrices by

$$\mathfrak{P} = \left\{ P = (p_{ij})_{i,j \in S} : p_{ij} \in [0,1], \sum_{k \in S} p_{ik} = 1 \quad \forall i,j \in S \right\}.$$

We devise an MCMC Metropolis scheme to generate an ensemble of transition matrices which is distributed according to the discrete likelihood function $\mathcal{L}_d$ restricted on the set $\mathfrak{P}$. Compared to the Bayesian approach, we do not assume any prior distribution of transition matrices.

A MCMC Metropolis scheme works basically as follows. Suppose you want to sample from a probability distribution which is induced by a density function $f \in L^1(\mathbb{R}^d)$. Let $x_\mathcal{C} \in \mathbb{R}^d$ be the current state under consideration in the ensemble. In the proposal step a new state $x_\mathcal{N} \in \mathbb{R}^d$ is generated. In the acceptance step the proposed state $x_\mathcal{N}$ is accepted with the probability

$$p_{acc} = \min \left\{ 1, \frac{f(x_\mathcal{N}) \cdot p(x_\mathcal{C} \to x_\mathcal{N})}{f(x_\mathcal{C}) \cdot p(x_\mathcal{N} \to x_\mathcal{C})} \right\}, \tag{7.5}$$

where $p(x_\mathcal{C} \to x_\mathcal{N})$ is probability of generating the state $x_\mathcal{N}$ conditional on the state $x_\mathcal{C}$ and $p(x_\mathcal{N} \to x_\mathcal{C})$ is defined analogously. If the new state is accepted than $x_\mathcal{N}$ is added to the ensemble and the scheme restarts with $x_\mathcal{N}$ as the current state. Otherwise, the current state $x_\mathcal{C}$ is added to the ensemble and is considered again in the next iteration of the scheme.

Let us in the following comment on several issues concerning the MCMC sampling procedure:

- The target density function $f \in L^1(\mathbb{R}^d)$ does not have to be normalized because only the ratio $f(x_\mathcal{N})/f(x_\mathcal{C})$ is involved in the acceptance probability in (7.5).

- The sampling of a probability distribution *restricted* on a subset of the state space, say $R \subset \mathbb{R}^d$, can easily be achieved by modifying the density function f according to

  $$f_R(x) \overset{def}{=} \mathbb{1}_R(x) f(x).$$

  If the MCMC sampling procedure is started with $x_\mathcal{C} \in R$ then the ratio in the acceptance probability in (7.5),

  $$\frac{f_R(x_\mathcal{N})}{f_R(x_\mathcal{C})} = \frac{1_R(x_\mathcal{N}) f(x_\mathcal{N})}{1_R(x_\mathcal{C}) f(x_\mathcal{C})} = \mathbb{1}_R(x_\mathcal{N}) \frac{f(x_\mathcal{N})}{f(x_\mathcal{C})},$$

  is well defined during the sampling procedure and the resulting ensemble is distributed according to f restricted on R.

- In principle, one can use any strategy for the generation of a new state in the proposal step as long as one is able to evaluate the probabilities $p(x_\mathcal{C} \to x_\mathcal{N})$ and $p(x_\mathcal{N} \to x_\mathcal{C})$. The choice of the proposal step strategy, however, is crucial for the efficiency and the convergence of the sampling procedure. For a discussion on these issues see, e.g., [13].

7.3. Ensemble of Transition Matrices via MCMC

We are interested in sampling the distribution induced by the discrete likelihood function $\mathcal{L}_d(P)$. In the following, it is convenient to represent the target density function $f(P) = \mathcal{L}_d(P)$ as

$$f(P) = e^{-g(P)} \text{ with } g(P) \overset{def}{=} -\log(\mathcal{L}_d(P)). \tag{7.6}$$

7.3.1. Dynamics on the Transition Matrix Space

For the generation of a proposal state $P_{\mathcal{N}}$ we exploit the fact that the non-normalized probability density function $\rho(P)$ of the invariant measure associated with the SDE

$$\mathrm{d}P_t = -\nabla g(P_t)dt + \sqrt{2}\ \mathrm{d}W_t \tag{7.7}$$

is given by

$$\rho(P) = e^{-g(P)} = \mathcal{L}_d(P),$$

where $P \in \mathbb{R}^{d^2}$ is understood as a d^2-dimensional vector and W_t is a d^2-dimensional standard Wiener process. A scheme for the generation of a proposal state $P_{\mathcal{N}}$ is obtained by discretizing the SDE in (7.7) by means of the Euler-Maruyama-scheme,

$$P_{\mathcal{N}} = P_{\mathcal{C}} - \nabla g(P_{\mathcal{C}})\Delta t + \sqrt{2\Delta t}\ \eta, \tag{7.8}$$

where $0 < \Delta t \in \mathbb{R}$ denotes the discretization time step and the random variable η is drawn from a d^2-dimensional standard Gaussian distribution with mean $0 \in d^2$ and covariance matrix $I = \mathrm{diag}(1, \ldots, 1) \in \mathbb{R}^{d^2 \times d^2}$. Unfortunately, the proposal step equation in (7.8) does not preserve the transition matrix property, i.e. $P_{\mathcal{N}} \notin \mathfrak{P}$, because the Gaussian random variable η is unbounded. One option is to choose a sufficiently small time discretization step Δt such that $p_{ij} \in [0, 1],\ 0 \le i, j \le d$ but in general $P_{\mathcal{N}}$ is not a stochastic matrix, i.e. $\sum_{m \in S} p_{im} \neq 1$.

7.3.2. MCMC on the Frequency Matrix Space

Motivation

As a preparation for an alternative approach, recall that if only an incomplete observation of a Markov chain with discrete state space S is available, the transition matrix $\hat{P} = (\hat{p}_{ij}),\ i, j \in S$ which most likely explains the data is given by

$$\hat{p}_{ij} = \frac{c_{ij}}{\sum_{k \in S} c_{ik}}, \tag{7.9}$$

where an entry c_{ij} of the frequency matrix $C = (c_{ij}),\ i, j \in S$ provides the number of observed transitions from i to j. The relation in (7.9) can be written in compact form,

$$\hat{P} = u(C),$$

where the function $u(C) : \mathbb{R}^{d^2} \mapsto \mathfrak{P}$ is defined as

$$u(C) \overset{def}{=} \left(\frac{c_{11}}{\sum_{m \in S} c_{1m}}, \ldots, \frac{c_{dd}}{\sum_{m=1}^{d} c_{dm}} \right) \in \mathfrak{P}.$$

To avoid any notational confusion with respect to the empirical frequency matrix, we will denote in the following a general frequency matrix by K.

The crucial idea is now to generate an ensemble of frequency matrices $\mathcal{K} = \{K \in \mathbb{R}_+^{d^2}\}$ via an MCMC procedure which is distributed according to the likelihood function $\mathcal{L}_d(u(C))$. We will show that the ensemble $\mathcal{P} = \{P = u(K) : K \in \mathcal{K}\}$ is distributed according to $\mathcal{L}_d(P)$.

Derivation of the MCMC Procedure on the Frequency Matrix Space

We consider a dynamics on the state space of frequency matrices,

$$\mathrm{d}K_t = -\nabla\tilde{g}(K_t)dt + \sqrt{2\beta^{-1}}\,\mathrm{d}W_t, \tag{7.10}$$

where $K_t = (k_{ij})_{i,j\in S} \in \mathbb{R}^{d^2}$, the factor β^{-1} can be seen as an artificial temperature. The function $\tilde{g} : \mathbb{R}^{d^2} \mapsto \mathbb{R}$ is defined according to

$$\tilde{g}(K) \overset{def}{=} g(u(K)),$$

where the function g is defined in (7.6). Then the probability density function $\rho(K)$ of the invariant distribution of (7.10) is given by

$$\rho(K) = e^{-\beta\tilde{g}(K)} = \left[\mathcal{L}_d(u(K))\right]^\beta . \tag{7.11}$$

The time discretization of (7.10) via the Euler-Maruyama-scheme yields an equation for the proposal step,

$$K_\mathcal{N} = K_\mathcal{C} - \nabla\tilde{g}(K_\mathcal{C})\Delta t + \sqrt{2\beta^{-1}\Delta t}\;\eta, \tag{7.12}$$

where the gradient $\nabla\tilde{g}(K)$ takes the form

$$\nabla\tilde{g}(K) = (\frac{c_1}{k_1} - \frac{c_{11}}{k_{11}}, \dots, \frac{c_d}{k_d} - \frac{c_{dd}}{k_{dd}})^T, \tag{7.13}$$

with $k_i = \sum_{m=1}^d k_{im}$ and Δt and η are as in (7.8).

It remains to derive an expression for the probability $p(K_\mathcal{C} \to K_\mathcal{N})$ but this immediately follows by realizing that the difference $\Delta K = K_\mathcal{N} - K_\mathcal{C}$ is distributed according to a d^2-dimensional Gaussian distribution with mean $-\Delta t\nabla\tilde{g}(K_\mathcal{C}) \in \mathbb{R}^{d^2}$ and covariance matrix $2\beta^{-1}\Delta tI \in \mathbb{R}^{d^2\times d^2}$. Consequently, the probability to generate the proposal state $K_\mathcal{N}$ while being in the current state $K_\mathcal{C}$ is

$$p(K_\mathcal{C} \to K_\mathcal{N}) = Z^{-1}\exp\left[-\frac{1}{4\beta^{-1}\Delta t}\|\Delta K + \nabla\tilde{g}(K_\mathcal{C})\Delta t\|^2\right],$$

where Z is a normalization factor.

In order to ensure that the matrix $u(K_\mathcal{N})$ is a transition matrix, i.e. $u(K_\mathcal{N}) \in \mathfrak{P}$, we generate an ensemble of frequency matrices restricted on the subset (cf. Sect. 7.2)

$$\mathfrak{K} = \left\{K \in R_+^{d^2} : k_i^- < \sum_{m=1}^d k_{im} < k_i^+\right\}, \tag{7.14}$$

where $0 < k_i^- < k_i^+,\ i = 1,\dots,d$. The particular choice of the boundary conditions for $\mathfrak{K}$ will become clear in Section 7.3.3.

Combining all issues, we finally end up with the Metropolis algorithm, as stated in Algorithm 9, to generate an ensemble of transition matrices distributed according to the discrete likelihood function $\mathcal{L}_d(P)$.

Algorithm 9 Metropolis algorithm

```
Input: Frequency matrix $C = (c_{ij})_{i,j \in S}$, number of MCMC steps $nMCMC$, time step $\Delta t$, temperature $\beta^{-1}$.
Output: Ensemble $\mathcal{P}$ of transition matrices.
(1) Initialize $K_{\mathcal{C}} := C$.
(2) FOR $n = 1$ TO $nMCMC$ DO
(3)   Generate proposal frequency vector $K_{\mathcal{N}} = (k_{ij})$:
          $K_{\mathcal{N}} = K_{\mathcal{C}} - (\frac{c_1}{k_1} - \frac{c_{11}}{k_{11}}, \dots, \frac{c_d}{k_d} - \frac{c_{dd}}{k_{dd}})^T + \sqrt{2\Delta t}\ \eta.$
(4)   Accept $K_{\mathcal{N}}$ with acceptance probability ($\Delta K = K_{\mathcal{N}} - K_{\mathcal{C}}$):
          $p_{acc} = \min \left\{ 1, \mathbb{1}_{\mathfrak{K}}(K_{\mathcal{N}}) \frac{\mathcal{L}_d(u(K_{\mathcal{N}})) \exp\left[-\frac{1}{4\beta^{-1}\Delta t}\|\Delta K + \nabla\tilde{g}(K_{\mathcal{C}})\Delta t\|^2\right]}{\mathcal{L}_d(u(K_{\mathcal{C}})) \exp\left[-\frac{1}{4\beta^{-1}\Delta t}\|-\Delta K + \nabla\tilde{g}(K_{\mathcal{N}})\Delta t\|^2\right]} \right\}.$
(5)   If $K_{\mathcal{N}}$ is accepted THEN set $K_{\mathcal{C}} := K_{\mathcal{N}}$.
(6)   Add $u(K_{\mathcal{C}})$ to the transition matrix ensemble $\mathcal{P}$.
(7) END FOR
```

7.3.3. Proof of Correctness

It remains to prove that resulting ensemble of transition $\mathcal{P} = \{u(\mathcal{K})\}$ is indeed distributed according to $\mathcal{L}_d(P)$.

Theorem 7.3.1. *Let $\mathcal{K} = \{K \in \mathfrak{K}\}$ be an ensemble of frequency matrices distributed according to $\mathcal{L}_d(u(K))$. Then the ensemble $\mathcal{P} = \{u(K) : K \in \mathcal{K}\}$ is distributed according to $\mathcal{L}_d(P)$.*

Proof. We prove that for all $P \in \mathfrak{P}$ holds

$$\mathbb{P}[u(K) = P] \propto \mathcal{L}_d(P).$$

Without loss of generality, we restrict ourselves to the first row vector $K^{(1)} = (k_{11}, \dots, k_{1d})$ of a frequency matrix $K \in \mathcal{K}$. For the sake of notational simplicity we write in the following

$$u(k_{11}, \dots, k_{1d}) = (\frac{k_{11}}{\sum_{m=1}^{d} k_{1m}}, \dots, \frac{k_{1d}}{\sum_{m=1}^{d} k_{1m}}),$$
$$\mathcal{L}_d(p_{11}, \dots, p_{1d}) = \prod_{j=1}^{d} (p_{1j})^{c_{1j}}.$$

Let $\mathfrak{P}^{(1)} = \{p = (p_{11}, \dots, p_{1d}) : p \in \mathbb{R}_+^d, \sum_{j=1}^{d} p_{1j} = 1\}$. Since $\mathfrak{P}^{(1)} \subset \mathbb{R}^d$ is an (d-1)-dimensional manifold we represent an element $p \in \mathfrak{P}^{(1)}$ by

$$p = (p_{11}, \dots, p_{1(d-1)}, 1 - \sum_{j=1}^{d-1} p_{1j}).$$

Furthermore, we will denote in the following by $\Pi(p), p \in \mathfrak{P}^{(1)}$ the projection onto the first (d-1) components of p, i.e,

$$\Pi(p) = (p_{11}, \dots, p_{1(d-1)}).$$

The crucial observation now is that due to the particular choice of the set $\mathfrak{K}$ in (7.14) we have

$$\{K^{(1)} : u(K^{(1)}) = (p_{11}, \dots, p_{1d})\} = \{(\alpha p_{11}, \dots, \alpha p_{1d}) : k_1^- < \alpha < k_1^+\}, \tag{7.15}$$

which motivates to consider the new observable $\tilde{K}^{(1)} = \tilde{T}(K^{(1)})$,

$$\begin{aligned} &\tilde{T} : \mathbb{R}_+^d \to \mathbb{R}_+ \times \Pi(\mathfrak{P}^{(1)}) \\ &\tilde{T}(k_{11}, \dots, k_{dd}) \mapsto (\alpha, p_{11}, \dots, p_{1(d-1)}), \\ &\quad \alpha = \sum_{m=1}^{d} k_{1m}, \quad p_{1j} = \frac{k_{1j}}{\sum_{m=1}^{d} k_{1m}}, \ j = 1, \dots, d-1. \end{aligned} \tag{7.16}$$

If we denote the probability density function associated with the new observable $\tilde{K}^{(1)}$ by $\tilde{\mathcal{L}}$ then it should be clear that

$$\mathbb{P}[u(K^{(1)}) = (p_{11}, \dots, p_{1d})] \propto \int_{k_1^-}^{k_1^+} \tilde{\mathcal{L}}(\alpha, p_{11}, \dots, p_{1(d-1)}) \mathrm{d}\alpha.$$

In Lemma 7.3.1 we show that $\tilde{\mathcal{L}}$ is simply given by

$$\tilde{\mathcal{L}}(\alpha, p_{11}, \dots, p_{1(d-1)}) = \mathcal{L}_d(p_{11}, \dots, p_{1d}) \alpha^{(d-1)},$$

where $p_{1d} = (1 - \sum_{j=1}^{d-1} p_{1j})$. But this immediately implies

$$\mathbb{P}[u(K^{(1)}) = (p_{11}, \dots, p_{1d})] \propto \mathcal{L}_d(p_{11}, \dots, p_{1d}).$$

and we are done. □

It remains to prove

Lemma 7.3.1.

$$\tilde{\mathcal{L}}(\alpha, p_{11}, \dots, p_{1(d-1)}) = \mathcal{L}_d(p_{11}, \dots, 1 - \sum_{j=1}^{d-1} p_{1j}) \alpha^{(d-1)}$$

Proof. The probability density function $\tilde{\mathcal{L}}(\tilde{K}^{(1)})$ associated with $\tilde{K}^{(1)} = (\alpha, p_{11}, \dots, p_{1(d-1)})$ is given by [58]

$$\tilde{\mathcal{L}}(\tilde{K}^{(1)}) = \mathcal{L}_d(u(\tilde{T}^{-1}(\tilde{K}^{(1)}))) \left| \det(J(\tilde{T}^{-1})(\tilde{K}^{(1)})) \right|, \tag{7.17}$$

where

$$\begin{aligned} &\tilde{T}^{-1} : \mathbb{R}_+ \times \Pi(\mathfrak{P}^{(1)}) \to \mathbb{R}_+^d \\ &\tilde{T}^{-1}(\alpha, p_{11}, \dots, p_{1(d-1)}) \mapsto (\alpha p_{11}, \dots, \alpha p_{1(d-1)}, \alpha(1 - \sum_{j=1}^{d-1} p_{1j})). \end{aligned} \tag{7.18}$$

The first factor in (7.17) reduces to

$$\begin{aligned} \mathcal{L}_d(u(\tilde{T}^{-1}(\tilde{K}^{(1)}))) &= \mathcal{L}_d(u(\alpha p_{11}, \dots, \alpha p_{1(d-1)}, \alpha(1 - \sum_{j=1}^{d-1} p_{1j}))) \\ &= \mathcal{L}_d(p_{11}, \dots, p_{1d}), \end{aligned}$$

where $p_{1d} = (1 - \sum_{j=1}^{d-1} p_{1j})$.

Finally, we compute the determinant in (7.17):

$$\det J(\tilde{T}^{-1}) = \begin{vmatrix} p_{11} & \alpha & 0 & 0 & \dots \\ p_{12} & 0 & \alpha & 0 & \dots \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ p_{1(d-1)} & 0 & \dots & \dots & \alpha \\ 1 - \sum_{j=1}^{d-1} p_{1j} & -\alpha & \dots & \dots & -\alpha \end{vmatrix}$$

$$= \begin{vmatrix} p_{11} & \alpha & 0 & 0 & \dots \\ p_{12} & 0 & \alpha & 0 & \dots \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ p_{1(d-1)} & 0 & \dots & \dots & \alpha \\ 1 & 0 & \dots & \dots & 0 \end{vmatrix}$$

$$=(-1)^{(d-1)} \begin{vmatrix} 1 & 0 & \dots & \dots & 0 \\ p_{11} & \alpha & 0 & 0 & \dots \\ p_{12} & 0 & \alpha & 0 & \dots \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ p_{1(d-1)} & 0 & \dots & 0 & \alpha \end{vmatrix} = (-1)^{(d-1)} \alpha^{(d-1)}.$$

□

7.4. Numerical Experiments

7.4.1. Dirichlet Distribution

In the first example we use the derived MCMC method to sample from a two-dimensional (non-normalized) Dirichlet-distribution

$$\mathcal{L}(p_1, p_2) = (p_1)^{c_1} \cdot (p_2)^{c_2} \cdot (1 - p_1 - p_2)^{c_3} \tag{7.19}$$

on the state space $D = \{p_1 + p_2 + p_3 = 1 : p_1, p_2, p_3 \geq 0\}$ with parameters $c_1, c_2, c_3 > 0$. For our numerical experiments we chose two different sets of parameters, namely $\mathcal{C}_1 = (c_1 = 3, c_2 = 8, c_3 = 10)$ and $\mathcal{C}_2 = (c_1 = 43, c_2 = 8, c_3 = 15)$. We sampled both distribution at the "temperature" $\beta^{-1} = 10$ via Algorithm 9 and generated an ensemble consisting of 10^6 transition matrices, respectively. As boundary conditions for the restriction $\mathfrak{K}$ in (7.14) we chose $k_1^- = (\sum_{j=1}^{3} c_j) - 5$ and $k_1^+ = (\sum_{j=1}^{3} c_j) + 5$. For the simulation with respect to the parameter set $\mathcal{C}_1$ we had for the time step $\Delta t = 10^{-3}$ an acceptance rate of 93% and with respect to $\mathcal{C}_2$ for $\Delta t = 10^{-2}$ an acceptance rate of 96%.

In Figure 7.1 we compare the distribution of the ensemble from the simulation with respect to the parameter set $\mathcal{C}_1$ (top right panel) with the corresponding analytical distribution (top left panel). For the sake of comparison, we normalized all distributions such that their respective maximal value is one. The distributions resulting for the parameter set $\mathcal{C}_2$ are given in the second row of Figure 7.1. One can see both distributions are well sampled.

Let us comment on the choice of the simulation parameters. The simulation's temperature $\beta^{-1} = 5$ ensures that even states with a very low statistical weight

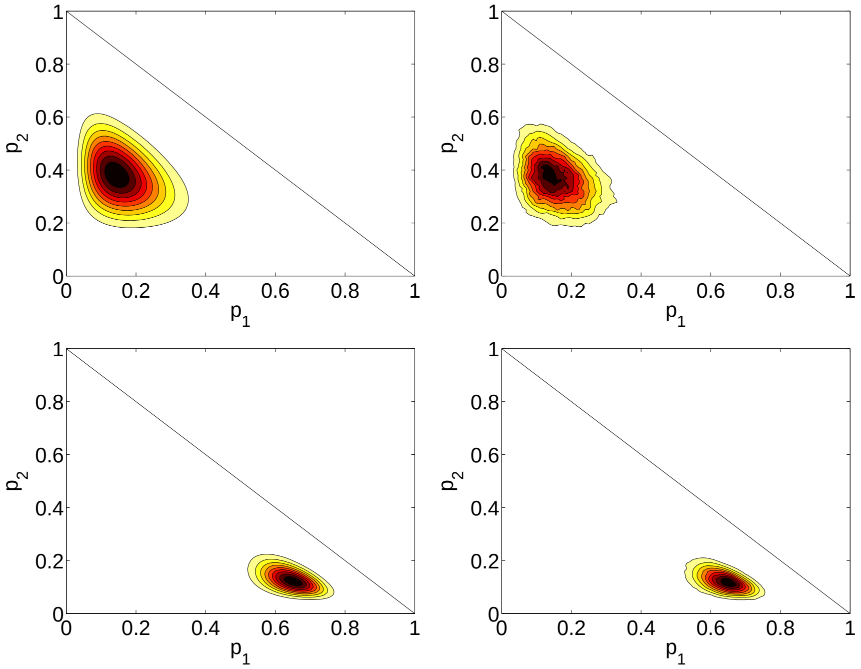

Figure 7.1.: We compare the distributions of the Dirichlet distribution in (7.19) (first column) with the distributions of the ensembles generated via Algorithm 9 (second column) with respect to the parameter set $\mathcal{C}_1 = (c_1 = 3, c_2 = 8, c_3 = 10)$ (first row) and $C_2 = (c_1 = 43, c_2 = 8, c_3 = 15)$ (second row). For example, the analytical distribution with respect to $\mathcal{C}_1$ attains its maximum at $(\frac{c_1}{c1+c2+c3}, \frac{c_2}{c1+c2+c3}) \approx (0.14, 0, 38)$.

with respect to the target distribution $\mathcal{L}_d(u(K))$ are sufficiently often proposed such that the variance is right reproduced. For realistic values of the parameters ($c_i > 100$), however, our extensive numerical experiments have shown that the Dirichlet distribution in (7.19) is already well sampled at a low temperature $\beta^{-1} = 1$.

7.4.2. Small Example

In this section we demonstrate the performance of the derived Algorithm 9 on a Markov chain with a small state space $S \cong \{1, \ldots, 25\}$. This example is constructed such that it allows to relate the element-wise variances of the resulting ensemble of committor functions to an underlying discretized potential landscape.

As exemplified in the Section 4.3.1, a Smoluchowski diffusion process in a potential landscape can be approximated by a Markov jump process where the infinitesimal generator L of the approximating Markov jump process results from a finite differences discretization scheme of the generator, associated with the diffusion process (cf. Sect. A.3). Doing so, a transition matrix can easily be obtained because the generator L generates a semigroup of transition matrices via $P(t) = \exp(tL)$. For a particular choice of $t > 0$ we will call $P(t) = \exp(tL)$ transition matrix.

For our numerical experiments, we utilized the generator given in (4.44) which results from an approximation of the Smoluchowski dynamics in the three-hole potential landscape. We approximated the diffusion (at temperature $\beta^{-1} = 1$) on a

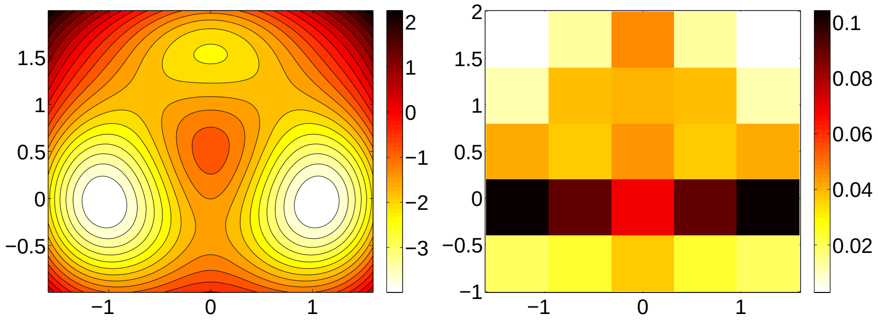

Figure 7.2.: Left: Contour plot of the three-hole potential (3.45). Right: Box plot of the stationary distribution associated with the 25-state Markov chain $P = \exp(1.2L)$.

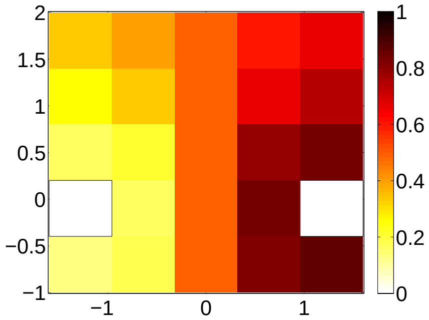

Figure 7.3.: Box plot of the committor function associated with the transition matrix $P = \exp(1.2L)$. As the sets A and B, we chose the two states with the highest stationary distribution. The set A consists of the state corresponding to the left white box and the set B consists of the state corresponding to the right white box.

5×5 mesh of the domain $\Omega = [-1.5, 1.5] \times [-1, 1.5]$ which results in a generator $L \in \mathbb{R}^{25\times 25}$ on a discrete state space of 25 states.

The potential landscape of the three-hole potential in (3.45) is illustrated as a contour plot in the left panel of Figure 7.2. In the right panel, we show a box plot of the stationary distribution of the transition matrix $P(1.2) = \exp(1.2 \cdot L) \in \mathbb{R}^{25\times 25}$. Although we used an extremely coarse-grained mesh (5×5), one can clearly see that the equilibrated dynamics of the Markov chain reflects the topology of the potential landscape, e.g., the two states in the Markov chain with highest stationary probability correspond to the two deep minima in the potential landscape, respectively. The discrete committor function with respect to $P(1.2)$ is illustrated in Figure 7.3. As the set A and B we chose the two states with the highest stationary distribution (depicted by white boxes). The main question we were interested in is of how the element-wise variances of a committor ensemble do depend on the length N of the observed time series of the Markov chain. For this purpose, we generated via Algorithm 9 a sequence of committor function ensembles $\{q_1^{(N)}, \dots, q_6^{(N)}\}$ for time series of length $N = 10^3, \dots, N = 10^8$ where the respective time series were all subsampled from a fixed realization of the Markov chain. For each ensemble we

N	10^3	10^4	10^5	10^6	10^7	10^8
κ	13.34	14.03	12.27	11.97	11.91	11.94

Table 7.1.: The condition number κ of the matrix $\hat{P}^{(N)} - I$ *after* elimination of the condition $q_i = 0, \forall i \in A, q_i = 1, \forall i \in B$ which arises from solving of the discrete committor equation (7.4). The table gives the condition number κ as a function of the length N of the considered time series. Results for time series all subsampled from the same realization.

sampled $m = 500000$ committor functions where we used the discretization time step $\Delta t = 10^{-3}$ in the proposal step equation (7.12). As boundary conditions for the restriction $\mathfrak{K}$ in (7.14) we chose $k_i^- = c_i - 15$ and $k_i^+ = c_i + 15$ where $c_i = \sum_{j=1}^{d} c_{ij}$. In all simulations we had an acceptance rate of about $> 93\%$.

In the following $\hat{P}^{(N)}$ denotes the MLE transition matrix resulting from the time series of length N and $\hat{q}^{(N)}$ is the associated committor function. The mean committor function of an ensemble $\{q_1^{(N)}, \dots, q_6^{(N)}\}$ is denoted by $\bar{q}^{(N)}$ and the variance by $var(\bar{q}^{(N)})$. In Figure 7.4 we illustrate the committor function $\hat{q}^{(N)}$ (first column), mean committor function $\bar{q}^{(N)}$ (second column) and its variance $var(\bar{q}^{(N)})$ (third column) for all ensembles, respectively. At first glance, one can see that the committor functions $\hat{q}^{(N)}$ are almost identical with the corresponding mean committor function $\bar{q}^{(N)}$ of the ensemble, except for the length $N = 10^3$. Beside the observation that the variance of the ensembles decreases by the same order of magnitude as the length N increases, the box plots in third column reveal that the states with the lowest stationary distribution exhibit the highest variance in the committor function. The observations are confirmed by the graphs shown in Figure 7.5. In the left panel, we plot the maximal variance $\|var(\bar{q}^{(N)})\|_\infty$ of the mean committor functions $\bar{q}^{(N)}$ as a function of the length N of the respective time series whereas in the right panel the error $\|\bar{q}^{(N)} - \hat{q}^{(N)}\|$ (measured in the 2-norm) between the mean committor function $\bar{q}^{(N)}$ and the committor function $\hat{q}^{(N)}$ is shown as a function of the length N of the respective time series.

Our numerical experiments have shown that the committor function $\hat{q}^{(N)}$ even for short time series ($N = 10^3$) almost coincides with the expected committor function with respect to the discrete likelihood function $\mathcal{L}_d$.

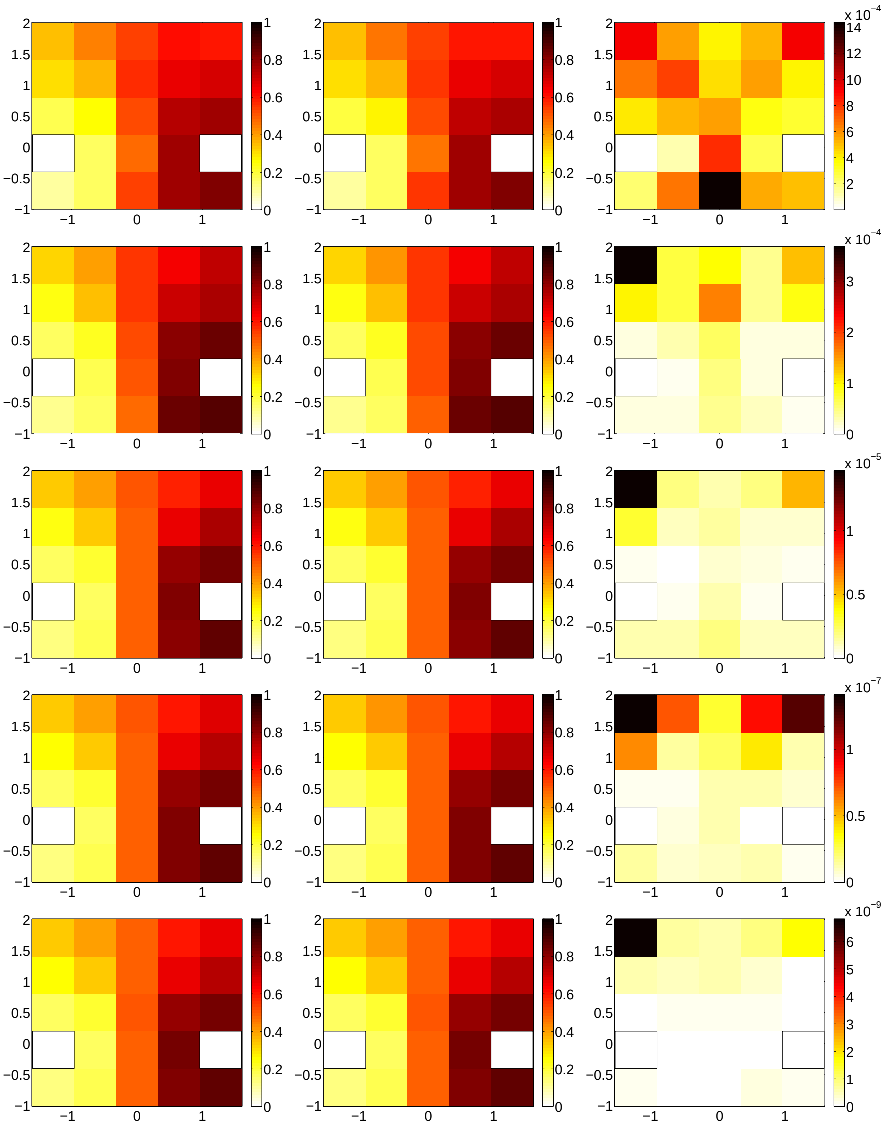

Figure 7.4.: Left column: Box plots of the committor functions $\hat{q}^{(N)}$ resulting from the MLE transition matrix $\hat{P}^{(N)}$ (7.3), respectively. Middle column: Box plots of the mean committor functions $\bar{q}^{(N)}$ of the committor function ensembles $\{q_1^{(N)}, \ldots, q_5^{(N)}\}$, respectively. Right column: Box plots of the variances $var(\bar{q}^{(N)})$ of the mean committor functions, respectively. Results for different lengths $N = 10^3$ (top), $\ldots, N = 10^7$ (bottom) of respective time series all subsampled from the same realization.

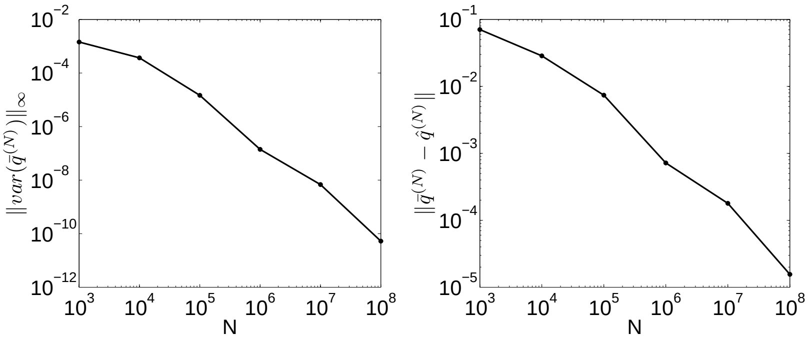

Figure 7.5.: Left: The maximal variance $\|var(\bar{q}^{(N)})\|_\infty$ of the mean committor functions $\bar{q}^{(N)}$ as a function of the length N of the respective time series. Right: The error $\|\bar{q}^{(N)} - \hat{q}^{(N)}\|$ (measured in the 2-norm) between the mean committor function and the committor function resulting from the MLE transition matrix $\hat{P}^{(N)}$ in (7.3) as a function of the length N of the respective time series. Results for time series all subsampled from the same realization.

7.4.3. Glycine

In the last example we apply the MCMC methods in order to estimate the uncertainties of the forward committor function q^+ in the glycine in solvent example from Section 4.3.2. We are aware that in the glycine-example the forward committor function q^+ is computed via an (estimated) generator L of a Markov jump process and *not* via the transition matrix of a Markov chain. Nevertheless, the MCMC method allows to get an idea of the uncertainties because q^+ is almost identical with the discrete committor function $\hat{q}^+$ based on the MLE $\hat{P}$ and computed via (7.4). Both committor functions are illustrated in the panels of Figure 7.6.

For the estimation of the variance of the committor function $\hat{q}^+$ we generated an ensemble of $7 \cdot 10^6$ transition matrices ($\Delta t = 10^{-5}$) and computed the element-wise variances of the resulting ensemble of committor functions. The boundary conditions for the restriction $\mathfrak{K}$ were the same as in the previous example. To be more precise, instead of generating a full transition (counts) matrix in each step of the simulation, we used the structure of the MLE $\hat{P}$ as a template, i.e., we only generated entries k_{ij} if $c_{ij} > 0$. In each iteration step of the Algorithm 9 we solved the discrete committor equation in (7.4) with respect to the current transition matrix. Finally, a clever update-scheme allowed us to compute the element-wise variances of the committor function ensemble $\{q^+_{MCMC}\}$ *on the fly* (see the end of this section).

The final variances are illustrated in the left panel of Figure 7.7 where the boxes are colored according to the log-values of the respective variances in order to emphasize the different orders of magnitudes. Again, the comparison of the variances element by element with the Gibbs energy of the Markov chain $\hat{P}$ reveals what intuitively should be clear; the states with high variance correspond to those with very high discrete free energy which is equivalent to a very small stationary distribution. In Figure 7.8 we show the maximal variance $\|var(\{q^+_{MCMC}\})\|_\infty$ as a function of the MCMC-steps.

We end this section by deriving the update-scheme for the "on the fly" com-

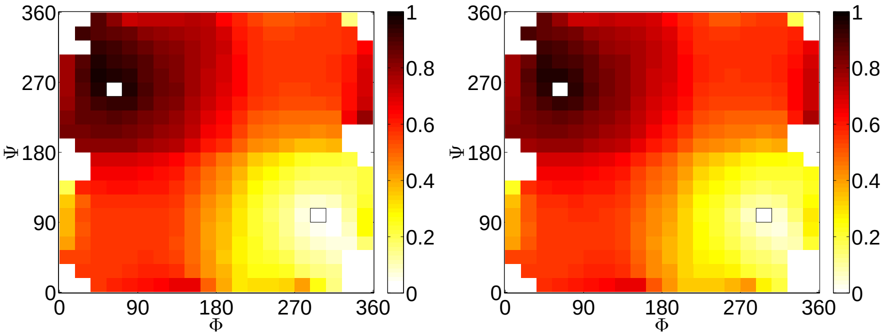

Figure 7.6.: Right: The panel shows the forward committor q^+ based on an estimated generator $\tilde{L}$ (cf. Sect. 4.3.2) and computed via (4.11). Left: The corresponding box plot of the discrete committor $\hat{q}^+$ based on the MLE $\hat{P}$ and computed via (7.4). As the set A we chose the box (shown as a white box with black boundary) which covers the peak of the restricted stationary distribution on the lower right conformation. The set B for the upper left conformation (shown as a white box) was chosen analogously. Results for an equidistant discretization of the torsion angle space into 20 × 20 boxes.

putation of the variances. We derive the scheme for a one-dimensional time series $(x_1, \ldots, x_N), x_i \in \mathbb{R}$. A short calculation shows that the estimator of the variance of the time series reduces to

$$\frac{1}{N+1}\sum_{i=1}^{N}\left(x_i - \sum_{j=1}^{N} x_j\right)^2 = \frac{1}{N+1}\left(s_1(N) - \frac{1}{N} s_2^2(N)\right), \qquad (7.20)$$

where $s_1(N) = \sum_{j=1}^N x_j^2$ and $s_2(N) = \sum_{j=1}^N x_j$. But this means if one is interested in the in the variance of the time series $(x_1, \ldots, x_N, x_{N+1})$ then only the sums s_1 and s_2 have to be updated and the right hand side in (7.20) yields the desired result with respect to $N' = N + 1$.

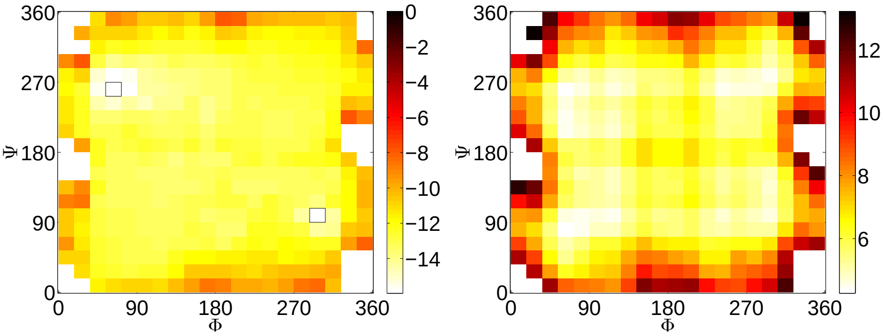

Figure 7.7.: The left panel illustrates the element-wise variances of the committor functions ensemble $\{q^+_{MCMC}\}$. In order to emphasize the variances' magnitudes of order, we chose a logarithmical scale. The comparison of the variances with the discrete free energy of the MLE Markov chain $\hat{P}$, as shown in the right panel, again reveals that the states with the highest variances correspond to those with the lowest statistical weights.

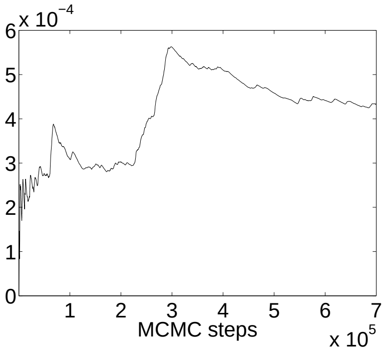

Figure 7.8.: The maximal variance $\|var(\{q^+_{MCMC}\})\|_\infty$ as the function of the MCMC steps.