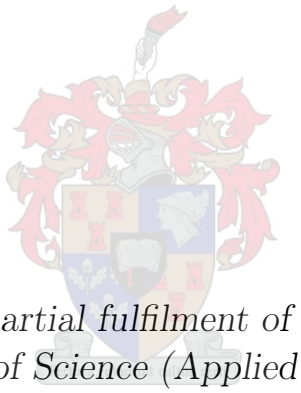
Variance Estimation for Markov Processes

by

Wessel Blomerus



Thesis presented in partial fulfilment of the requirements for the degree of Master of Science (Applied Mathematics) in the Faculty of Science at Stellenbosch University

Supervisor: Prof. Hugo Touchette

March 2021

Declaration

By submitting this thesis electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Date: February, 2021 .....

Abstract

Variance Estimation for Markov Processes

W Blomerus

Department of Mathematical Sciences,
Stellenbosch University,
Stellenbosch, South Africa.

Thesis: MSc (Applied Mathematics)

February 2021

We study the asymptotic variance of additive functionals of Markov processes, used in statistics and stochastic modelling as estimators of model parameters. The observations generated by these processes are correlated, which complicates the estimation of the asymptotic variance. In practice, methods for estimating the asymptotic variance are based on either estimating the correlation function or the segmentation of the additive observable (batch mean method). In this thesis, we propose and study three new estimators, based on a link between the asymptotic variance, large deviation theory, and an equation of probability theory called the Poisson equation. The first two estimators rely on the fact that the solution of the Poisson equation can be represented as a conditional expectation. The third estimator is based on a stochastic approximation of the solution of the Poisson equation, suggested by recent works in large deviation theory, which describe the solution as an eigenfunction that can be iteratively estimated in an 'online' way as a simulation unfolds. We illustrate these three estimators for simple Markov processes, including Markov chains and diffusion processes, for which the asymptotic variance is exactly known.

Uittreksel

Variansieberaming van Markov-Prosesse

("Variance Estimation for Markov Processes")

W Blomerus

Departement Wiskundige Wetenskappe,
Universiteit van Stellenbosch,
Stellenbosch, Suid-Afrika.

Tesis: MSc (Toegepaste Wiskunde)

Februarie 2021

Ons bestudeer die asimptotiese variansie van additiewe funksionale vir Markov-prosesse wat gebruik word in statistiek en stogastiese modellering as beramers van model parameters. Dié prosesse genereer gekorreleerde waarnemings wat die beraming van die asimptotiese variansie kompliseer. In die praktyk word metodes om die asimptotiese variansie te beraam gebaseer op óf die beraming van die korrelasiefunksie óf die segmentasie van die additiewe waarneming (lot gemiddeldes metode). In hierdie tesis stel ons drie nuwe beramers voor wat gebaseer is op die verwantskap tussen die asimptotiese variansie, teorie van groot afwykings en 'n vergelyking van waarskynlikheidsteorie, genaamd die Poisson-vergelyking. Die eerste twee beramers is gebaseer op die feit dat the oplossing van die Poisson-vergelyking in terme van 'n voorwaardelike verwagting uitgebeeld kan word. Die derde beramer is gebaseer op 'n stogastiese benadering van die Poisson-vergelyking se oplossing. Hierdie benadering word voorgestel in onlangse werk in die teorie van groot afwykings, waarin die oplossing beskryf word as 'n eiefunksie wat iteratief benader kan word in 'n "aanlyn"-manier soos die simulasie ontvou. Ons gebruik hierdie drie beramers op eenvoudige Markov-prosesse, naamlik Markov-kettings en diffusieprosesse, waar die asimptotiese varianse bekend is.

Acknowledgements

First and foremost, I wish to acknowledge and express my sincere gratitude to my supervisor, Prof. Hugo Touchette, for his continuous support during my master's degree. The completion of my thesis would not have possible without his patience, guidance, advice and encouragement. Our stimulating conversations provided me with countless eureka moments and a newfound motivation to pursue my PhD.

During my studies, I was funded by the National Institute for Theoretical Physics (NITheP). I am grateful for their financial support.

I am thankful for my entire family's support leading up to and during my master's degree. A special thank you to my mother Louisa and father Jaco for paving the way towards my education and supporting me through challenging times. Thank you to my brother Tiaan for making the COVID-19 lockdown and working from home bearable.

Lastly, I would like to thank my friends for providing me with much needed social distractions. Each coffee break played an important role in maintaining the balance between my academic and social life.

Contents

List of Figures

Chapter 1

Introduction

In this thesis we study the asymptotic variance of time-integrated functionals of Markov processes. We provide a review of existing estimation techniques used to estimate the asymptotic variance and propose as our main result new estimators that can be used in an online way, that is, estimators that update the asymptotic variance according to newly observed data as the process unfolds. In this chapter we provide background regarding our thesis topic which serves as a motivation for the study. Thereafter we state the objective of our research as well as give an outline of the structure of the thesis.

1.1 Statistical estimators

In statistics and stochastic modelling, we are often faced with the task of estimating parameters of a model using noisy data obtained from an experiment or a simulation. Mathematically, we can formulate this task in a general way as that of estimating the expectation $\mathbb{E}[g(X)]$ of some function g of a random variable X representing the state of the system in some steady state. The data obtained from an experiment or simulation is then represented as a sequence $X_1, X_2, \ldots, X_n$ of n observations or measurements, from which $\mathbb{E}[g(X)]$ is to be estimated [1, 2, 3].

In the simplest case, the sequence of observations can be assumed to be (or is as a matter of design) a sequence of independent and identically distributed (i.i.d.) random variables, from which it is natural to estimate $\mathbb{E}[g(X)]$ using the sample mean

$$\hat{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} g(X_i), \tag{1.1}$$

The hat notation of $\hat{\mu}_n$ refers to the empirical estimator of $\mu = \mathbb{E}[g(X)]$. Under the i.i.d. model, this estimator has the desired property of being consistent, that is, it converges to the expectation as the number of observations or data n increases. In probability theory, this convergence is expressed by the law of large numbers, which states that the sample mean converges (in probability) to the expected value as $n \to \infty$ [4].

In practice, we are limited to a finite number of observations and therefore we want to ascertain how close the sample mean is to the true value of the parameter to be estimated,

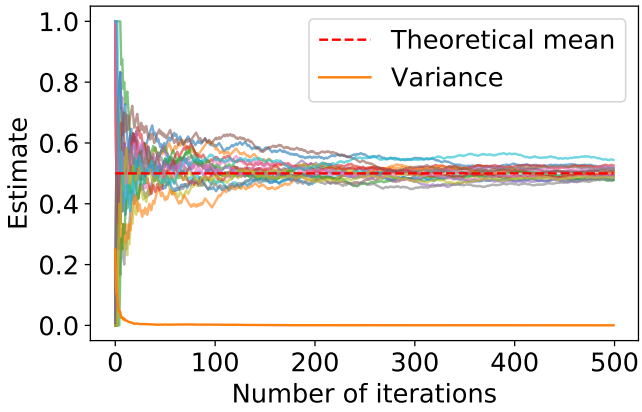


Figure 1.1: Sample means (1.1) and variance (orange) of Bernoulli processes representing the fraction of heads in a sequence of coin flips. We use $p = 0.5$ in the numerical experiment as the coin probability and see, indeed, that the sample mean convergences to 0.5 with the variance decreasing to 0 as n is increased.

as represented by $\mathbb{E}[g(X)]$. Since the estimator is a random variable, this translates to finding the probability that the estimator is a certain distance away from the exact value. As the number of observations increases, the estimated value falls within some range or spread around the true value of the parameter.

If we assume that the estimator is normally distributed around the true value, then the width of the spread is directly related to the width of the normal distribution. This assumption is confirmed by the central limit theorem [4] when the variance of the observations in the i.i.d. model is finite. Then the distribution of the sample mean is approximately Gaussian for large n and centered around the expected value of the distribution sampled from. The variance of the Gaussian decreases with n, the number of observations, as σ_A^2/n, where σ_A^2 is a constant known as the asymptotic variance. This constant, which is the focus of this thesis, can also be estimated in the i.i.d. model using the unbiased estimator

$$\hat{\sigma}_n^2 = \frac{1}{n-1}\sum_{i=1}^{n-1}\left(g(X_i) - \hat{\mu}_n\right)^2 \tag{1.2}$$

or the more common estimator

$$\hat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^{n}\left(g(X_i) - \hat{\mu}_n\right)^2, \tag{1.3}$$

whose bias decreases to 0 as $n \to \infty$. Thus, from the same sequence of observations or data, we obtain an estimate of the parameter that we are interested in together with an estimate of its variance, which serves as an error bar for the estimated parameter.

To illustrate these results, let us consider the simple task of determining whether a coin is fair or biased. Suppose we describe consecutive coin flips as a binary sequence, where a 1 or 0 represents a coin landing on heads or tails, respectively. This can also be

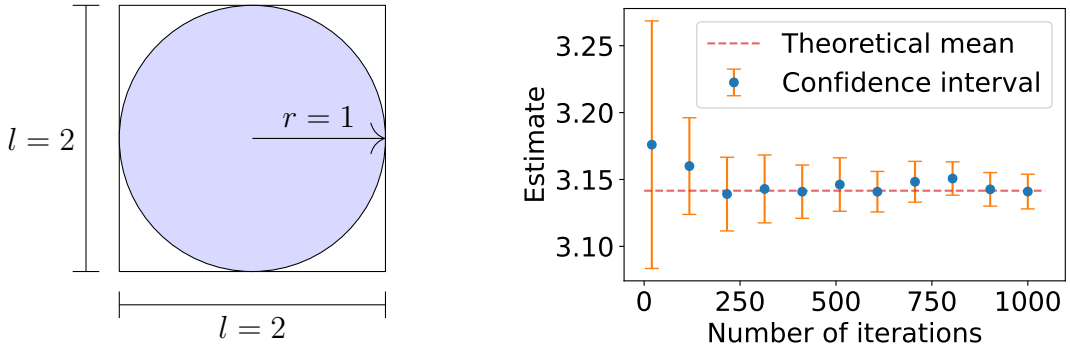


Figure 1.2: Random estimation of π. (Left) Random points are generated uniformly in the square and counted as 4 if they fall in the disk (blue). (Right) Convergence of $\hat{\pi}_n$ (blue) to the known constant π (dashed red) with confidence intervals (orange) as the number of iterations increases.

seen as a Bernoulli process where p is the probability that the coin lands on heads. Such a sequence is i.i.d. and, by construction, its sample mean (1.1) determines the fraction of heads observed. Therefore, a sequence generated by a fair coin must have a sample mean converging to 0.5 as the number of heads must show up, on average, half of the time.

As an example, we calculate the sample mean as a function of the number of iterations n for 20 coin flip sequences with $p = 0.5$. From the results shown in Fig. 1.1, we see that the sample mean of each sequence does in fact converge to the expected value of a fair coin (dashed red). Despite the random nature of each sequence, the law of large numbers guarantees the overall convergence to the expected value. To quantify the convergence we calculate the variance of the sample means as the number of iterations increases (orange) using (1.3) and see that quickly it converges to 0.

As a second example, let us consider another classical problem, namely, that of estimating the constant π using the experiment shown in Fig. 1.2 (left), which consists of a disk of radius $r = 1$ inscribed in a square within which random uniform points (X_i, Y_i) are generated. From basic geometry, it is easy to see that a consistent estimator of π is given by

$$\hat{\pi}_n = \frac{4}{n} \sum_{j=1}^{n} \mathbb{1}((X_j, Y_j) \text{ in circle}), \tag{1.4}$$

where n is the number of observed points and $\mathbb{1}(\cdot)$ is the indicator function (equal to 1 if the point falls in the disk and 0 otherwise). The results shown in Fig. 1.2 (right) illustrate that $\hat{\pi}_n$ (blue) converges to π (dashed red) as the number of points generated increases. In order to determine the confidence intervals (orange), we calculate the variance using (1.3). Confidence intervals are important estimators as they represent an estimated range of possible values that could contain the true expected value, in this case, π.

1.2 Correlated samples

In practice, the i.i.d. assumption rarely holds when modeling problems that arise from real life or practical situations, as the state of a system at a given time usually depends on its state at some previous time, resulting in correlations in the observations of this state.

To model these correlations, we need to describe the sequence $X_1, X_2, \ldots, X_n$ as a stochastic process, which in the simplest case (and in this thesis) is taken to be a Markov process. Instances where this type of process has been used include:

- In finance and economics, to model the growth of asset prices and stock values [5]. Stock values, in particular, are often modelled with a Markov process in continuous time called the geometric Brownian motion [6].
- In meteorology, where weather generators are used to simulate the daily air temperature and amount of precipitation [7, 8]. An example of this is where the current amount of precipitation is used to assign the day to a state in a Markov chain to estimate the precipitation of the next day.
- In renewable energy for forecasting wind power generated by turbines or predicting the power generation of solar cells [9]. For instance, the type and amount of cloud cover present is classified and assigned to states of a Markov chain in order to estimate the batteries' charge at a future instance [10].

These models present various parameters of interest that can be estimated, as in the i.i.d. case, from a sequence of observations or states $X_1, X_2, \ldots, X_n$ using the same sample mean given in (1.1), even if the states are related in a Markov way. The consistency of this estimator is ensured in the Markov case by an extension of the law of large numbers called the ergodic theorem, which states, under some conditions on the process considered, that the sample mean converges in probability to a stationary expectation [2]. Moreover, there is a generalisation of the central limit theorem for Markov processes that guarantees that the distribution of the sample mean converges to a Gaussian distribution for large n [11, 12, 13]. A crucial difference, however, between the i.i.d. and Markov cases is that the asymptotic variance σ_A^2 is not determined by the estimator (1.3), due to the correlation between observations.

Estimating the asymptotic variance for correlated random variables is a non-trivial task. In practice, there are two conventional approaches for estimating σ_A^2. The first is based on estimating the correlation between observations as the asymptotic variance is defined by the autocorrelation function [14]. The second is based on decorrelating the observations by partitioning the sequence containing all the observations into smaller subsequences or batches and then calculating the mean of the batches [15, 16, 17]. It can be shown that the sequence containing the batch means is i.i.d., which simplifies the estimation of σ_A^2 to a block version of the 'naïve' variance estimator in (1.3) [18].

These methods are used in many applications, but do come with some challenges, which we will discuss in some detail in the next chapters. In fact, discussing some of these methods, Bratley [19, p. 101] notes that

> No completely satisfactory method for analyzing the output of steady-state simulations has yet been devised. Certainly, no consensus has been reached regarding the relative merits of existing methods. The above recommendations are lukewarm, but we have nothing better to suggest – other than perhaps to reassess whether a steady-state simulation should be carried out.

One problem, as we will see, is that these methods contain parameters, such as batch size and truncation length, whose accuracy are dependent on the process considered. A second problem is that these methods might struggle to produce accurate estimates for cases where the amount of data is finite or limited.

1.3 Objectives and contributions

The main objective of this thesis is to present and test new methods for estimating the asymptotic variance of sample means arising in the context of Markov processes. We are particularly interested in 'online' methods or algorithms that estimate the asymptotic variance in 'real time' as the process considered unfolds or as observations are gathered. To this end, we use a connection between the asymptotic variance and a probabilistic version of the Poisson equation, first considered by Neveu in 1972 [20], to present three new estimators of the asymptotic variance.

The first two estimators are based on representing the solution of the Poisson equation as a conditional expectation, following results from Dynkin, Pardoux and Veretennikov [21, 22], which can be computed numerically using different realisations of the data or one long stream of data. The third estimator is based on a stochastic approximation of the solution of the Poisson equation, which can be used in an online way on a stream of data to iteratively calculate the asymptotic variance. Each of these methods are implemented on Markov processes and their accuracy evaluated.

1.4 Outline

The thesis is structured as follows. In Chapter 2 we give a general introduction of Markov processes as well as discuss the asymptotic variance associated with their observables. We also introduce and implement the large deviation theory required to theoretically determine the asymptotic variance. In Chapter 3 we review conventional estimation techniques currently used in practice to estimate the asymptotic variance and illustrate them by means of examples. Chapter 4 contains the theory regarding the Poisson equation associated with a Markov process and its observable and presents the link between the Poisson solution and the asymptotic variance. In Chapter 5 we propose and evaluate numerical estimators based on the theory covered in Chapter 4. We conclude the thesis in Chapter 6 with a summary of our findings and list open questions regarding the proposed online estimator.

Chapter 2

Asymptotic variance of Markov additive functionals

We present in this chapter the Markov models used in this thesis, focusing on two particular classes: Markov chains evolving in discrete time and Markov diffusions evolving in continuous time defined by stochastic differential equations. We explain for both classes how the asymptotic variance of additive functionals is determined from the covariance function, and illustrate the results on various examples that will be studied again in the rest of the thesis. The material of this part is taken from [23, 12, 24]. We also introduce another method, based on large deviation theory, to obtain the asymptotic variance, which we illustrate with examples and validate with known results. We base our introduction for this part on the material found in [25, 26, 27].

2.1 Markov chains

Let $X_1, X_2, \ldots, X_n$ be a sequence of n random variables (RVs) taking values in a set $\mathcal{X}$, called the state space. This sequence forms a Markov chain if its joint probability

$$P(x_1, x_2, \ldots, x_n) = P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n) \tag{2.1}$$

is given by

$$P(x_1, x_2, \ldots, x_n) = P(X_1 = x_1)P(X_2 = x_2|X_1 = x_1) \ldots P(X_n = x_n|X_{n-1} = x_{n-1}), \tag{2.2}$$

where $P(X_1 = x_1)$ is the initial distribution of the first state and $P(X_n = x_n|X_{n-1} = x_{n-1})$ is the conditional probability associated with transitioning from state X_{n-1} to state X_n at the n-th transition [2]. This factorization of the joint probability defines the Markov or 'memoryless' property, where the probability of transitioning from one state to another only depends on the current state. This is usually emphasized by writing the sequence of RVs as

$$X_1 \to X_2 \to \cdots \to X_n \tag{2.3}$$

to show that X_n only depends on X_{n-1} and that the sequence, as a whole, can be seen as a trajectory of states evolving in discrete time $i = 1, 2, \ldots, n$.

For simplicity, we assume that the state space $\mathcal{X}$ is discrete, and write $P(X_n = x_n)$ as

$$P(X_n = i) = \pi_i^{(n)}, \quad i \in \mathcal{X}, \tag{2.4}$$

which can be seen as components of a vector $\pi^{(n)}$. Similarly, the conditional probability $P(X_n = x_n | X_{n-1} = x_{n-1})$ can be written in matrix form with

$$p(i,j)(n) = P(X_n = j | X_{n-1} = i), \quad i \in \mathcal{X} \quad \text{and} \quad j \in \mathcal{X}, \tag{2.5}$$

representing the (i,j)th entry of a matrix $\mathbf{P}(n)$. The resulting matrix is called the transition matrix and has the obvious property that

$$\sum_{j \in \mathcal{X}} p(i,j)(n) = 1 \quad \forall i \in \mathcal{X}. \tag{2.6}$$

Throughout the thesis, we consider Markov chains that are time-homogeneous, meaning that the transition matrix $\mathbf{P}(n)$ does not depend on n. More specifically, the matrix remains constant with respect to time and, as a result, we can write

$$P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n) = \pi(1)_{x_1} p(x_1, x_2) \ldots p(x_{n-1}, x_n). \tag{2.7}$$

By summing over states, one can verify that the probability of transitioning from i to j in m time steps is written as

$$P(X_{n+m} = j | X_n = i) = P(X_{m+1} = j | X_1 = i) = (\mathbf{P}^m)_{ij}, \tag{2.8}$$

where $\mathbf{P}^m$ is $\mathbf{P}$ to the power m.

From (2.7) we can also infer that the distribution $P(X_n = i) = \pi_i^{(n)}$ can be expressed as

$$\begin{aligned} \pi_i^{(n)} &= \sum_k P(X_{n-1} = k) P(X_n = i | X_{n-1} = k) \\ &= \sum_k \pi_k^{(n-1)} p(k,i) \\ &= (\pi^{(n-1)} \mathbf{P})_i. \end{aligned} \tag{2.9}$$

By recursively applying this equation, we obtain an expression that describes the evolution of a Markov chain

$$\pi^{(n)} = \pi^{(1)} \mathbf{P}^n, \tag{2.10}$$

where the row vector $\pi^{(1)}$ represents the initial distribution over all states.

The behaviour of $\pi^{(n)}$ as $n \to \infty$ depends on the transition matrix. For the thesis, we assume that the Markov chain is ergodic. This means technically that the chain is aperiodic and positive recurrent [28] and ensures that the distribution $\pi^{(n)}$ of X_n 'evolves' to a unique distribution π, called the ergodic or stationary distribution. Thus,

$$\pi = \lim_{n \to \infty} \pi^{(1)} \mathbf{P}^n \tag{2.11}$$

starting from any distribution $\pi^{(1)}$ for X_1 [2]. The unique stationary distribution is thus a fixed point of $\mathbf{P}$, which means that it can be calculated as

$$\pi = \pi \mathbf{P}. \tag{2.12}$$

Thus π is the right eigenvector of $\mathbf{P}$ with eigenvalue of 1.

2.2 Additive functionals of Markov chains

The problem considered in this thesis is to study the convergence of estimators having the general form

$$S_n(g) = \frac{1}{n}\sum_{i=1}^{n} g(X_i), \tag{2.13}$$

where $g : \mathcal{X} \to \mathbb{R}$ is a real-valued function of the X_i's. We also call $S_n(g)$ an additive functional of the Markov chain, an observable, or simply a sample mean.

Assuming that the Markov chain is ergodic, we have that $S_n(g)$ converges in probability to the value

$$\pi(g) = \mathbb{E}_\pi[g(X)] = \sum_{i\in\mathcal{X}} \pi_i g(i), \tag{2.14}$$

where $\mathbb{E}_\pi[\cdot]$ denotes the expectation with respect to the stationary distribution π. This convergence is known as the ergodic theorem [12], which can be expressed mathematically as

$$\lim_{n\to\infty} P(|S_n(g) - \pi(g)| > \varepsilon) = 0 \tag{2.15}$$

for all $\varepsilon > 0$. As a shorthand, we also write $S_n(g) \to \pi(g)$ in probability. This generalizes the law of large numbers, which applies to independent and identically distributed (i.i.d.) RVs, to Markov chains.

As for sequences of i.i.d. RVs, there is also a central limit theorem associated with $S_n(g)$, which basically states that

$$S_n(g) \xrightarrow{\text{dist}} \mathcal{N}\left(\pi(g), \frac{\sigma_A^2}{n}\right) \tag{2.16}$$

as $n \to \infty$, where $\xrightarrow{\text{dist}}$ means convergence in distribution and $\mathcal{N}(\mu, \sigma^2)$ denotes a normal distribution with mean μ and variance σ^2. To be more precise,

$$\lim_{n\to\infty} P\left(\frac{nS_n(g) - n\pi(g)}{\sqrt{n}\sigma_A}\right) = \mathcal{N}(0,1). \tag{2.17}$$

This means, in essence, that the distribution of $S_n(g)$ is approximately Gaussian close to $\pi(g)$ and has a variance that decreases with n, in agreement with (2.16). The constant σ_A^2 is called the asymptotic variance and is the main focus of the thesis. Note that if we re-center g with respect to the stationary distribution, by introducing

$$\bar{g} = g - \pi(g),$$

we can rewrite these results as $S_n(\bar{g}) \to 0$ in probability and

$$S_n(\bar{g}) \xrightarrow{\text{dist}} \mathcal{N}\left(0, \frac{\sigma_A^2}{n}\right), \tag{2.18}$$

respectively. This re-centering will be important in the remainder of the thesis. Note that σ_A^2 is the same for $S_n(g)$ and $S_n(\bar{g})$.

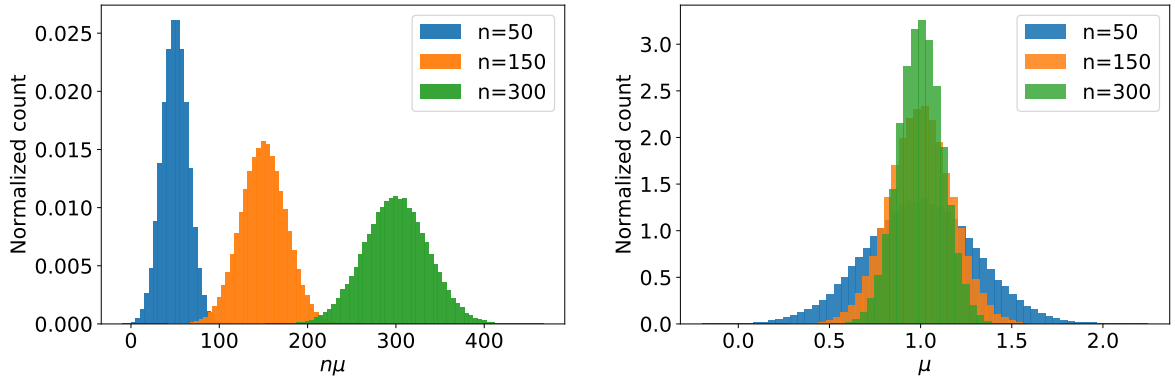


Figure 2.1: (Left) Empirical distributions of a Gaussian sum nS_n with $n = [50, 150, 300]$ RVs having $\mu = 1$ and $\sigma^2 = 2$. (Right) Visualization of the law of large numbers for a normalized Gaussian sum S_n for $n = [50, 150, 300]$, $\mu = 1$ and $\sigma^2 = 2$.

To substantiate the concepts described above, we illustrate (2.15) and (2.16) by means of a simple example. We calculate the sum nS_n with $g(X) = X$ for $n = [50, 150, 300]$ RVs sampled from $\mathcal{N}(1, 2)$. From the results shown in Fig. 2.1 (left), we see that the sum is distributed according to a normal distribution with mean $n\mu$ and variance $n\sigma^2$. The law of large numbers states that the sample mean converges almost surely to the expected value of the sampled distribution as $n \to \infty$, which is clearly shown in Fig. 2.1 (right).

For a general Markov chain, we can express the asymptotic variance as

$$\begin{aligned}\sigma_A^2 &= \lim_{n\to\infty} \frac{1}{n} \mathrm{Var}\left(\sum_{i=1}^{n} g(X_i)\right) \\ &= \lim_{n\to\infty} \frac{1}{n}\left[\sum_{i=1}^{n} \mathrm{Var}(g(X_i)) + \sum_{j=1}^{n}\sum_{k=1}^{n} \mathrm{Cov}(g(X_j), g(X_k))\right], \end{aligned} \tag{2.19}$$

where $j \neq k$ in the double sum,

$$\mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] \tag{2.20}$$

is the variance and

$$\mathrm{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] \tag{2.21}$$

is the covariance function. Since the double sum reproduces terms twice, we can write

$$\sigma_A^2 = \lim_{n\to\infty} \frac{1}{n}\left[\sum_{i=1}^{n} \mathrm{Var}(g(X_i)) + 2\sum_{1\leq j<k\leq n} \mathrm{Cov}(g(X_j), g(X_k))\right]. \tag{2.22}$$

Moreover, if we assume that the Markov chain is stationary, the two sums converge, by

the Cesàro mean property, to

$$\sigma_A^2 = \mathrm{Var}_\pi(g(X_i)) + 2\sum_{k=1}^{\infty} \mathrm{Cov}_\pi(g(X_i), g(X_{i+k})) \tag{2.23}$$

$$= \mathbb{E}_\pi[(g(X_i) - \pi(g))^2] + 2\sum_{k=1}^{\infty} \mathbb{E}_\pi[(g(X_i) - \pi(g))(g(X_{i+k}) - \pi(g))], \tag{2.24}$$

where the choice of i is arbitrary and $\mathrm{Var}_\pi(\cdot)$ and $\mathrm{Cov}_\pi(\cdot)$ denotes the variance and covariance calculated with the stationary distribution, respectively. In terms of $\bar{g}$ we have equivalently

$$\sigma_A^2 = \mathbb{E}_\pi[\bar{g}(X_1)^2] + 2\sum_{k=2}^{\infty} \mathbb{E}_\pi[\bar{g}(X_1)\bar{g}(X_k)]. \tag{2.25}$$

See Theorem 17.5.3 of [28] for a proof.

2.3 Examples

We illustrate the concepts introduced up to this point by means of three examples. We are interested in the sample mean S_n with $g(X_i) = X_i$ of a Markov chain $(X_i)_{i=1}^n$ and show how to obtain the ergodic average $\pi(X)$ and its asymptotic variance.

2.3.1 Two-state Markov chain

The first example that we consider is a two-state Markov chain with states $X_i \in [0, 1] = \mathcal{X}$ and transition matrix

$$\mathbf{P} = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix}, \tag{2.26}$$

where $0 < \alpha, \beta \leq 1$. It can be shown from (2.12) that, if $\alpha \notin \{0, 1\}$ and $\beta \notin \{0, 1\}$, then the Markov chain $X_1, X_2, \ldots, X_n$ is ergodic and has a unique stationary distribution given as

$$\pi = \begin{pmatrix} \frac{\beta}{\alpha+\beta} & \frac{\alpha}{\alpha+\beta} \end{pmatrix}. \tag{2.27}$$

Therefore, we find

$$\mathbb{E}_\pi[X] = \frac{\alpha}{\alpha+\beta}. \tag{2.28}$$

To calculate the asymptotic variance we firstly determine the covariance given by (2.23) and (2.24) as

$$\begin{aligned} \mathrm{Cov}_\pi(X_i, X_{i+k}) &= \mathbb{E}_\pi\left[\left(X_i - \frac{\alpha}{\alpha+\beta}\right)\left(X_{i+k} - \frac{\alpha}{\alpha+\beta}\right)\right] \\ &= \mathbb{E}_\pi\left[X_i X_{i+k} - \frac{\alpha^2}{(\alpha+\beta)^2}\right]. \end{aligned} \tag{2.29}$$

Next, we explicitly evaluate $\mathbb{E}_\pi[X_i\, X_{i+k}]$ and note that the only nonzero contribution is when both X_i and X_{i+k} occupy state 1 at each respective time. For the case where $k = 1$,

staying at state 1 for one iteration is the only valid transition and gives $\mathbb{E}_\pi[X_i\, X_{i+1}] = \frac{\alpha}{\alpha+\beta}(1-\beta)$, leading to

$$\mathrm{Cov}_\pi(X_i, X_{i+1}) = \frac{\alpha}{\alpha+\beta}(1-\beta) - \frac{\alpha^2}{(\alpha+\beta)^2} = \frac{\alpha\beta(1-\alpha-\beta)}{(\alpha+\beta)^2}. \tag{2.30}$$

Repeating the process of evaluating $\mathbb{E}_\pi[X_i\, X_{i+k}]$ for all possible realizations, where $k = 2, 3, \ldots$ leads to a general formula of the form

$$\mathrm{Cov}_\pi(X_i, X_{i+k}) = \frac{\alpha\beta(1-\alpha-\beta)^k}{(\alpha+\beta)^2} \tag{2.31}$$

for $k \geq 0$. Lastly, we need the variance of the stationary distribution. We obtain it by either setting $k = 0$ in (2.31) or realising that the stationary distribution resembles a Bernoulli distribution, yielding

$$\mathrm{Var}_\pi(X) = \frac{\alpha\beta}{(\alpha+\beta)^2}. \tag{2.32}$$

Following (2.24), we then obtain the asymptotic variance as

$$\sigma_A^2 = \frac{\alpha\beta}{(\alpha+\beta)^2} + 2\sum_{k=1}^{\infty}\left[\frac{\alpha\beta(1-\alpha-\beta)^k}{(\alpha+\beta)^2}\right] = \frac{\alpha\beta(2-\alpha-\beta)}{(\alpha+\beta)^3}. \tag{2.33}$$

From the central limit theorem, we therefore have that, for n large but finite,

$$S_n = \frac{1}{n}\sum_{i=1}^{n} X_i \tag{2.34}$$

is approximately normally distributed with mean $\frac{\alpha}{\alpha+\beta}$ and variance $\frac{1}{n}\frac{\alpha\beta(2-\alpha-\beta)}{(\alpha-\beta)^3}$.

2.3.2 Autoregressive model

An autoregressive (AR) process is a higher-order Markov model in which the state $X_n \in \mathbb{R}$ depends linearly on the previous states $X_{n-1}, X_{n-2}, \ldots$. In the simplest case, called the autoregressive process of order 1 or AR(1), we have

$$X_n = c + \rho X_{n-1} + Y_n, \tag{2.35}$$

where c is a constant and Y_n is independent and identically distributed according to $\mathcal{N}(0, \varphi^2)$. The stochastic contribution of Y_n allows the model to represent stochastic differential equations. Writing the recursion (2.35) from X_0, leads to

$$X_{n+1} = c\sum_{i=0}^{n}\rho^i + \rho^{n+1}X_0 + \sum_{i=0}^{n}\rho^{n-i}Y_{i+1}. \tag{2.36}$$

By evaluating the limit of (2.36) as $n \to \infty$, it is evident that we should have $|\rho| < 1$ to ensure convergence to a stationary distribution. In such a case, the first term on the

right-hand side is a convergent geometric series in ρ and the contribution of the second term is negligible as $n \to \infty$. In the limit, the resulting form of the process is thus given by

$$X_{n+1} = \frac{c}{1-\rho} + \sum_{i=0}^{n} \rho^{n-i} Y_{i+1}. \tag{2.37}$$

The linear combination of independent normal variables Y_n results in a normal distribution with a mean of zero. Therefore, as $n \to \infty$ the AR(1) is described by a normal distribution with mean

$$\lim_{n\to\infty} \mathbb{E}[X_{n+1}] = \frac{c}{1-\rho} \tag{2.38}$$

and variance

$$\lim_{n\to\infty} \mathrm{Var}(X_{n+1}) = \lim_{n\to\infty} \sum_{i=0}^{n} \rho^{2(n-i)} \mathrm{Var}(Y_{i+1}) = \lim_{n\to\infty} \varphi^2 \sum_{i=0}^{n} \rho^{2(n-i)} = \frac{\varphi^2}{1-\rho^2}. \tag{2.39}$$

Following [23], we use known properties of the covariance and (2.35) to determine the covariance between two AR(1) RVs k iterations apart as

$$\mathrm{Cov}(X_{n+k}, X_n) = \mathrm{Cov}(c + \rho X_{n+k-1} + Y_{n+k}, X_n) = \rho \mathrm{Cov}(X_{n+k-1}, X_n). \tag{2.40}$$

By continuing with this recursion, we are able to expand the covariance with respect to k as

$$\mathrm{Cov}(X_{n+k}, X_n) = \rho\, \mathrm{Cov}(X_{n+k-1}, X_n) = \cdots = \rho^k\, \mathrm{Cov}(X_n, X_n), \tag{2.41}$$

leading to the asymptotic variance

$$\sigma_A^2 = \frac{\varphi^2}{1-\rho^2} \left(1 + 2 \sum_{k=1}^{\infty} \rho^k \right) = \frac{\varphi^2}{1-\rho^2} \frac{1+\rho}{1-\rho} = \frac{\varphi^2}{(\rho-1)^2}. \tag{2.42}$$

From (2.16), we thus have that

$$S_n = \frac{1}{n} \sum_{i=1}^{n} X_i \tag{2.43}$$

is approximated by a normal distribution with mean $\frac{c}{1-\rho}$ and variance $\frac{1}{n}\frac{\varphi^2}{(\rho-1)^2}$, when n is large but finite.

2.3.3 Metropolis algorithm

The third example that we discuss is a general Markov chain, called the Metropolis algorithm, used in simulations to sample a given distribution π, called the target distribution [29]. The algorithm starts with an initial (random) value for X_0 and proceeds to generate the subsequent states as follows:

1. For the n-th iteration, a new state X'_n is proposed as

$$X'_n = X_{n-1} + \delta X, \tag{2.44}$$

where δX is sampled from a proposal distribution q which is symmetric. The variance of the proposal distribution governs the size of the proposed move and influences the efficiency of the algorithm (for a detailed discussion, see [30]).

2. The proposed move is accepted or rejected according to the acceptance probability, defined as

$$P(X_{n-1} \to X'_n) = P(X'_n|X_{n-1}) = \min\left(1\,,\, \frac{\pi(X'_n)}{\pi(X_{n-1})}\right). \tag{2.45}$$

3. If the move is accepted, then X_{n-1} is updated to $X_n = X'_n$; otherwise the value is left unchanged that is, $X_n = X_{n-1}$.

It can be checked that the transition probability (2.45) defines an ergodic Markov chain, whose stationary distribution is the target distribution π.

The Metropolis algorithm is widely used to estimate the ergodic expectation $\mathbb{E}_\pi[g(X)]$ of a function g of X by computing the sample mean estimator $S_n(g)$ for n large enough. The asymptotic variance can also be obtained in principle from the covariance, following (2.23); however, this depends in general on the target distribution $\pi(\cdot)$ and choice of distribution q for δX. As a result, we are unable to theoretically determine the asymptotic variance. Other methods can be used in practice, which will be discussed in Chap. 3.

2.4 Markov diffusions

The second class of Markov processes that we consider are continuous-state and continuous-time processes defined by a stochastic differential equation (SDE) of the form

$$dX_t = F(X_t)dt + \sigma dW_t, \tag{2.46}$$

where

- $X_t \in \mathbb{R}^n$ is the state of the process a time t;
- $t \in [0, T]$ is the continuous time index;
- $F : \mathbb{R}^n \to \mathbb{R}^n$ is the force or drift of the process, describing the deterministic evolution of X_t in the absence of noise ($\sigma = 0$);
- $W_t \in \mathbb{R}^m$ is a vector of independent Brownian motions, whose increments dW_t are distributed according to $\mathcal{N}(0, dt)$. This term is responsible for the stochastic nature of the evolution of X_t. In general, the dimensions of X_t and W_t may differ;
- $\sigma : \mathbb{R}^n \to \mathbb{R}^n \times \mathbb{R}^m$ is the noise matrix that controls the variability or amplitude of the noise present in the process. This matrix causes the dimensions of X_t and σW_t to agree, resulting in valid vector addition.

For simplicity, we consider the case where X_t and W_t have the same dimensions, i.e. $n = m$. Additionally, we assume that σ depends on neither time nor X_t. The process satisfies the Markov property as the evolution of X_t only depends on the transformation F of the current state X_t to which noise is added. This is evident if we write the SDE (2.46) as an evolution equation for a discrete time step Δt as

$$X_{n+1} = X_n + F(X_n)\Delta t + \sigma\sqrt{\Delta t}Z, \tag{2.47}$$

where $X_n = X_{\Delta t\, n}$, with a slight abuse of notation, and $Z \sim \mathcal{N}(0,1)$. This recursive equation is called the Euler-Maruyama scheme and is used to numerically approximate the solution of a SDE. From (2.47), it is clear that a diffusion process discretized in time can be seen as a discrete-time Markov chain with a continuous state space.

The analog of $\pi^{(n)}$ for a diffusion X_t is the probability density $p(X_t = x) = p(x,t)$, which is known to evolve in time according to the Fokker-Planck equation given by

$$\frac{\partial}{\partial t}p(x,t) = -\nabla \cdot (F(x)p(x,t)) + \frac{1}{2}\nabla \cdot D\nabla p(x,t), \tag{2.48}$$

starting from an initial probability density $p(x,0)$ [31]. This is the diffusion analog of the evolution equation (2.10) for $\pi^{(n)}$. Here the matrix $D = \sigma\sigma^T$ is symmetric and is called the covariance matrix.

Since the Fokker-Planck equation is linear, we can express it as

$$\frac{\partial}{\partial t}p(x,t) = \mathcal{L}^\dagger p(x,t) \tag{2.49}$$

in terms of a Fokker-Planck operator $\mathcal{L}^\dagger$ given by

$$\mathcal{L}^\dagger = -\nabla \cdot F + \frac{1}{2}\nabla \cdot D\nabla. \tag{2.50}$$

The dual $\mathcal{L}$ of $\mathcal{L}^\dagger$, with respect to the inner product

$$\langle f, p\rangle = \int_{\mathbb{R}^n} f(x)p(x)d^n x, \tag{2.51}$$

is called the generator of the process X_t and is expressed as

$$\mathcal{L} = F \cdot \nabla + \frac{1}{2}\nabla \cdot D\nabla. \tag{2.52}$$

The generator is similar to the transition matrix of a Markov chain as it determines the evolution of expectations according to

$$\frac{\partial}{\partial t}\mathbb{E}[g(X_t)] = \mathbb{E}[(\mathcal{L}g)(X_t)]. \tag{2.53}$$

The continuous-time analog of the transition matrix is in fact the conditional probability density

$$P_t(x,y) = p(X_{s+t} = y, X_s = x) \tag{2.54}$$

which specifies the probability of transitioning from state x to y in t time. This conditional probability density is also called the propagator and can be expressed in operator form as

$$P_t(x,y) = \left(e^{\mathcal{L}t}\right)(x,y). \tag{2.55}$$

For more information on the properties of $\mathcal{L}$ and $\mathcal{L}^\dagger$, see [32].

As for a Markov chain, a diffusion process can be ergodic and thus have a unique stationary probability density π such that

$$\lim_{t\to\infty} p(x,t) = \pi(x) \tag{2.56}$$

is satisfied from any initial probability density $p(x,0)$ for X_0. Since π is time independent, it must satisfy

$$\mathcal{L}^\dagger \pi(x) = 0, \tag{2.57}$$

which is the continuous-time analog of (2.12).

An SDE in $\mathbb{R}^n$ has a stationary distribution in the particular case where the drift $F(x)$ can be written as the gradient of a potential function, that is

$$F(x) = -\nabla U(x) \tag{2.58}$$

and σ is proportional to the identity matrix, i.e., $\sigma = \epsilon I$. These processes are called gradient SDEs and have a stationary distribution known to be the Gibbs distribution, given by

$$\pi(x) = c \exp\left[-\frac{2U(x)}{\epsilon^2}\right], \tag{2.59}$$

where c is a normalization constant. This can be verified by solving (2.57) using the gradient drift in (2.58). The equation given in (2.57), which is often called the stationary Fokker-Planck equation, has the form of an eigenvalue problem where $\pi(x)$ is the eigenfunction with eigenvalue 0.

2.5 Additive functionals of Markov diffusions

The principles discussed in Sec. 2.2 regarding additive functionals can be generalised to continuous-time and continuous-space process, where the summations over n are replaced by integrals over a time T. Thus an additive functional or observable for X_t can be written as

$$S_T(g) = \frac{1}{T}\int_0^T g(X_t)dt. \tag{2.60}$$

Assuming that the process is ergodic, we have that $S_T(g)$ converges in probability to the stationary expectation, denoted as before as

$$\pi(g) = \mathbb{E}_\pi[g(X)] = \int_{\mathbb{R}^n} g(x)\pi(x)dx. \tag{2.61}$$

As for Markov chains, this convergence is known as the ergodic theorem [2] and is mathematically expressed as

$$\lim_{T\to\infty} P(|S_T(g) - \pi(g)| > \varepsilon) = 0 \tag{2.62}$$

for all $\varepsilon > 0$. The central limit theorem associated with additive functionals of a Markov chain can be similarly restated for $S_T(g)$ as

$$S_T(g) \xrightarrow{\text{dist}} \mathcal{N}\left(\pi(g), \frac{\sigma_A^2}{T}\right) \tag{2.63}$$

as $T \to \infty$, assuming that $\pi(g) < \infty$ [13]. To be more precise,

$$\lim_{T\to\infty} P\left(\frac{TS_T(g) - T\pi(g)}{\sqrt{T}\sigma_A}\right) = \mathcal{N}(0,1). \tag{2.64}$$

As before, we are interested in the asymptotic variance σ_A^2 and we can rewrite these results in terms of $\bar{g} = g - \pi(g)$. This re-centers the real-valued function with respect to the stationary distribution, such that

$$S_T(\bar{g}) \xrightarrow{\text{dist}} \mathcal{N}\left(0, \frac{\sigma_A^2}{T}\right) \tag{2.65}$$

as $T \to \infty$ and

$$\lim_{T\to\infty} P\left(\frac{TS_T(\bar{g})}{\sqrt{T}\sigma_A}\right) = \mathcal{N}(0,1). \tag{2.66}$$

Similarly to Sec. 2.2, the asymptotic variance for a general Markov diffusion process can be written as

$$\begin{aligned}\sigma_A^2 &= \lim_{T\to\infty} \frac{1}{T}\text{Var}\left(\int_0^T g(X_s)ds\right) \\ &= \lim_{T\to\infty} \frac{1}{T}\int_0^T\int_0^T \text{Cov}(g(X_u), g(X_v))dudv,\end{aligned} \tag{2.67}$$

where $u \neq v$ and $\text{Var}(\cdot)$ and $\text{Cov}(\cdot)$ are the continuous expressions of (2.20) and (2.21), respectively. As before, the double integral can be simplified such that

$$\sigma_A^2 = \lim_{T\to\infty} \frac{2}{T}\int_{0\le u<v\le T} \text{Cov}(g(X_u), g(X_v))dudv. \tag{2.68}$$

If the process is stationary, the result simplifies to

$$\sigma_A^2 = 2\int_0^\infty \text{Cov}_\pi(g(X_0), g(X_t))dt \tag{2.69}$$

and equivalently in terms of $\bar{g}$

$$\sigma_A^2 = 2\int_0^\infty \mathbb{E}_\pi[\bar{g}(X_0)\bar{g}(X_t)]dt. \tag{2.70}$$

2.6 Asymptotic variance using large deviation theory

We have shown that the asymptotic variance can be obtained from the knowledge of the covariance function. We now discuss another method for obtaining the asymptotic variance of additive functionals of Markov processes, based on the theory of large deviations [25, 26, 27]. The method is described for diffusions, but it is also applicable to Markov chains. Thus, we consider the observable $S_T(g)$ applied to an ergodic diffusion process X_t, defined by the SDE (2.46) with drift F and noise matrix σ. Note that, for notation purposes, we simplify the representation of the observable $S_T(g)$ given in (2.60) to S_T.

The method is based on the basic observation of large deviation theory, which is that the probability density of S_T, written as $p(S_T = s)$, in many cases has the asymptotic form

$$p(S_T = s) \approx e^{-TI(s)} \tag{2.71}$$

as $T \to \infty$ or, in practice, when $T \gg 1$. This approximation is referred to as the *large deviation principle* (LDP) and means that the dominant contribution to $p(S_T = s)$ is a decaying exponential in T. The function $I(s)$, which governs the rate of decay of $p(S_T = s)$, is called the rate function and is expressed as

$$I(s) = \lim_{T\to\infty} -\frac{1}{T} \ln p(S_T = s). \tag{2.72}$$

For an ergodic Markov process with unique stationary distribution π, there exists a unique s^* such that $I(s^*) = 0$. In general, $I(s) \geq 0$ and this means by normalization of $p(S_T = s)$ that

$$s^* = \lim_{T\to\infty} \mathbb{E}[S_T] = \int_{\mathcal{X}} \pi(x) g(x)\, dx. \tag{2.73}$$

Thus the ergodic value of S_T is the zero s^* of the rate function $I(s)$.

The Gärtner-Ellis theorem is typically used to obtain the rate function and is based on the scaled cumulant generating function (SCGF), defined as

$$\lambda(k) = \lim_{T\to\infty} \frac{1}{T} \ln \mathbb{E}[e^{TkS_T}]. \tag{2.74}$$

The theorem states that, if $\lambda(k)$ exists for $k \in \mathbb{R}$ and is differentiable in k, then S_T satisfies a large deviation principle (2.71) with the rate function given as the Legendre-Fenchel transform of $\lambda(k)$:

$$I(s) = \sup_{k\in\mathbb{R}} \{ks - \lambda(k)\}. \tag{2.75}$$

Rate functions obtained via the Gärtner-Ellis theorem are convex. In many cases, $\lambda(k)$ is strictly convex, in which case (2.75) reduces to the Legendre transform

$$I(s) = k(s)s - \lambda(k(s)), \tag{2.76}$$

where $k(s)$ is the unique solution of $\lambda'(k) = s$ [33].

The SCGF has the following properties at $k = 0$:

- From the definition of the SCGF (2.74) and the fact that probability distributions are normalized (i.e. $\mathbb{E}[1] = 1$), we have $\lambda(0) = 0$.
- The expected value of S_T can be expressed as

$$\lambda'(0) = \lim_{T\to\infty} \frac{\mathbb{E}[S_T\, e^{TkS_T}]}{\mathbb{E}[e^{TkS_T}]}\bigg|_{k=0} = \lim_{T\to\infty} \mathbb{E}[S_T] = s^*. \tag{2.77}$$

- The asymptotic variance follows from taking the derivative of $\lambda'(k)$ by means of the quotient rule. Here

$$\lambda''(0) = \lim_{T\to\infty} T(\mathbb{E}[S_T^2] - \mathbb{E}[S_T]^2) = \lim_{T\to\infty} \frac{1}{T} \mathrm{Var}(S_T) = \sigma_A^2. \tag{2.78}$$

We are particularly interested in the relation between the rate function and σ_A^2 when both $\lambda(k)$ and $I(s)$ and are strictly convex. Then the Legendre transform (2.76) implies

$$I''(s^*) = \frac{1}{\lambda''(0)} = \frac{1}{\sigma_A^2}. \tag{2.79}$$

For a detailed discussion of the properties of the SCGF and the rate function, see Sec. 3.5 of [25].

Next, we turn our attention to calculating the SCGF of ergodic Markov processes. It can be shown, by using the Feynman-Kac formula [26], that $\lambda(k)$ corresponds to the dominant eigenvalue of a linear operator $\mathcal{L}_k$, called the tilted generator, given by

$$\mathcal{L}_k = F \cdot \nabla + \frac{1}{2}\nabla \cdot D\nabla + kg, \tag{2.80}$$

where g is the test function that appears in the additive functional $S_T(g)$, as defined in (2.60). The dual $\mathcal{L}_k^\dagger$ of $\mathcal{L}_k$ with respect to the inner product shown in (2.51) is

$$\mathcal{L}_k^\dagger = -\nabla \cdot F + \frac{1}{2}\nabla \cdot D\nabla + kg. \tag{2.81}$$

In general, these operators are not Hermitian, due to $(\nabla \cdot F)^\dagger = -F \cdot \nabla$ being anti-symmetric. This complicates the spectral problem associated with calculating $\lambda(k)$. In fact, to find the dominant eigenvalue, one has to consider both the direct eigenvalue problem

$$\mathcal{L}_k r_k(x) = \lambda(k) r_k(x), \tag{2.82}$$

where $r_k(x)$ is the "right" eigenfunction and the dual eigenvalue problem

$$\mathcal{L}_k^\dagger l_k(x) = \lambda(k) l_k(x), \tag{2.83}$$

where $l_k(x)$ is the "left" eigenfunction. These two eigenfunctions must follow

$$r_k(x) l_k(x) \to 0 \tag{2.84}$$

as $|x| \to 0$ and must decay sufficiently fast that $l_k r_k$ is integrable in $\mathbb{R}^n$. In practice, we impose

$$\int_{\mathbb{R}^n} r_k(x) l_k(x)\, d^n x = 1 \tag{2.85}$$

as well as

$$\int_{\mathbb{R}^n} l_k(x)\, d^n x = 1. \tag{2.86}$$

The condition (2.84) basically follows from the duality between $\mathcal{L}_k$ and $\mathcal{L}_k^\dagger$, which is equivalent to performing integration by parts, where $r_k(x)l_k(x)$ is a boundary term. For more details see [26].

Observing (2.82) and (2.83) for $k = 0$, we note that $\mathcal{L}_{k=0}^\dagger = \mathcal{L}^\dagger$, so that $l_{k=0} = \pi$ and $\mathcal{L}_{k=0} = \mathcal{L}$, so that $r_{k=0} = 1$. This leads to both normalization conditions reducing to the same expression, namely, $\int_{\mathbb{R}^n} \pi(x)\, d^n x = 1$.

2.7 Symmetrization

As mentioned in the previous section, finding the SCGF is complicated due to the fact that the tilted generator $\mathcal{L}_k$ is, in general, not Hermitian. However, $\mathcal{L}_k$ can be transformed to a Hermitian operator by performing a unitary transformation in the case where $\mathcal{L}_k$ has a real spectrum, which is the case when an SDE is gradient. This simplification is called a symmetrization and produces an operator given by

$$\mathcal{H}_k = \pi^{1/2}\mathcal{L}_k\pi^{-1/2}, \tag{2.87}$$

where $\pi(x)$ is the Gibbs distribution shown in (2.59). If one substitutes $\mathcal{L}_k$ with (2.80) and replace $F = -\nabla U$, we see that $\mathcal{H}_k$ has the form

$$\mathcal{H}_k = \frac{\sigma^2}{2}\Delta - V_k, \tag{2.88}$$

where $\Delta = \nabla^2$ represents the Laplacian and

$$V_k(x) = \frac{|\nabla U(x)|^2}{2\sigma^2} - \frac{\Delta U(x)}{2} - kg(x) \tag{2.89}$$

is an effective potential. Since $\mathcal{H}_k$ and $\mathcal{L}_k$ are unitarily related, the dominant eigenvalue $\lambda(k)$ is the same. Therefore, the complicated spectral problem shown in (2.82) reduces to

$$\mathcal{H}_k\psi_k = \lambda(k)\psi_k, \tag{2.90}$$

where the eigenfunction ψ_k is related to r_k and l_k by

$$\psi_k(x) = \pi(x)^{1/2}r_k(x) \quad \text{and} \quad \psi_k(x) = \pi(x)^{-1/2}l_k(x). \tag{2.91}$$

This simplifies the boundary condition of the eigenvalue problem shown in (2.85) to

$$\int_{\mathbb{R}^n} \psi_k(x)^2 \, d^n x = 1. \tag{2.92}$$

2.8 Examples

We demonstrate the ideas developed in the previous sections by calculating the asymptotic variance for different examples. Firstly, we revisit the two-state Markov chain to confirm the calculated asymptotic variance, expressed in (2.33). Thereafter we determine the asymptotic variance of two dynamical observables for a simple Markov linear diffusion, called the Ornstein-Uhlenbeck process.

2.8.1 Two-state Markov chain

The first example that we consider is a special case of the two-state Markov chain, described in Sec. 2.3.1, obtained for $\alpha = \beta$, so that the transition matrix is

$$\mathbf{P} = \begin{pmatrix} 1-\alpha & \alpha \\ \alpha & 1-\alpha \end{pmatrix}, \tag{2.93}$$

where $0 \leq \alpha \leq 1$. It can be shown from (2.12) that, if $\alpha \neq 0, 1$, then the Markov chain $X_1, X_2, \ldots, X_n$ is ergodic and has a unique stationary distribution given by

$$\pi = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \end{pmatrix}. \tag{2.94}$$

As before, we are interested in the sample mean

$$S_n = \frac{1}{n} \sum_{i=1}^{n} X_i \tag{2.95}$$

of the Markov chain. To calculate the expected value of S_n and the asymptotic variance associated with the estimate, we turn our attention to the SCGF. For a Markov chain with discrete time and discrete state space, the SCGF $\lambda(k)$ translates to the natural logarithm of the dominant eigenvalue associated with the tilted transition matrix

$$\mathbf{P}_k = \begin{pmatrix} 1-\alpha & \alpha \\ \alpha e^k & (1-\alpha)e^k \end{pmatrix}, \tag{2.96}$$

as discussed in [25]. The dominant eigenvalue of this matrix can be explicitly calculated and is given as

$$\lambda(k) = \ln \left[\frac{1}{2} \left((1+e^k)(1-\alpha) + \sqrt{(\alpha + \alpha e^k - e^k - 1)^2 - 4\left(e^k - 2\alpha e^k\right)} \right) \right]. \tag{2.97}$$

From this result, we then follow (2.77) and use (2.97) to determine the expected value as

$$\mathbb{E}_\pi[X] = \lim_{n\to\infty} \mathbb{E}[S_n] = \lambda'(0) = \frac{1}{2}. \tag{2.98}$$

To calculate the asymptotic variance we determine the second derivative of (2.97) and set $k = 0$, as stated in (2.78). This results in

$$\sigma_A^2 = \frac{1-\alpha}{4\alpha}. \tag{2.99}$$

These results agree with the theoretical results obtained in Sec. 2.3.1, when $\alpha = \beta$.

2.8.2 Ornstein-Uhlenbeck process with linear observable

The one-dimensional Ornstein-Uhlenbeck process satisfies the SDE

$$dX_t = -\gamma X_t \, dt + \sigma \, dW_t, \tag{2.100}$$

where $X_t \in \mathbb{R}$, $W_t \in \mathbb{R}$, $\gamma > 0$ represents a friction constant and $\sigma > 0$ is the noise amplitude. This is a gradient SDE with force $F(x) = -\gamma x$ derived from the quadratic potential $U(x) = \gamma x^2/2$, so that its stationary distribution is a Gibbs distribution (2.59) given by

$$\pi(x) = \sqrt{\frac{\gamma}{\pi\sigma^2}} \exp\left[-\frac{\gamma x^2}{\sigma^2} \right]. \tag{2.101}$$

For this process, we are interested in the asymptotic variance of the additive functional S_T defined as

$$S_T = \frac{1}{T}\int_0^T X_t\, dt, \tag{2.102}$$

which corresponds to the area per unit time under the trajectory $(X_t)_{t=0}^T$. The tilted generator for the considered gradient process and observable is

$$\mathcal{L}_k = -\gamma x\frac{d}{dx} + \frac{\sigma^2}{2}\frac{d^2}{dx^2} + kx. \tag{2.103}$$

Although the operator is not Hermitian, it is known that it has a real spectrum [26]. Therefore, we can use the symmetrization to determine the Hermitian operator

$$\mathcal{H}_k = \frac{\sigma^2}{2}\frac{d^2}{dx^2} - V_k(x), \tag{2.104}$$

where

$$V_k(x) = \frac{\gamma^2 x^2}{2\sigma^2} - \frac{\gamma}{2} - kx \tag{2.105}$$

is the effective potential.

The spectral problem (2.90) for the operator $\mathcal{H}_k$ shown in (2.104) resembles, up to a sign, the time-independent Schrödinger equation of the one-dimensional quantum harmonic oscillator, where the potential is shifted [26]. The shift is determined by finding the x-value where the minimum occurs and the potential is equivalently expressed as

$$V_k(x) = \frac{\gamma^2}{2\sigma^2}(x - x^*)^2 - \frac{\gamma}{2} - \frac{k^2\sigma^2}{2\gamma^2}, \tag{2.106}$$

where the minimum potential $x^* = k\sigma^2/\gamma^2$. The harmonic oscillator has known eigenvalues $\lambda_n(k)$ and eigenfunctions $\psi_k^{(n)}$ [34]. By identifying $\gamma = \hbar\omega$ and $\sigma = \hbar/\sqrt{m}$ and applying the shift $x \to x + x^*$ to the quantum problem we find

$$\lambda_n(k) = \frac{k^2\sigma^2}{2\gamma^2} - n\gamma \tag{2.107}$$

for $n = 0, 1, 2, \ldots$ and

$$\psi_k^{(n)} = \frac{1}{\sqrt{2^n n!}}\left(\frac{\gamma}{\pi\sigma^2}\right)^{1/4}\exp\left[\frac{-\gamma(x-x^*)^2}{2\sigma^2}\right]H_n\left(\frac{\sqrt{\gamma}}{\sigma}(x-x^*)\right), \tag{2.108}$$

where H_n are the Hermite polynomials. The dominant eigenvalue of the oscillator is analogous to its ground state energy ($n = 0$). Consequently, the SCGF is given by

$$\lambda(k) = \lambda_0(k) = \frac{k^2\sigma^2}{2\gamma^2} \tag{2.109}$$

as shown in Fig 2.2 (left). The eigenfunction corresponding to $\lambda(k)$ is

$$\psi_k(x) = \psi_n^{(0)}(x) = \left(\frac{\gamma}{\pi\sigma^2}\right)^{1/4}\exp\left[-\frac{\gamma(x - k\sigma^2/\gamma^2)^2}{2\sigma^2}\right]. \tag{2.110}$$

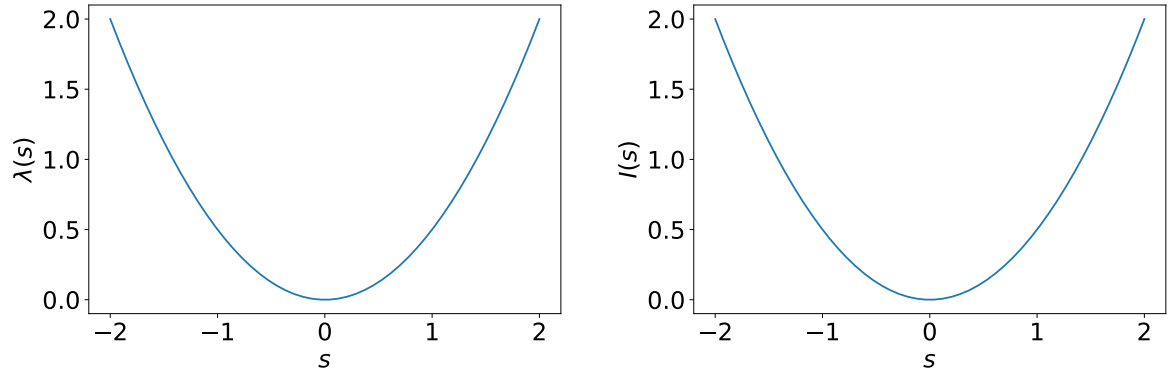


Figure 2.2: (Left) SCGF of a linear observable S_T for the Ornstein-Uhlenbeck process with $\gamma = 1$ and $\sigma = 1$. (Right) Rate function $I(s)$ of S_T.

From this result, we compute the eigenfunctions $r_k(x)$ and $l_k(x)$ of the SCGF by substituting (2.110) and (2.101) into the expressions (2.91). After normalizing the expressions according to (2.85) and (2.86), we then find

$$r_k(x) = \exp\left[\frac{kx}{\gamma} - \frac{3\sigma^2 k^2}{4\gamma^3}\right] \tag{2.111}$$

and

$$l_k(x) = \sqrt{\frac{\gamma}{\pi\sigma^2}}\exp\left[-\frac{\gamma\left(2x - \sigma^2 k/\gamma^2\right)^2}{4\sigma^2}\right]. \tag{2.112}$$

We observe that the product between $r_k(x)$ and $l_k(x)$ decays to zero as $|x| \to \infty$, which agrees with (2.84). Additionally, $r_0(x) = 1$ and $l_0(x) = \pi(x)$, as mentioned in Sec. 2.6.

The SCGF (2.109) is differentiable and strictly convex for all $k \in \mathbb{R}$, as illustrated in Fig. 2.2 (left). Therefore, the calculation of the rate function $I(s)$ reduces to a Legendre transform according to (2.76), leading to

$$I(s) = \frac{\gamma^2 s^2}{2\sigma^2}. \tag{2.113}$$

The rate function is parabolic, as shown in Fig. 2.2 (right), which indicates that S_T has Gaussian fluctuations around $s^* = 0$. The obtained value for s^* is consistent with the property (2.77), where

$$\lim_{T\to\infty} \mathbb{E}[S_T] = \mathbb{E}_\pi[X] = 0 \tag{2.114}$$

with π given in (2.101). Following (2.78), we determine the asymptotic variance of the observable as

$$\sigma_A^2 = \lambda''(0) = \frac{1}{I''(0)} = \frac{\sigma^2}{\gamma^2}, \tag{2.115}$$

which determines the width of the probability distribution of S_T according to the relation displayed in (2.63).

This result is consistent with the fact that the covariance function of the Ornstein-Uhlenbeck process can be described as

$$R(t) = \mathrm{Cov}_\pi(X_0, X_{0+t}) = \frac{\sigma^2}{2\gamma} e^{-\gamma t} \tag{2.116}$$

for $t \geq 0$ [35]. Following (2.69), we are able to theoretically calculate the asymptotic variance as

$$\sigma_A^2 = 2\int_0^\infty \frac{\sigma^2}{2\gamma} e^{-\gamma t} = \frac{\sigma^2}{\gamma^2}, \tag{2.117}$$

which agrees with (2.115).

2.8.3 Ornstein-Uhlenbeck process with quadratic observable

For our third example, we consider the same Ornstein-Uhlenbeck process X_t as before, but now we shift our interest to the quadratic observable

$$S_T = \frac{1}{T}\int_0^T X_t^2\, dt. \tag{2.118}$$

The stationary distribution π of the process remains the same, given by (2.101). The process, with a quadratic observable, now has a tilted generator of the form

$$\mathcal{L}_k = -\gamma x \frac{d}{dx} + \frac{\sigma^2}{2}\frac{d^2}{dx^2} + kx^2, \tag{2.119}$$

with Hermitian counterpart

$$\mathcal{H}_k = \frac{\sigma^2}{2}\frac{d^2}{dx^2} - V_k(x), \tag{2.120}$$

where the effective potential is given by

$$V_k(x) = x^2\left(\frac{\gamma^2}{2\sigma^2} - k\right) - \frac{\gamma}{2}. \tag{2.121}$$

This potential is equivalent again to the time-independent Schrödinger equation of a quantum harmonic oscillator. To ensure that the spectral problem (2.90) has a well defined spectrum, we must have $\gamma^2/2\sigma^2 \geq k$. The eigenvalues $\lambda_n(k)$ and eigenfunctions $\psi_k^{(n)}$ of the quantum harmonic oscillator translated to our problem are now given by

$$\lambda_n(k) = -\left(n + \frac{1}{2}\right)\sqrt{\gamma^2 - 2k\sigma^2} + \frac{\gamma}{2} \tag{2.122}$$

for $n = 0, 1, 2, \ldots$ and

$$\psi_k^{(n)}(x) = \frac{1}{\sqrt{2^n n!}}\left(\frac{\sqrt{\gamma^2 - 2k\sigma^2}}{\pi\sigma^2}\right)^{1/4} \exp\left[-\frac{\sqrt{\gamma^2 - 2k\sigma^2}}{2\sigma^2}x^2\right] H_n\left(\frac{\sqrt[4]{\gamma^2/\sigma^2 - 2k}}{\sqrt{\sigma}}x\right), \tag{2.123}$$

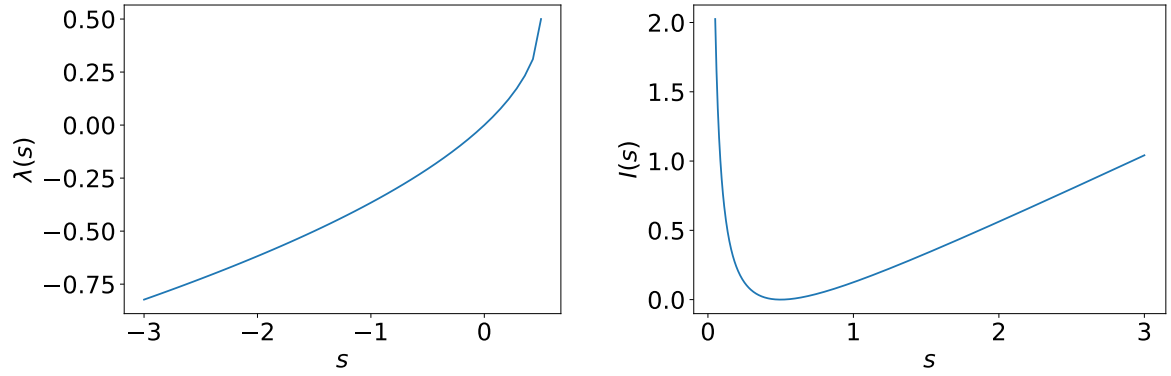


Figure 2.3: (Left) SCGF of a quadratic observable S_T for the Ornstein-Uhlenbeck process with $\gamma = 1$ and $\sigma = 1$. (Right) Rate function $I(s)$ of S_T.

respectively. The dominant eigenvalue, shown in Fig 2.3 (left), is thus

$$\lambda(k) = -\frac{1}{2}\sqrt{\gamma^2 - 2k\sigma^2} + \frac{\gamma}{2} \tag{2.124}$$

with the corresponding normalized eigenfunction

$$\psi_k(x) = \left(\frac{\sqrt{\gamma^2 - 2k\sigma^2}}{\pi\sigma^2}\right)^{1/4} \exp\left[-\frac{\sqrt{\gamma^2 - 2k\sigma^2}}{2\sigma^2}x^2\right]. \tag{2.125}$$

From this result, we find the expressions for $r_k(x)$ and $l_k(x)$, as before, by substituting (2.125) and (2.101) into (2.91) and normalizing according to (2.85) and (2.86). This yields

$$r_k(x) = \left(\frac{2\sqrt{\gamma^2 - 2k\sigma^2}}{\sqrt{\gamma^2 - 2k\sigma^2} + \gamma}\right)^{\frac{1}{2}} \exp\left[-\frac{x^2}{2\sigma^2}\left(\sqrt{\gamma^2 - 2k\sigma^2} - \gamma\right)\right] \tag{2.126}$$

and

$$l_k(x) = \left(\frac{\sqrt{\gamma^2 - 2k\sigma^2} + \gamma}{2\pi\sigma^2}\right)^{\frac{1}{2}} \exp\left[-\frac{x^2}{2\sigma^2}\left(\sqrt{\gamma^2 - 2k\sigma^2} + \gamma\right)\right]. \tag{2.127}$$

As before, the SCGF satisfies the conditions of the Gärtner-Ellis theorem and is strictly convex. This allows us to determine the rate function $I(s)$ via the Legendre transform of (2.124). The resulting rate function is given by

$$I(s) = \frac{\gamma^2}{2s\sigma^2}\left(s - \frac{\sigma^2}{2\gamma}\right)^2 \tag{2.128}$$

for $s > 0$. The rate function is shown in Fig. 2.3 (right) and illustrates that the expected value of the observable is $s^* = \sigma^2/(2\gamma)$, mathematically expressed in (2.77). Observing the graph, we notice that the rate function is linear with gradient $\gamma^2/(2\sigma^2)$ when $s \to \infty$ and

diverges as $1/s$ when $s \to 0$. This causes asymmetrical tails in the probability distribution of S_T; however, we still expect Gaussian fluctuations around s^*. The asymptotic variance is determined using (2.78) as

$$\sigma_A^2 = \lambda''(0) = \frac{1}{I''(1/2)} = \frac{\sigma^4}{2\gamma^3} \tag{2.129}$$

and describes the local Gaussian fluctuations around the concentration point $s^* = \pi(X^2) = \sigma^2/(2\gamma)$ of S_T as $T \to \infty$.

Note that obtaining the asymptotic variance from the covariance using (2.69) is not possible here, because the covariance function in this case is unknown. This emphasizes the advantage of using the large deviation method to determine the asymptotic variance, as we can calculate σ_A^2 without the use of the covariance function.

Chapter 3

Asymptotic variance estimation techniques

This chapter serves as a review of known methods for estimating the asymptotic variance of additive functionals of Markov processes. We present the most conventional methods used in practice, which are based on either the estimation of the covariance function or the segmentation of the additive observable into batch means. These methods will be used thereafter as benchmarks for the new methods that we present in Chapters 4 and 5. Using specific examples of Markov processes, similar to the examples previously considered, we give guidelines and address the weaknesses of each method. The methods that we discuss in this chapter are based on discrete-time Markov processes; however, continuous-time Markov processes are implicitly included, since the time discretization of a time-continuous Markov diffusion process is a Markov chain.

3.1 Covariance estimation

In this section we describe a variance estimation technique based on estimating the covariance [23, 36, 14], which appears in the asymptotic variance formula (2.23). The technique is based on our calculation in Sec. 2.2, which showed that σ_A^2 can be expressed for a stationary Markov chain as

$$\sigma_A^2 = \gamma_0 + 2\sum_{k=1}^{\infty}\gamma_k = \gamma_0(1 + 2\sum_{k=1}^{\infty}c(k)) = \gamma_0\tau. \tag{3.1}$$

Here the k-lag autocovariance of the functionals $g(X_1), g(X_2), \ldots$ is given as

$$\gamma_k = \mathrm{Cov}(g(X_t), g(X_{t+k})) = \mathbb{E}[(g(X_t) - \mathbb{E}_\pi[g(X)])(g(X_{t+k}) - \mathbb{E}_\pi[g(X)])], \tag{3.2}$$

$c(k) = \gamma_k/\gamma_0$ and τ are called the autocorrelation function and correlation time, respectively. Lastly,

$$\gamma_0 = \mathbb{E}[(g(X_t) - \mathbb{E}_\pi[g(X)])^2] \tag{3.3}$$

is referred to as the naïve variance, since it does not take into account the correlations in the state. However, if $g(X_1), g(X_2), \ldots$ are i.i.d., then the naïve variance is the asymptotic variance.

The natural estimator of the autocovariance function γ_k is

$$\hat{\gamma}_k = \frac{1}{n-k}\sum_{i=1}^{n-k}[g(X_i) - \hat{\mu}_n][g(X_{i+k}) - \hat{\mu}_n], \tag{3.4}$$

where the hat indicates that $\hat{\gamma}$ is an empirical estimate of γ_k. Similarly, $\hat{\mu}_n$ is the estimated sample mean given by (1.1), and n is the number of observed values. This leads to an estimator of the asymptotic variance σ_A^2 given as

$$\hat{\sigma}_A^2 = \hat{\gamma}_0 + 2\sum_{k=1}^{\infty}\hat{\gamma}_k. \tag{3.5}$$

Equivalently, we can rewrite this result in terms of the estimated autocorrelation function $\hat{c}(k) = \hat{\gamma}_k/\hat{\gamma}_0$:

$$\hat{\sigma}_A^2 = \hat{\gamma}_0\left(1 + 2\sum_{k=1}^{\infty}\hat{c}(k)\right). \tag{3.6}$$

The autocorrelation function is a dimensionless version of the autocovariance function and takes on values between -1 and 1: a value close 0 is indicative of no correlation, whereas values close to -1 or 1 represents a strong correlation [2].

For a Markov process with a fixed number of observed values n, the estimators $\hat{\gamma}_k$ and $\hat{c}(k)$ are not reliable when k is large. As we consider larger k-lags, the number of sample values contributing to the estimate decreases. When the number of observed values is small, the estimator can not converge. This, in turn, leads to inaccurate estimates of the asymptotic variance.

The first approach to improve the accuracy of $\hat{\sigma}_A^2$ is to improve the convergence of the estimator $\hat{\gamma}_k$. This is done by replacing the denominator of $n-k$ in (3.4) with n [36]. Increasing the denominator to n decreases the contribution of $\hat{\gamma}_k$ for large k and results in more accurate approximations when estimating the asymptotic variance by means of (3.6). The second approach is to discard the inaccurate estimates of $\hat{\gamma}_k$ when k is large by truncating the infinite sum up to a finite integer value M. The resulting estimator is given by

$$\hat{\sigma}_A^2 = \hat{\gamma}_0\left(1 + 2\sum_{k=1}^{M}\hat{c}(k)\right), \tag{3.7}$$

where $1 \leq M \leq n$ and is based on the observation that $\hat{\gamma}_k$, and subsequently $\hat{c}(k)$, tend to zero as $k \to \infty$ [14]. The value M is a hyperparameter (controlled by the practitioner) and is used to set a threshold for when $\hat{\gamma}_k$ is either close to zero or too noisy. This increases the accuracy of $\hat{\sigma}_A^2$; however, determining the optimal value M is a difficult task as (3.7) contains estimates of $\hat{c}(k)$ that randomly fluctuate.

Additionally, to 'smooth' the estimator of the asymptotic variance, we can use some weight function $w(k)$ that further reduces the contribution of $\hat{\gamma}_k$ at large k. This is expressed as

$$\hat{\sigma}_A^2 = w(0)\hat{\gamma}_0 + 2\sum_{k=1}^{M}w(k)\hat{\gamma}_k, \tag{3.8}$$

where $0 \leq w(k) \leq 1$; see [32] for desired properties of $w(k)$.

A variation to the covariance estimation technique is to assume that the autocorrelation function of the observed process decays exponentially as $k \to \infty$ [14]. This can be expressed mathematically as

$$c_{\text{exp}}(k) = e^{-\upsilon kh}, \tag{3.9}$$

where υ is a constant with units time^{-1} and h is the sampling interval of the process. The assumption allows us to estimate the correlation time τ using an exponential function, which leads to the asymptotic variance being approximated by a rescaled naïve variance:

$$\hat{\sigma}_A^2 = \hat{\gamma}_0 \left(1 + 2\sum_{k=1}^{\infty} \hat{c}_{\text{exp}}(k)\right) \approx \hat{\gamma}_0 \hat{\tau}_{\text{exp}}, \tag{3.10}$$

where τ_{exp} is the exponentially approximated correlation time, as in (3.1). The exponential autocorrelation function (3.9) decreases to 0 for large k. Hence, poor estimates of γ_k when k is large does not influence the accuracy of the estimated asymptotic variance, as is the case in (3.6). We can use auto-regression methods to obtain υ, where a higher auto-regression method can lead to better σ_A^2 approximations [14]. Alternatively, we can estimate τ using

$$\hat{\tau} = 1 + 2\sum_{k=1}^{M} \hat{c}(k), \tag{3.11}$$

as done in (3.7).

As suggested in the introduction, we are interested in determining the asymptotic variance of an additive functional by using steady-state analysis. An ideal estimator for this type of analysis would estimate the asymptotic variance in parallel with the process as it unfolds. This allows us to have an estimated value of the asymptotic variance at each time-step. As the process evolves over time, each new observation is used to update the estimated asymptotic variance and we refer to such an estimator as being *online*. The covariance estimation technique requires all the data to estimate the asymptotic variance and therefore cannot be classified as an online estimator.

3.2 Examples

We illustrate the general covariance estimation technique based on (3.1) by calculating the asymptotic variance of the mean area of the Ornstein-Uhlenbeck process and auto-regressive process.

3.2.1 Ornstein-Uhlenbeck process

We consider the Ornstein-Uhlenbeck process, defined in (2.100), with parameters $\gamma = 0.5$ and $\sigma = 2$. We are interested in the linear observable S_T of the process, such that $g(x) = x$. The k-lag autocorrelation function $\hat{c}(k)$ is determined by calculating $\hat{\gamma}_k/\hat{\gamma}_0$ according to (3.4) for $k \geq 0$. As shown in Fig. 3.1 (left), this results in an empirical estimate (blue) of the theoretical autocorrelation function $c(k)$ (green), derived from (2.116). We see that

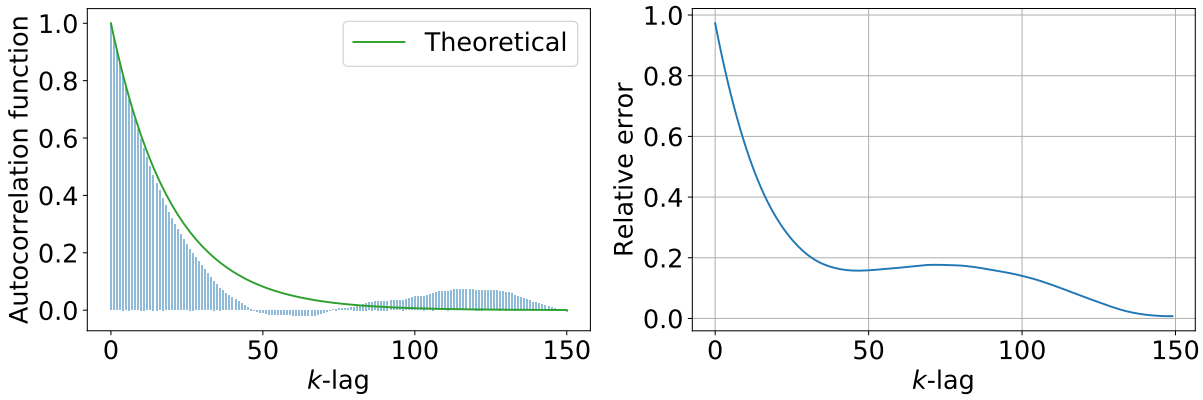


Figure 3.1: (Left) Estimate of the autocorrelation function (blue) for the Ornstein-Uhlenbeck process with $\gamma = 0.5$ and $\sigma = 2$, compared with the theoretical autocorrelation function (green). (Right) Relative error of the empirically calculated asymptotic variance as a function of k-lag.

$\hat{c}(k)$ accurately estimates the theoretical autocorrelation function when k is small. As k increases, the number of observed values for each estimate decreases. This leads to unreliable estimates when k is large, depicted by the fluctuations in the tail of $\hat{c}(k)$.

Next, we follow (3.6) and sum over all the k-lags of $\hat{c}(k)$ in order to estimate the known asymptotic variance (2.115). The relative error (RE) is used to compare the estimated asymptotic variance $\hat{\sigma}_A^2$ with the theoretical asymptotic variance σ_A^2 calculated before in (2.117). The RE gives an indication of the accuracy of the estimator and is determined by the formula

$$\text{RE} = \frac{|\hat{\sigma}_A^2 - \sigma_A^2|}{\sigma_A^2}. \tag{3.12}$$

As previously mentioned by (3.7), in order to improve the accuracy of $\hat{\sigma}_A^2$ we need to truncate the infinite sum present in (3.6) to the Mth lag. Using $\hat{c}(k)$ from Fig. 3.1 (left), one expects that the optimal truncation would occur near $k = 50$. However, for this example, the calculated $\hat{\gamma}_k$ is already unreliable before $k = 50$ and therefore inaccurately represents the theoretical autocorrelation function. Here the optimal M-value would incorporate unreliable estimators and achieve an accurate $\hat{\sigma}_A^2$ when $k = 150$, shown in Fig. 3.1 (right). As the theoretical autocorrelation function is rarely known, determining the optimal M-value from the estimated autocorrelation function is a non-trivial task.

We can improve the quality of the estimator by increasing the total number of observed values for the Ornstein-Uhlenbeck process. However, this solution is not always possible in practice as the number of available observations may be limited.

3.2.2 Auto-regression model AR(1)

We now calculate the asymptotic variance of a linear additive functional of an AR(1) process for which the autocorrelation function decays exponentially. We introduced the process in Sec. 2.3.2 and from the calculations shown in (2.41), we know that the autoco-

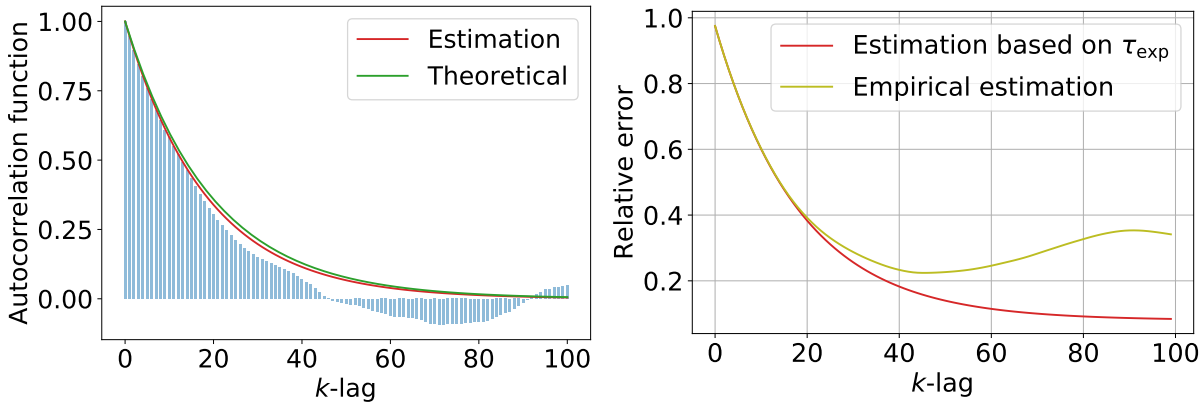


Figure 3.2: (Left) Theoretical autocorrelation function of an auto-regression model AR(1) with $\rho = 0.95$ (green) estimated based on τ_{exp} (red). (Right) Relative error of estimator based on τ_{exp} (red) and the empirical estimator (yellow) as a function of k-lag.

variance function of an AR(1) process is given by

$$\gamma_k = \gamma_0 \rho^k, \tag{3.13}$$

for $c = 0$ and $|\rho| < 1$. Next, we note that the autocorrelation function $c(k)$ is expressed as an exponential in k and following (3.9) we rewrite γ_k as

$$\gamma_k \approx \gamma_0 \, e^{-\upsilon k h}. \tag{3.14}$$

Comparing these results shows that

$$\rho = e^{-\upsilon h} \tag{3.15}$$

and that $\hat{\sigma}_A^2$ converges to a single positive value as $k \to \infty$. For the example, we consider $h = 1$ and estimate ρ using the k-lag autocovariance function, such that

$$\hat{\rho} = \frac{\hat{\gamma}_1}{\hat{\gamma}_0}, \tag{3.16}$$

as shown in (2.40). Once we have $\hat{\rho}$, the asymptotic variance is estimated following

$$\hat{\sigma}_A^2 \approx \hat{\gamma}_0 \tau_{\text{exp}} = \hat{\gamma}_0 \left(1 + 2 \sum_{k=1}^{\infty} \hat{\rho}^k \right) = \hat{\gamma}_0 \frac{1 + \hat{\rho}}{1 - \hat{\rho}}. \tag{3.17}$$

We illustrate these results by estimating the autocorrelation function of an AR(1) process with $\rho = 0.95$. Form the estimations of our example (Fig. 3.2 (left)), we see that the estimator based on an autocorrelation function that decays exponentially (red) is a more accurate representation of the theoretical autocorrelation function (green), than the empirical estimator of $c(k)$ (blue) given by the normalized (3.4). Note that the estimator based on τ_{exp} only depends on $\hat{\gamma}_0$ and $\hat{\gamma}_1$. Therefore, it does not suffer from unreliable

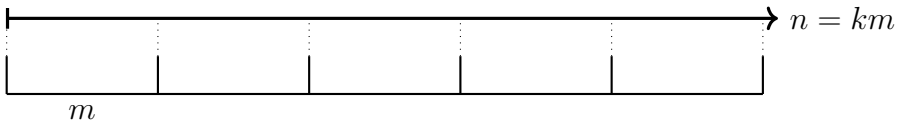


Figure 3.3: Total number of observations n divided into k batches of size m.

values when k is large, because (3.14) tends to zero as $k \to \infty$. However, any inaccuracy in either of these two estimates will have a large influence on the final σ_A^2 estimated.

Figure 3.2 (right) illustrates a comparison between the estimator based on τ_{exp} (red) as shown in (3.10) and the empirical estimator (yellow) as shown in (3.6) for an AR(1) process with $\rho = 0.95$. In order to calculate RE (3.12), we require the theoretical asymptotic variance corresponding to an AR(1) process, which is given by (2.42). From these results, it is again evident that the empirical estimator fluctuates as k increases. This is due to the increased fluctuations of $\hat{\gamma}_k$ when k is large. On the other hand, the estimator based on τ_{exp} does not fluctuate owing to the exponential approximation and this results in a better estimator of σ_A^2 as $k \to \infty$.

3.3 Batch means method

In this section we describe another method for calculating the asymptotic variance, called the non-overlapping batch means (NBM) method [16, 15, 37, 17]. This method is based on the simple observation that a sample mean such as

$$S_n = \frac{1}{n} \sum_{i=1}^{n} g(X_i) \tag{3.18}$$

can be rewritten as

$$S_n = \frac{1}{k} \sum_{i=1}^{k} \bar{Y}_i, \tag{3.19}$$

where

$$\bar{Y}_i = \frac{1}{m} \sum_{j=1}^{m} g(X_{(i-1)m+j}) \tag{3.20}$$

is the ith block or batch sample mean involving m RVs of the Markov chain starting from $X_{(i-1)m+1}$. The segmentation of the whole sequence $X_1, X_2, \ldots, X_n$ into these k blocks is illustrated in Fig. 3.3 and assumes that $n = km$.

The advantage of writing the sample mean in block form is that the batch means $\bar{Y}_i$ are expected to become i.i.d. in the double limit where $n \to \infty$ and $m \to \infty$. This is confirmed by the functional limit theorem [16, 18], which shows that the batch means $\bar{Y}_i$ are uncorrelated and normally distributed as $m \to \infty$ while k is constant. Therefore in this limit, we can express σ_A^2 as the naïve variance of the batch means, leading to the estimator

$$\hat{\sigma}_A^2 = \frac{m}{k} \sum_{i=1}^{k} (\bar{Y}_i - \hat{\mu}_n)^2. \tag{3.21}$$

Note that, if $k \to \infty$ for a constant m and correlated observations, then the naïve variance estimator will converge to an inaccurate estimate.

When the total number of observations n is fixed, both m and k are finite. Therefore, in order to gain either accuracy or consistency there is a trade-off between the batch size and the number of batches, respectively. For highly correlated samples, the estimator requires a large batch size m to decorrelate the samples and to produce an accurate estimate. An increase in batch size leads to a decrease in k the numbers of batches, due to the fixed number of observations n. If there is an insufficient number of batches, then the estimates will become inconsistent. The batch size for which there is a perfect trade-off between accuracy and consistency with respect to error is referred to as the optimal batch size.

Unfortunately, no consensus has yet been reached on the optimal batch size or batch number. Authors have proposed sampling plans where $m = n^{\theta}$ for some $\theta \in (0, 1)$ [38]. Others suggest that the batch size should be related to the autocorrelation time as the observations are effectively decorrelated after τ observations, i.e., $m > \tau$ [39]. Some authors argue that there is little statistical reason to use more than 30 batches and suggest using $k = 30$ as a rule of thumb [36]. In practice, the consequences related to batch correlation are more serious than the consequences of fewer batches, i.e., a large batch size should be prioritized over the number of batches to ensure uncorrelated batch means [23].

Figure 3.4 illustrates the influence of the batch size on the estimated asymptotic variance for the Ornstein-Uhlenbeck process with $\gamma = \sigma = 1$. We calculate the asymptotic variance for a fixed number of batches, while increasing the number of observations. This leads to an increase in the number of observations within each batch. For a reference, the NBM estimator (blue) is compared with the naïve variance estimator (orange). It is clear from the results that the NBM estimator achieves more accurate estimates than the naïve estimation as n is increased. These results are expected as the naïve variance estimator performs poorly when the observations are correlated. The accuracy of the NBM estimator improves as we consider more observations. An increase in n leads to a larger batch size. This in turn reduces the correlation between batches and produces more accurate estimates.

The accuracy of the NBM method can be improved by determining the optimal batch number. Calculating the optimal batch number allows us to demonstrate the inaccuracy caused from the trade-off between the number of batches k and batch size m. For an Ornstein–Uhlenbeck process with $\gamma = 0.1$ and $\sigma = 1$, we determine $\hat{\sigma}_A^2$ for S_T with $g(x) = x$. Following (3.21), we estimate $\hat{\sigma}_A^2$ for an increasing number of batches, while keeping n fixed. The estimation error is calculated by comparing the estimated asymptotic variance at each batch size with the theoretical asymptotic variance, derived in Sec. 2.8.2. Repeating this procedure allows us to calculate the average error that occurs for each batch number. The error function resulting from the chosen parameters has a convex shape (Fig. 3.5). The minimum (or turning point) of the error function is $k = 34$ and represents the optimal batch number. When $k > 34$, the batch size is too small to properly decorrelate the observations and this results in inaccurate NBM estimates. On the other hand, if $k < 34$, there are not enough batches to consistently estimate the asymptotic variance.

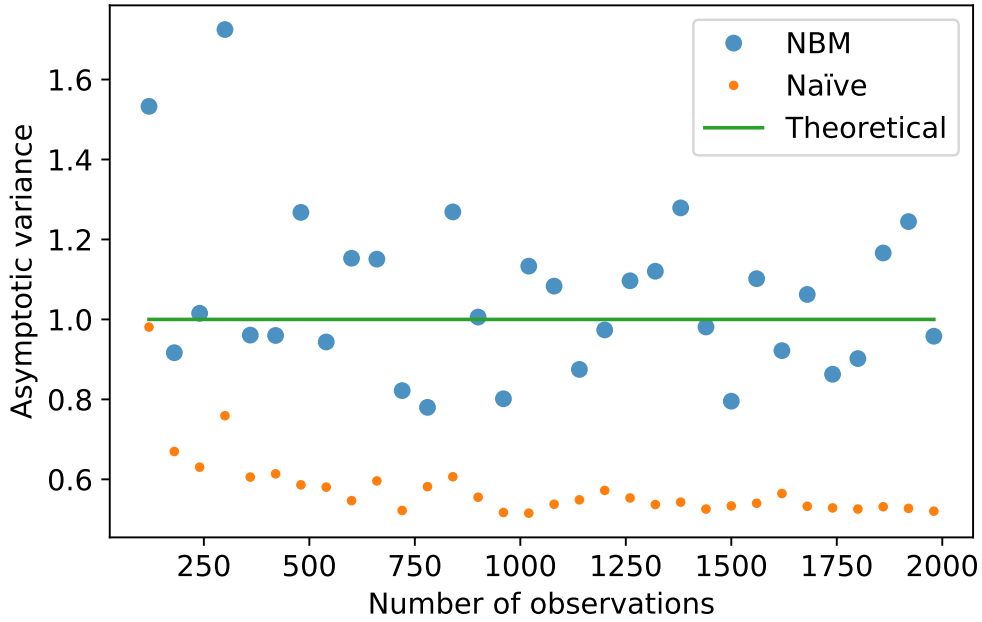


Figure 3.4: Comparison of the naïve variance estimation (orange) and NBM estimator using 50 batches (blue) to approximate the theoretical asymptotic variance (green) of the Ornstein–Uhlenbeck process with $\gamma = 1$ and $\sigma = 1$

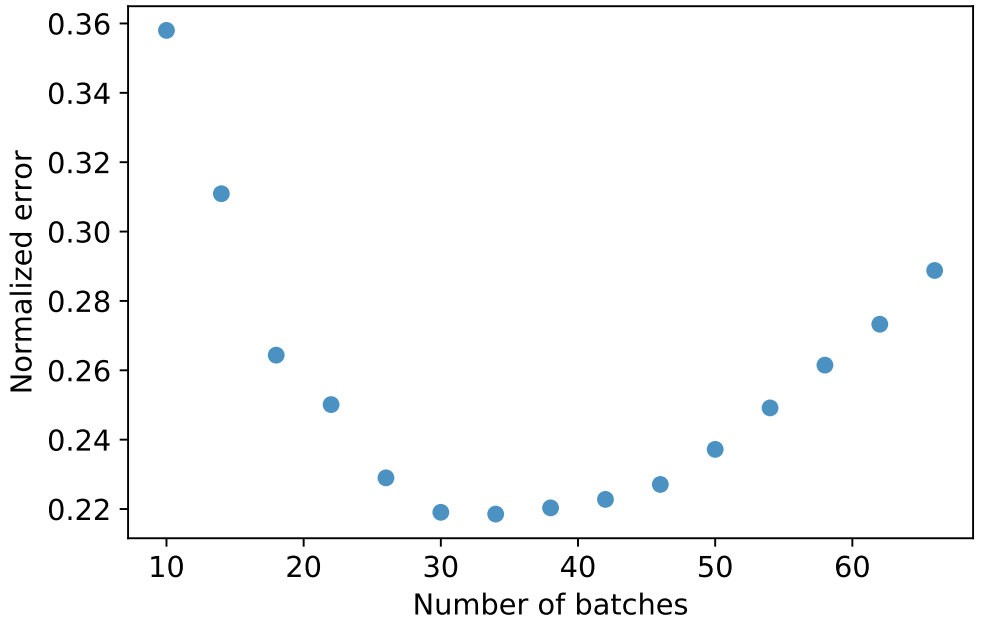


Figure 3.5: The normalized error using the NBM estimator to approximate the asymptotic variance of a Ornstein–Uhlenbeck process with $\gamma = 0.1$ and $\sigma = 1$

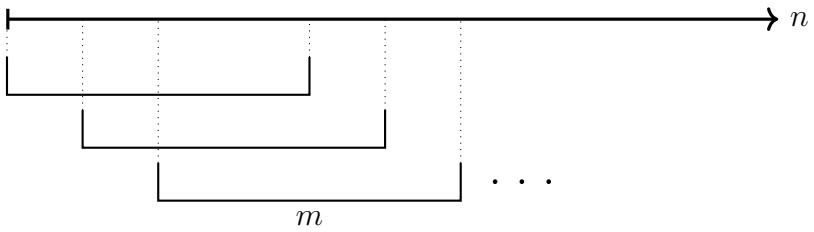


Figure 3.6: Total number of observations n divided into $(n - m + 1)$ overlapping batches each of size m.

The batch means method is a simple yet effective method of estimating the asymptotic variance of a sample mean. However, it is not an online estimator as it requires all the observations of the process during its calculation. Additionally, we need to optimise the batch number to minimize the estimation error. This is only done by recalculating $\hat{\sigma}_A^2$ for different batch sizes, which is computationally expensive. In practice, the optimization is a difficult task as the theoretical variance in usually not known.

3.3.1 Overlapping batch means

Instead of dividing all the observations into adjacent blocks, as considered before and shown in Fig. 3.3, we can divide the observations into overlapping blocks or batches. This variation is called the overlapping batch means method (OBM) and is illustrated in Fig. 3.6, where the total number of observations n is divided into $(n - m + 1)$ overlapping batches each of size m. Assuming that the means of the overlapping batches are uncorrelated, the variance of a sample mean is estimated as

$$\hat{\sigma}_A^2 = \frac{m}{n-m+1} \sum_{j=1}^{n-m+1} (\bar{Y}_j^{(\mathrm{OBM})} - \hat{\mu}_n)^2, \tag{3.22}$$

where the mean of the ith overlapping block $\bar{Y}_i$ is

$$\bar{Y}_i^{(\mathrm{OBM})} = \frac{1}{m} \sum_{j=1}^{m} g(X_{i+j-1}). \tag{3.23}$$

This estimator may seem counter-intuitive as the batches overlap and should therefore be correlated. However, it can be shown that the OBM estimator is approximately equal to the $w(k)$-weighted estimator, shown in (3.8) [37]. Hence, the batch dependence is not a concern even with a large amount of overlap [40].

We study the influence the number of batches has on (3.22) by considering the case where $n \to \infty$ and m is constant. The increase in n results in more batches contributing to the estimate of σ_A^2, which leads to a more reliable estimator. On the other hand, if we keep n constant and increase m, the estimator $\bar{Y}_i^{(\mathrm{OBM})}$ will become more accurate, leading to a more accurate $\hat{\sigma}_A^2$. Therefore, when the number of observations is finite the OBM has the same batch size and batch number trade-off, as discussed before.

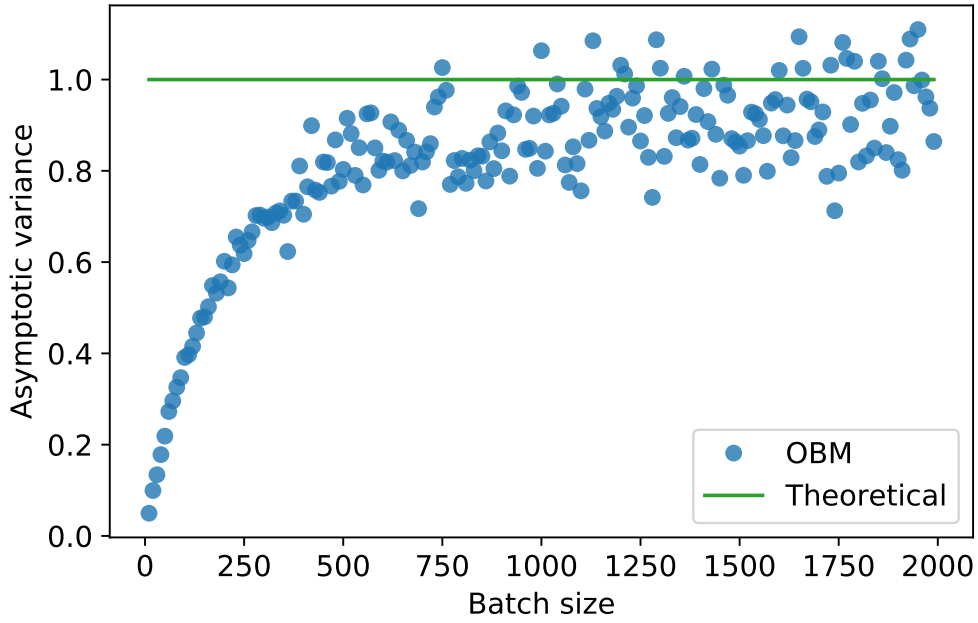


Figure 3.7: Influence of increasing batch sizes on the OBM estimate (blue) of theoretical asymptotic variance (green) for the Ornstein-Uhlenbeck process with $\gamma = 1$ and $\sigma = 1$.

When comparing the NBM and OBM, one can show that the distribution of the OBM estimator has only 2/3 the variance of the NBM estimator, given the batch means are independent and normally distributed [37]. Furthermore, it appears that the OBM estimator is less sensitive to batch size than the NBM estimator [15]. The OBM estimator is theoretically more efficient; however the NBM estimator is computationally more efficient as it requires $O(n)$ computations and $O(1)$ storage, whereas OBM requires $O(n)$ computations and $O(m)$ storage [39]. Hence, the choice of method depends on the correlation between observations. For a process with highly correlated observations, one is inclined to use NBM, the reason being that for highly correlated observations, a larger m is required to decorrelate the batch means and the NBM will be computationally more efficient. However, if the observations are less correlated, one is inclined to use OBM. Here a smaller batch size m will sufficiently decorrelate the observations and one will gain accuracy at the slight expense of computational storage.

We illustrate the influence of batch size on the OBM estimator by estimating the asymptotic variance of S_T with $g(x) = x$ for a fixed number of observations generated by the Ornstein-Uhlenbeck process with $\gamma = \sigma = 1$ (Fig. 3.7). The estimated asymptotic variance is repetitively calculated using (3.22) while increasing the batch size (blue) and these results are compared to the known theoretical variance (green).

When the number of observations is fixed, a small batch size results in a large number of batches. The large number of batches leads to consistent estimates; however, the batch size is too small to accurately estimate the asymptotic variance. This is clear from the results shown in Fig. 3.7 as the OBM estimator consistently produces an inaccurate

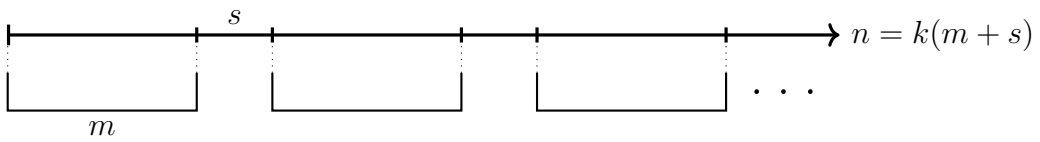


Figure 3.8: Total number of observations n divided into batches of size m spaced s apart.

estimate of the asymptotic variance when the batch sizes are small. As the batch size increases, the number of batches decreases. This causes the estimator's accuracy to increase at the cost of consistency. As the batch size increases, the variance of the normally distributed estimates increases. The increasing variability will ultimately result in inaccurate estimates.

3.3.2 Spaced batch means

Another variation aims to further decrease the batch correlation present in the batch means method by inserting fixed spaces between batches. This variation is called the spaced batch means method (SBM) and is illustrated by Fig. 3.8, where the total number of observations n is divided into batches of size $m - s$ spaced s apart. For simplicity, we consider the case where m divides the total number of observations n exactly.

Only the first $m - s$ observations of each batch contributes to the batch mean, while the remaining s observations are ignored. Therefore the expected value is estimated as

$$\hat{\mu}'_n = \frac{1}{k}\sum_{i=1}^{k} \bar{Y}_j^{(\mathrm{SBM})}, \tag{3.24}$$

where the ith spaced batch mean $\bar{Y}_j^{(\mathrm{SBM})}$ has the form

$$\bar{Y}_i^{(\mathrm{SBM})} = \frac{1}{m-s}\sum_{j=1}^{m-s} g(X_{(i-1)m+j}). \tag{3.25}$$

Similarly to before, we assume that the spaced batch means are uncorrelated and the variance is estimated according to the naïve variance estimator:

$$\hat{\sigma}_A^2 = \frac{m-s}{k}\sum_{j=1}^{k}(\bar{Y}_j^{(\mathrm{SBM})} - \hat{\mu}'_n)^2. \tag{3.26}$$

We note that, for $s = 0$, the SBM estimator reduces to the NBM estimator and we consequently compare the two methods. When s and m are known, the computational requirements of the two methods are estimated to be the same [39]. The addition of spaces to the NBM results in an extra parameter which needs to be specified by the practitioner. Despite the fact that the spaces between blocks lead to more decorrelated batch means, it does not significantly improve the accuracy of $\hat{\sigma}_A^2$ [41]. In practice, it is generally better to use all the observations when estimating the expected value of a process. Therefore, NBM is usually preferred over the SBM, due to the risk of inaccurate estimation resulting from ignored observations.

Chapter 4

Asymptotic variance and the Poisson equation

It is known that the asymptotic variance σ_A^2 of an observable $S_n(g)$ or its centered version $S_n(\bar{g})$ can be related to the solution of a differential equation called the Poisson equation [42, 28, 43]. In this chapter, we present this link, which will serve as the theoretical basis for the new method that we propose in Chap. 5 for estimating σ_A^2.

Considering the cases of Markov chains and diffusions separately, we introduce the Poisson equation and show how its solution determines σ_A^2. We also discuss two representations of this solution in terms of conditional expectation and as the r_k eigenfunction discussed before in connection with large deviations. As before, we illustrate all the results discussed with examples involving the two-state Markov chain and the Ornstein-Uhlenbeck process.

4.1 Continuous-time Poisson equation

We begin our presentation with the case of ergodic diffusions X_t described by the SDE introduced in Sec. 2.4, whose generator, we recall, is

$$\mathcal{L} = F \cdot \nabla + \frac{1}{2} \nabla \cdot D \nabla, \tag{4.1}$$

where $D = \sigma\sigma^T$. For this process, we also consider as before the observable

$$S_T(g) = \frac{1}{T} \int_0^T g(X_t) dt, \tag{4.2}$$

which converges in probability to

$$\pi(g) = \mathbb{E}_\pi[g(X)] = \int_{\mathbb{R}^n} g(x)\pi(x)dx, \tag{4.3}$$

π being the stationary distribution of X_t.

The Poisson equation associated with X_t and g is the differential equation defined as

$$\mathcal{L}\phi(x) = -[g(x) - \pi(g)], \tag{4.4}$$

where $\mathcal{L}$ is the generator of X_t, as introduced in Sec. 2.4, and $\phi(x) : \mathbb{R}^n \to \mathbb{R}$ is the solution of (4.4) that we aim to determine. In terms of $\bar{g} = g - \pi(g)$, we can simply write this equation as

$$\mathcal{L}\phi(x) = -\bar{g}(x). \tag{4.5}$$

The Poisson equation may have more than one solution, when a solution exists at all [42]. We consider here solutions that are square integrable with respect to π, i.e. $\mathbb{E}_\pi[\phi^2] < \infty$, which ensures that ϕ exists and is almost everywhere bounded or slowly increasing [22]. This condition is similar to $\pi(|\phi|) < \infty$, which is used by [42] to show that a solution to the Poisson equation exists. It can also be shown that if $\phi_1(x)$ and $\phi_2(x)$ are both solutions to the Poisson equation such that $\pi(|\phi_1| + |\phi_2|) < \infty$, then $\phi_1(x) + \phi_2(x) = c$ that is, the two solutions differ by a constant c. This is proved in Chap. 17 of [28].

Solving the Poisson equation is a non-trivial task, especially in high dimensional spaces. In practice, we will attempt to estimate $\phi(x)$ using the fact that it can be represented in terms of conditional expectations as

$$\phi(x) = \int_0^\infty \mathbb{E}_x[\bar{g}(X_t)]dt, \tag{4.6}$$

where

$$\mathbb{E}_x[\bar{g}(X_t)] = \mathbb{E}[\bar{g}(X_t)|X_0 = x] = \int dy P_t(x, y)\bar{g}(y). \tag{4.7}$$

This result was derived by Pardoux and Veretennikov in [22] for the case of unbounded ergodic diffusions in $\mathbb{R}^n$ and generalizes a result previously derived by Dynkin for bounded diffusions [21]. In the latter case, we have instead

$$\phi(x) = \int_0^\tau \mathbb{E}_\pi[\bar{g}(X_t)]dt, \tag{4.8}$$

where τ is the first hitting time on the boundary of the bounded process starting from $X_0 = x$. In this thesis, the general integral representation of the Poisson solution in (4.6) will be referred to as the Dynkin-Pardoux-Veretennikov (DPV) representation.

With this result, we are now ready to show how ϕ is related to the asymptotic variance σ_A^2. For this purpose, we recall the expression of the asymptotic variance derived in Sec. 2.5 as

$$\sigma_A^2 = 2\int_0^\infty \mathbb{E}_\pi[\bar{g}(X_0)\bar{g}(X_t)]dt. \tag{4.9}$$

The expectation present in the result can be expressed in terms of an integral over the state space, such that

$$\sigma_A^2 = 2\int_0^\infty dt \int_{\mathbb{R}^n} dx\ \pi(x)\mathbb{E}_x[\bar{g}(x)\bar{g}(X_t)]. \tag{4.10}$$

Moreover, if we change the order of the terms and integrals in the result we find

$$\sigma_A^2 = 2\int_{\mathbb{R}^n} dx\ \pi(x)\bar{g}(x)\phi(x) = 2\mathbb{E}_\pi[\bar{g}(X)\phi(X)], \tag{4.11}$$

where $\phi(x)$ replaced the expectation according to the DPV representation given in (4.6). An important aspect of this result is that σ_A^2 is calculated in the absence of the covariance function by using a simple expectation. This is a useful property and its importance will become clear in the next chapter.

As a side note, we recall that the Poisson solution is unique up to a constant c. To illustrate the constant's influence on σ_A^2, let us introduce a Poisson solution

$$\phi_1(x) = \phi(x) + c,$$

where $\mathbb{E}_\pi[\phi_1(x)^2] < \infty$. Evaluating the asymptotic variance (4.11) with ϕ_1, we see that

$$\sigma_A^2 = 2\mathbb{E}_\pi[\bar{g}(X)(\phi(X) + c)] = 2\mathbb{E}_\pi[\bar{g}(X)\phi(X)] + 2c\mathbb{E}_\pi[\bar{g}(X)] = 2\mathbb{E}_\pi[\bar{g}(X)\phi(X)]. \quad (4.12)$$

Hence, a constant c will have no influence on σ_A^2 as $\bar{g}$ is centered with respect to the stationary distribution.

4.2 Discrete-time Poisson equation

For completeness, we reproduce the results discussed above for a discrete ergodic Markov chain X_n with transition matrix $\mathbf{P}$ and stationary distribution π. Similarly as before, the observable has the form

$$S_n = \frac{1}{n}\sum_{i=1}^{n} g(X_i). \quad (4.13)$$

For the process and observable, the Poisson equation is now defined as

$$(I - \mathbf{P})\phi = (I - \mathbf{1}\pi)g, \quad (4.14)$$

where I denotes the identity matrix, $\mathbf{1}$ represents a column vector consisting of ones and ϕ is the Poisson solution. Note that now g and ϕ are column vectors with components g_i and ϕ_i, respectively. Equivalently in terms of $\bar{g} = g - \pi(g)$, we have

$$(\mathbf{P} - I)\phi = -\bar{g}. \quad (4.15)$$

Similarly to the continuous case, the Poisson solution ϕ is not unique and in general may not exist. We consider solutions such that $\mathbb{E}_\pi[\phi^2] < \infty$, which ensures that ϕ exists and is unique up until a constant [42]. Moreover, as before, it is possible to express ϕ in terms of the conditional expectation as

$$\phi(x) = \sum_{i=1}^{\infty} \mathbb{E}_x[\bar{g}(X_i)] \quad (4.16)$$

if X_i evolves in an unbounded state space or as

$$\phi(x) = \sum_{i=1}^{\tau(x)} \mathbb{E}_x[\bar{g}(X_i)] \quad (4.17)$$

if X_i evolves in a bounded space [43]. In the latter case, $\tau(x)$ is as before the first hitting time of the boundary when $X_0 = x$.

From these results, we can prove in a similar manner to the diffusion case that the asymptotic variance can be expressed in terms of the Poisson solution as

$$\sigma_A^2 = 2\pi(\bar{g}\phi) - \pi(\bar{g}^2) = 2\sum_x \pi(x)\bar{g}(x)\phi(x) - \sum_x \pi(x)\bar{g}(x)^2. \tag{4.18}$$

This expression is slightly more complicated than (4.11) because of the discrete-time nature of the process, see [20].

4.3 Poisson solution using large deviation theory

We describe in this section an alternative method to obtain the Poisson solution ϕ based on large deviation theory [44, 45, 46]. For simplicity, we discuss this representation for an ergodic diffusion X_t in $\mathbb{R}$ with stationary distribution π.

We recall that the rate function $I(s)$ of an additive observable S_T of X_t is obtained from the SCGF $\lambda(k)$ associated with the spectral problem described in Sec. 2.6 based on the tilted generator $\mathcal{L}_k$. For additive functionals with a (centered) test function $\bar{g}$, we specifically have

$$\mathcal{L}_k = \mathcal{L} + k\bar{g}, \tag{4.19}$$

where $\mathcal{L}$ is the generator of X_t given in (4.1).

In terms of the eigenfunction r_k associated with this operator, we can prove that

$$\phi(x) = \left.\frac{\partial_k r_k(x)}{r_k(x)}\right|_{k=0} = \partial_k r_k(x)|_{k=0} = \lim_{k\to 0}\frac{r_k(x)-1}{k}, \tag{4.20}$$

where the second equation follows since $r_0(x) = 1$ and ∂_k denotes the partial derivative with respect to k. This result comes from applying the chain rule to the right eigenfunction given in (2.82) and follows as

$$\partial_k[(\mathcal{L} + k\bar{g})r_k] = \mathcal{L}\partial_k r_k(x) + \bar{g}r_k(x) + k\bar{g}\partial_k r_k(x) = \partial_k\lambda(k)r_k(x) + \lambda(k)\partial_k r_k(x). \tag{4.21}$$

Next, we consider the limit $k \to 0$ and note that $\partial_k\lambda(k) = \lambda'(k)$ and $\lambda(0) = 0$. The resulting expression simplifies to

$$\mathcal{L}\partial_k r_k(x)|_{k=0} + \bar{g} = \lambda'(0). \tag{4.22}$$

Since $\bar{g}$ is centered, (2.77) shows that $\lambda'(0) = 0$. Thus,

$$\mathcal{L}\partial_k r_k(x)|_{k=0} = -\bar{g}(x), \tag{4.23}$$

which is the Poisson equation (4.5) for $\phi = \partial_k r_k|_{k=0}$.

This proof can be repeated for a discrete process using (4.15) as a starting point. The spectral problem in this case is the same and leads with (4.21) to

$$(\mathbf{P} - I)\partial_k r_k|_{k=0} = -\bar{g}. \tag{4.24}$$

4.4 Examples

We confirm the asymptotic variance of previously introduced examples using the different representations of the Poisson solution discussed in the previous sections.

4.4.1 Independent RVs

The first example that we consider is a sequence $X_1, X_2, \ldots, X_n$ of RVs sampled independently and identically from a normalised distribution $\pi(x)$. The distribution $\pi(x)$ stays constant over time, and therefore serves as the stationary distribution. This can also be seen as a Markov chain in which the transitions of the process are independent so that

$$P(X_{n+1} = y | X_n = x) = p(x, y) = \pi(y). \tag{4.25}$$

This allows us to rewrite the left-hand side of the Poisson equation, given by (4.15), as

$$(\mathbf{P} - I)\phi(x) = \sum_y p(x, y)\phi(y) - \phi(x) = \sum_y \pi(y)\phi(y) - \phi(x) = \mathbb{E}_\pi[\phi(x)] - \phi(x). \tag{4.26}$$

As discussed previously, the Poisson solution is unique up until a constant. Therefore, we can recenter the solution by choosing the constant equal to $-\pi(\phi)$, which leads to $\mathbb{E}_\pi[\phi(x)] = 0$. Then the resulting Poisson equation is

$$(\mathbf{P} - I)\phi(x) = -\phi(x) = -\bar{g}(x) \tag{4.27}$$

and so $\phi(x) = \bar{g}(x)$. Applying these results to the asymptotic variance (4.18) yields

$$\sigma_A^2 = 2\pi(\bar{g}^2) - \pi(\bar{g}^2) = \pi(\bar{g}^2), \tag{4.28}$$

which is the naïve variance of a centered observable.

4.4.2 Two-state Markov chain

The Poisson equation of the two-state Markov chain introduced in Sec. 2.3.1 is

$$\begin{pmatrix} -\alpha & \alpha \\ \beta & -\beta \end{pmatrix} \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix} = \begin{pmatrix} -\alpha/(\alpha + \beta) \\ 1 - \alpha/(\alpha + \beta) \end{pmatrix}, \tag{4.29}$$

where the right-hand side of the matrix equation represents the centered function $\bar{g}$ associated with $g(x) = x$. The Poisson solution is determined by solving the system of equations above and is given as

$$\phi = \begin{pmatrix} 1 \\ 1 + 1/(\alpha + \beta) \end{pmatrix}. \tag{4.30}$$

Next, we use (4.18) to calculate the asymptotic variance associated with the sample mean:

$$\begin{aligned}\sigma_A^2 &= 2\sum_x \pi(x)\bar{g}(x)\phi(x) - \sum_x \pi(x)\bar{g}(x)^2 \\ &= 2\left(-\frac{\alpha\beta}{(\alpha+\beta)^2} + \frac{\alpha\beta(\alpha+\beta+1)}{(\alpha+\beta)^3}\right) - \left(\frac{\alpha^2\beta}{(\alpha+\beta)^3} + \frac{\alpha\beta^2}{(\alpha+\beta)^3}\right) \\ &= \frac{\alpha\beta(2-\alpha-\beta)}{(\alpha+\beta)^3}. \end{aligned} \tag{4.31}$$

This result is consistent with the asymptotic variance obtained in Sec. 2.3.1.

4.4.3 Ornstein-Uhlenbeck process with linear observable

For our third example we revisit the Ornstein-Uhlenbeck process X_t, introduced in Sec. 2.8.2. The Poisson equation associated with the process and linear observable is given by

$$-\gamma x\frac{d}{dx}\phi(x) + \frac{\sigma^2}{2}\frac{d^2}{dx^2}\phi(x) = -x. \tag{4.32}$$

Note that $\mathbb{E}_\pi[x] = 0$ and therefore no re-centering is required. The solution to this second order differential equation has the general form

$$\phi(x) = \frac{x}{\gamma} + c_1 + c_2\frac{\sqrt{\pi}\sigma\,\mathrm{Erfi}(x\sqrt{\gamma}/\sigma)}{2\sqrt{\gamma}}, \tag{4.33}$$

where c_1 and c_2 are constants and $\mathrm{Erfi}(\cdot)$ is the error function. One can graphically confirm that c_2 needs to be equal to zero for $\phi(x)^2$ to be integrable with respect to the stationary distribution π. Once the integrability condition is satisfied, it is known that the Poisson solution is unique up until a constant [43], so we can set $c_1 = 0$. Hence, the solution of the Poisson equation is $\phi(x) = x/\gamma$ and the asymptotic variance related to the observable is

$$\sigma_A^2 = 2\mathbb{E}_\pi[X\phi(X)] = \frac{2}{\gamma}\mathbb{E}_\pi[X^2] = \frac{\sigma^2}{\gamma^2}. \tag{4.34}$$

The result is consistent with the asymptotic variance calculated in Sec. 2.8.2 using large deviation theory.

At this point we can also confirm the DPV representation of the Poisson solution by using the propagator of the Ornstein-Uhlenbeck process, given by [31] as

$$P_t(x,y) = \sqrt{\frac{\gamma}{\pi\sigma^2(1-e^{-2\gamma t})}}\exp\left(\frac{-\gamma(y-xe^{-\gamma t})^2}{\sigma^2(1-e^{-2\gamma t})}\right). \tag{4.35}$$

From this result, we calculate the conditional expectation as

$$\mathbb{E}_x[X_t] = \mathbb{E}[X_t|X_0 = x] = \int_{-\infty}^{\infty} P_t(x,y)y\,dy = xe^{-\gamma t} \tag{4.36}$$

and integrate this result over time to obtain the Poisson solution as

$$\phi(x) = \int_0^\infty x e^{-\gamma t} dt = \frac{x}{\gamma}. \tag{4.37}$$

This solution is consistent with the Poisson solution obtained in (4.33).

Finally, let us use large deviation theory to determine the Poisson solution in order to verify the results obtained above. In (2.111), we showed that the right eigenfunction associated with a linear observable for the Ornstein-Uhlenbeck process is

$$r_k(x) = \exp\left[\frac{kx}{\gamma} - \frac{3\sigma^2 k^2}{4\gamma^3}\right]. \tag{4.38}$$

Following (4.20), we calculate the derivative with respect to k and set $k = 0$:

$$\partial_k r_k(x)|_{k=0} = \left(\frac{x}{\gamma} - \frac{3\sigma^2 k}{2\gamma^3}\right) \exp\left[\frac{kx}{\gamma} - \frac{3\sigma^2 k^2}{4\gamma^3}\right]\Bigg|_{k=0} = \frac{x}{\gamma}. \tag{4.39}$$

This is consistent with the previously determined expressions for $\phi(x)$.

4.4.4 Ornstein-Uhlenbeck process with quadratic observable

For our last example, we consider the Ornstein-Uhlenbeck process X_t with quadratic observable, as discussed introduced in Sec. 2.8.3. We note that $\mathbb{E}_\pi[x^2] = \sigma^2/(2\gamma)$. Therefore, the Poisson equation associated with the process and observable needs to be re-centered:

$$-\gamma x \frac{d}{dx}\phi(x) + \frac{\sigma^2}{2}\frac{d^2}{dx^2}\phi(x) = -\left(x^2 - \frac{\sigma^2}{2\gamma}\right). \tag{4.40}$$

The Poisson solution ϕ is calculated by solving the second order differential equation above and is given by

$$\phi(x) = \frac{x^2}{2\gamma} + c_1 + c_2 \frac{\sqrt{\pi}\sigma \mathrm{Erfi}(x\sqrt{\gamma}/\sigma)}{2\sqrt{\gamma}}, \tag{4.41}$$

where c_1 and c_2 are constants. Similar to the discussion in Sec. 4.4.3, we need both constants to be equal to zero. Thus, the result reduces to $\phi(x) = x^2/(2\gamma)$ and one can calculate the asymptotic variance as

$$\sigma_A^2 = 2\mathbb{E}_\pi\left[\left(X^2 - \frac{\sigma^2}{2\gamma}\right)\phi(X)\right] = \frac{1}{2\gamma}\mathbb{E}_\pi\left[X^4 - \frac{\sigma^2}{2\gamma}X^2\right] = \frac{\sigma^4}{2\gamma^3}. \tag{4.42}$$

This result is consistent with the asymptotic variance calculated in Sec. 2.8.3.

As before, we also confirm the DPV representation of the Poisson solution by using the propagator of the Ornstein-Uhlenbeck process, given in (4.35). The conditional expectation is determined as

$$\mathbb{E}_x[X_t] = \mathbb{E}[X_t|X_0 = x] = \int_{-\infty}^\infty P_t(x,y)\left(y^2 - \frac{\sigma^2}{2\gamma}\right) dy = \frac{2x^2\gamma - \sigma^2}{2\gamma} e^{-2\gamma t}. \tag{4.43}$$

Next, we integrate the conditional expectation over time to obtain

$$\phi(x) = \int_0^\infty \frac{2x^2\gamma - \sigma^2}{2\gamma} e^{-2\gamma t} dt = \frac{x^2}{2\gamma} - \frac{\sigma^2}{4\gamma^2}. \tag{4.44}$$

This solution is the same as the previously determined Poisson solution up to a constant and follows the integrability condition, i.e. $\phi(x)$ follows (4.41) with $c_1 = -\sigma^2/(4\gamma^2)$ and $c_2 = 0$. As found in (4.12), the constant c_1 has no influence on the outcome of the asymptotic variance. This can be confirmed by calculating σ_A^2 as

$$\sigma_A^2 = 2\int_{-\infty}^{\infty} \phi(x)(x^2 - \frac{\sigma^2}{2\gamma})\pi(x)dx = 2\int_{-\infty}^{\infty} \left(\frac{x^2}{2\gamma} - \frac{\sigma^2}{4\gamma^2}\right)\left(x^2 - \frac{\sigma^2}{2\gamma}\right)\pi(x)dx = \frac{\sigma^4}{2\gamma^3}, \tag{4.45}$$

which is consistent with the known asymptotic variance for this example.

Finally, we verify the results using large deviation theory. In (2.126) we calculated the right eigenfunction associated with a quadratic observable for the Ornstein-Uhlenbeck process as

$$r_k(x) = c \exp\left[-\frac{x^2}{2\sigma^2}\left(\sqrt{\gamma^2 - 2k\sigma^2} - \gamma\right)\right], \tag{4.46}$$

where c represents a normalization constant. As the normalization constant does not contribute to $\phi(x)$, we omit it and calculate $\partial_k r_k(x)$ as

$$\partial_k r_k(x) = \frac{x^2}{2\sqrt{\gamma^2 - 2k\sigma}} \exp\left[-\frac{x^2}{2\sigma^2}\left(\sqrt{\gamma^2 - 2k\sigma^2} - \gamma\right)\right]. \tag{4.47}$$

Lastly, we follow (4.20) to determine the Poisson solution as

$$\phi(x) = \left.\frac{\partial_k r_k(x)}{r_k(x)}\right|_{k=0} = \frac{x^2}{2\gamma}, \tag{4.48}$$

which is consistent with the Poisson solution obtained throughout the example.

Chapter 5

Online asymptotic variance estimators

In this chapter we propose and evaluate three methods for numerically estimating the asymptotic variance based on the Poisson equation, discussed in Chap. 4. The first two methods are based on the DVP representation, discussed in Sec. 4.1, and deals with estimating the conditional expectation using either multiple trajectories or one long trajectory. The third method is based on a stochastic approximation of the Poisson solution, as determined from (4.20), which relates that solution to the dominant eigenvector of the tilted generator. A similar stochastic representation was considered in [47] and allows us here to iteratively estimate the Poisson solution and, in turn, the asymptotic variance in an 'online' way using a single trajectory of a Markov process. The accuracy of these methods is evaluated with the two-state process and the Ornstein-Uhlenbeck process, for which we have calculated the exact asymptotic variance in the previous chapters.

5.1 Dynkin-Pardoux-Veretennikov representation of the Poisson solution

We consider an ergodic Markov process X_t that converges to a stationary distribution π, as introduced in Sec. 2.4. The observable of the process is the same as introduced in Sec. 2.5, namely,

$$S_T(g) = \frac{1}{T}\int_0^T g(X_t)dt, \tag{5.1}$$

where g is assumed to be centered with respect to the stationary distribution.

The first estimator of the asymptotic variance that we consider is based on the DPV representation of the Poisson solution ϕ, given by

$$\phi(x) = \int_0^\infty \mathbb{E}_x[g(X_t)]dt, \tag{5.2}$$

presented in Sec. 4.1 [22, 21]. From this result, it is natural to define an estimator of ϕ by generating L independent trajectories $\{X_t^{(j)}\}_{j=1}^L$ of the process started at x, that is, $X_0^{(j)} = x$, and to estimate the conditional expectation in (5.2) by summing these copies

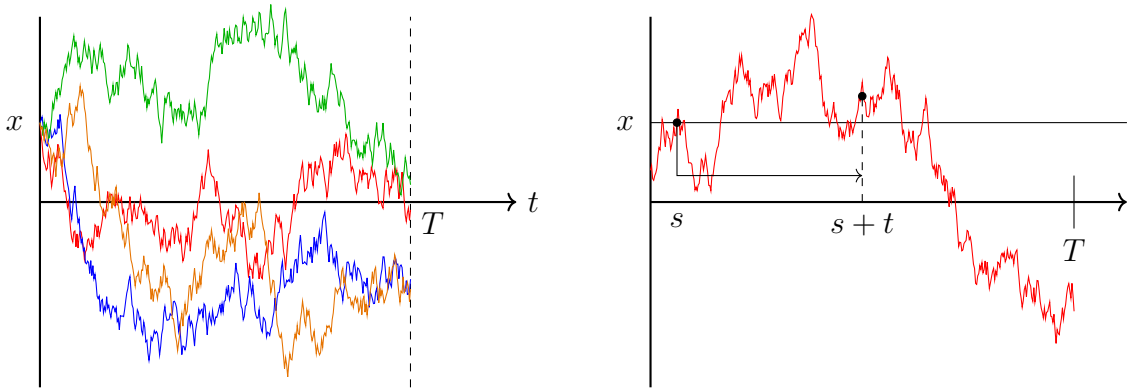


Figure 5.1: Illustration of the two approaches for estimating the conditional expectation (5.2). (Left) Simulation of many copies or replications of trajectories of length t starting at x. (Right) Simulation of one, long trajectory of length T, from which we extract segments of length t starting at x.

or replications of the process at different times, as illustrated in Fig. 5.1 (left). This leads us to define our first estimator of ϕ as

$$\hat{\phi}_{L,T}(x) = \int_0^T dt \frac{1}{L} \sum_{j=1}^{L} g(X_t^{(j)}), \tag{5.3}$$

where $X_t^{(j)}$ denotes the j-th replication of the process. This estimator depends on L, the number of replications, and T, the time used to truncate the infinite sum in (5.2). In the double limit where $L \to \infty$ and $T \to \infty$, we should have

$$\lim_{\substack{L\to\infty \\ T\to\infty}} \hat{\phi}_{L,T}(x) = \phi(x) \tag{5.4}$$

in probability, pointwise for all x. This follows from the law of large numbers, since the replications are independent.

In practice, we use fixed values of L and T, and increase these values until $\hat{\phi}_{L,T}$ seems to stabilize. We also simulate X_t by discretizing the process with respect to time for a fixed step-size Δt. Thus for one point x, the estimation of $\hat{\phi}_{L,T}(x)$ requires that we simulate L trajectories containing $T/\Delta t$ points. This can become computationally expensive when we consider a wide range of x, while increasing L and T for convergence.

Since we assume the process is ergodic, the conditional expectation can be estimated using a single trajectory to locate the times at which X_t hits the value x and then record the state of the process after a period or duration t, as illustrated in Fig. 5.1 (right). This defines a number of segments such that $X_s = x$ and $X_{s+t} = y$, which can be used to approximate the conditional expectation as

$$\hat{\phi}(x,t)_T = \frac{1}{N_{\text{seg}}} \sum_{i=1}^{N_{\text{seg}}} g(X_{s+t}^{(i)}), \tag{5.5}$$

where N_{seg} is the number of segments and $X_{s+t}^{(i)}$ is the last value of the i-th segment. Alternatively, we can write

$$\hat{\phi}(x,t)_T = \frac{1}{N_{\text{seg}}} \int_1^{T-t} ds\, g(X_{s+t}) \delta_{X_s,x}. \tag{5.6}$$

From this estimator, we then define our second estimator of ϕ as

$$\hat{\phi}(x)_T = \int_0^T dt\, \hat{\phi}(x,t)_T, \tag{5.7}$$

This estimator now only depends on T, the truncation time, which is also the length of the simulated trajectory, discretised as before in steps of Δt. In practice, we must also ease the constraint $X_s = x$, appearing in (5.6) as $\delta_{X_s,x}$, by accepting observations X_s that fall within a small range around x. Provided that this range is small enough, we then have

$$\lim_{T\to\infty} \hat{\phi}_T(x) = \phi(x) \tag{5.8}$$

in probability by the ergodic theorem.

Once we have an estimator of the Poisson solution, we estimate the asymptotic variance using (4.11) as

$$\hat{\sigma}_A^2 = 2 \int dx\, \hat{\pi}(x) g(x) \hat{\phi}(x), \tag{5.9}$$

where, for simplicity, we write $\hat{\phi}$ as either the estimator in (5.3), which depends on L and T, or the estimator in (5.7), which depends on T. In this formula, $\hat{\pi}$ is also the estimator of the invariant density π, calculated either with L replications or as an ergodic expectation, that is, as the histogram of the values visited by X_t as $t \to \infty$.

5.2 Application

We illustrate the convergence of the two estimators defined in the previous section by considering the Ornstein-Uhlenbeck process, introduced in Sec. 2.8.2, and the linear observable

$$S_T = \frac{1}{T} \int_0^T X_t dt, \tag{5.10}$$

introduced in Sec. 2.5. The Poisson solution corresponding to the process and observable is known to be $\phi(x) = x/\gamma$, as proved in Sec. 4.4.3.

We estimate ϕ for the Ornstein-Uhlenbeck process for the parameters $\gamma = 1$ and $\sigma = 1$ using both the replication and single trajectory methods. In the first method, we use $N = 2000$ replications, $\Delta t = 0.01$ and $T_1 = 50$. In the single trajectory method, we use a longer time $T_2 = 100000$ to make up for the fact that only one trajectory is considered, and accumulate the conditional expectation of x with values falling in the range of $[x - 0.01, x + 0.01]$. The results are shown in Fig. 5.2 and are compared with the exact Poisson solution (in blue).

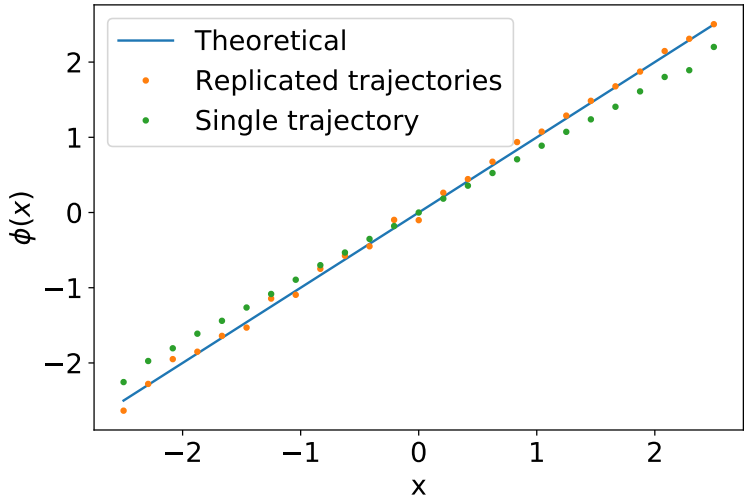


Figure 5.2: Estimated Poisson solution of the Ornstein-Uhlenbeck process with linear observable using replicated trajectories and single trajectory methods. Parameters: $\gamma = 1$, $\sigma = 1$, $\Delta t = 0.01$, $N = 2000$ $T_1 = 50$ and $T_2 = 100000$

The limitation of the single trajectory method is that it estimates the Poisson solution using only the states visited by the trajectory. Thus, $\hat{\phi}(x)_T$ is more accurate in the region where there are more observations (i.e. states visited) contributing to the estimation of the stationary density and is, conversely, less accurate in the region where there are less observations. There is no such limitation for the replicated trajectory method since each simulation is started at x over the entire region. However, the clear advantage of the single trajectory method is that we only need to simulate one trajectory. Moreover, ϕ values far from the expected value of the process are not so important, since for the asymptotic variance, ϕ is weighted over π.

This difference between the replicated and single trajectory methods leads to different estimated variance, namely, $\hat{\sigma}_A^2 = 1.06$ and $\hat{\sigma}_A^2 = 0.89$, respectively, compared to the exact value which is $\sigma_A^2 = 1$. Hence, for this example a higher accuracy is achieved by estimating $\phi(x)$ using replicated trajectories. However, we stress again that the single trajectory is more efficient as it does not require the simulation of many different trajectories. As a result, it can be applied if all we have is a single trajectory either simulated or recorded experimentally.

5.3 Stochastic approximation of the Poisson solution

The third method that we discuss requires only one trajectory, similarly to the second method discussed in the previous section, but does not require that we scan that trajectory with different time windows or lags. Rather, it estimates the Poisson solution $\phi(x)$ by using the successive states forming the trajectories, which means that it can be applied in an 'on-the-fly' or 'online' way as a trajectory is simulated or recorded experimentally.

The method is based on the result (4.20) discussed in Sec. 4.3, which links $\phi(x)$ to the eigenfunction related to the dominant eigenvalue and SCGF $\lambda(k)$ of the tilted generator $\mathcal{L}_k$. It also uses a stochastic approximation technique, developed recently in [47], based on a discretization of the semi-group associated with the operator $\mathcal{L}_k$.

To explain this technique, we need to introduce the semi-group P_t, which is an operator acting on smooth function $h : X \to \mathbb{R}$ according to

$$P_t h(x) = \int dy\, P_t(x,y) h(y) = \mathbb{E}_x[h(X_t)] = \mathbb{E}[h(X_t)|X_0 = x], \tag{5.11}$$

where, as noted before, $\mathbb{E}_x[h(X_t)]$ is the conditional expectation of $h(X_t)$ for the process started at $X_0 = x$. In this expression, $P_t(x,y)$ is the probability density of X_t given $X_0 = x$, that is, $P_t(x,y) = p(X_t = y|X_0 = x)$ or $P_t(x,y) = p(X_{s+t} = y|X_s = x)$ by stationarity of the process. From the semi-group, the generator $\mathcal{L}$ of X_t is defined as

$$\mathcal{L}h(x) = \lim_{t\to 0} \frac{P_t h(X_t) - h(x)}{t}. \tag{5.12}$$

From the stationarity property, we can also write

$$\mathcal{L}h(x) = \lim_{\Delta t\to 0} \frac{\mathbb{E}[h(X_{t+\Delta t})|X_t = x] - h(x)}{\Delta t}. \tag{5.13}$$

Taking the expectation of this result for any initial value leads to an expression which describes the action of the generator as an expectation, given by

$$\frac{d}{dt}\mathbb{E}[h(X_t)] = \mathbb{E}[\mathcal{L}h(X_t)]. \tag{5.14}$$

Note that $\mathcal{L}$ is the same operator as the one given in (2.52): it is a linear differential operator that acts on a function h to produce a new function that we write as $\mathcal{L}h(x)$, x being the argument of the new function.

Since we are interested in estimating $r_k(x)$, the operator that is of interest to us is not $\mathcal{L}$ but $\mathcal{L}_k$, the tilted generator of X_t and observable S_T. Appendix A.2 of [48] shows that this operator is also associated with a semi-group P_t^k, called the Feynman-Kac semi-group, defined by

$$Q_t^k h(x) = \mathbb{E}[h(X_t)e^{tkS_t}|X_0 = x]. \tag{5.15}$$

For $k = 0$, we recover $Q_t^0 = P_t$ and, therefore, $\mathcal{L}_k = \mathcal{L}$, in accordance with the expression of $\mathcal{L}_k$ given in (2.80).

As mentioned in Sec. 2.6, $\mathcal{L}_k$ also describes the direct eigenvalue problem (2.82) involved in the calculation of the SCGF $\lambda(k)$ of S_T. Following [47], we assume that $\mathcal{L}_k$ acts on functions that are square-integrable with respect to π and that it has a dominant eigenfunction r_k such that $r_k > 0$. Since r_k is an eigenfunction of $\mathcal{L}_k$, we have

$$Q_t^k r_k = e^{t\lambda(k)} r_k \tag{5.16}$$

for all time $t > 0$. Moreover, it can be proved (see references cited in Sec. 3 of [47]) that, if we apply Q_t^k to an arbitrary smooth function h, then we obtain in the long-time limit

$$e^{-t\lambda(k)}(Q_t^k h) \xrightarrow{t\to\infty} r_k \int_{\mathcal{X}} h(x) l_k(x) dx, \tag{5.17}$$

where $\mathcal{X}$ is the space containing X_t and l_k is the eigenfunction dual to r_k, defined by the dual spectral problem (2.90). This limit is a functional version of the power method used in numerical analysis to find the dominant eigenvalue of matrices. Here, we will use it to find the dominant eigenvalue and SCGF $\lambda(k)$.

For this purpose, we must discretize the action of the semi-group Q_t^k in time to write

$$Q_{\Delta t}^k h(x) = \mathbb{E}[e^{kg(X_t)\Delta t} h(X_{t+\Delta t})|X_t = x] = e^{kg(x)\Delta t}\mathbb{E}[h(X_{t+\Delta t})|X_t = x]. \tag{5.18}$$

The advantage of the discretization is that the limit (5.17) can be rewritten in terms of the iteration number n as

$$(Q_{\Delta t}^k)^n h(x) \sim e^{n\Delta t\lambda(k)} r_k(x) \int_{\mathcal{X}} h(y) l_k(y) dy \tag{5.19}$$

for $n \to \infty$, where $\sim$ denotes asymptotic equivalence in n. This shows that $\lambda(k)$ and r_k can be computed by recursively applying $Q_{\Delta t}^k$ to an initial guess h. To be more precise, let us define

$$h^{n+1}(x_n) = Q_{\Delta t}^k h^n(x_n) = e^{kg(x_n)\Delta t}\mathbb{E}[h^n(X_{n+1})|x_n], \tag{5.20}$$

where x_n is the state of X_t at the n-th time step (with discretization Δt) and X_{n+1} is the next state. Then we must have $h^n \to r_k$ as $n \to \infty$ from any initial function h^0. Note that the superscript on h is not a power, but the index of the iteration.

Similarly to the previous section, we could attempt to estimate the conditional expectation in (5.20) by simulating L copies or replicates of the random variable X_{n+1} from $X_n = x_n$ and take

$$h^{n+1}(x_n) \approx e^{kg(x_n)\Delta t} \frac{1}{L} \sum_{j=1}^{L} h^n(x_{n+1}^{(j)}), \tag{5.21}$$

as an approximation of the next iterate h^{n+1}, $x_{n+1}^{(j)}$ being the j-th copy of X_{n+1}. Although this approach is relatively easy to implement, it requires a recalculated expectation at each iteration. This could lead to a computationally expensive estimator and does not allow us to construct an online estimation of the asymptotic variance.

Here, we use instead a common approximation of the conditional expectation, referred to as a stochastic approximation, which amounts to writing

$$h^{n+1}(x_n) \approx e^{kg(x_n)\Delta t} h^n(x_{n+1}), \tag{5.22}$$

where x_{n+1} is the simulated or observed value of X_{n+1} starting from $X_n = x_n$. We note that this value is distributed according to $P_{\Delta t}(x_n, x_{n+1})$. Expanding the exponential to first order in Δt, we then have

$$h^{n+1}(x_n) \approx [1 + kg(x_n)\Delta t]\, h^n(x_{n+1}). \tag{5.23}$$

The stochastic approximation in (5.22) might appear to be very crude, but is known to yield the correct expectation for ergodic processes in the limit $n \to \infty$ when it is used in conjunction with an annealing scheme that changes the iteration (5.20) to

$$h^{n+1} = h^n + a_n(Q_{\Delta t}^k h^n - h^n) = (1 - a_n)h^n + a_n Q_{\Delta t}^k h^n, \tag{5.24}$$

where a_n is a sequence of decreasing positive numbers. By approximating the expectation related to Q_t^k as in (5.23), we then have

$$h^{n+1}(x_n) = (1 - a_n)h^n(x_n) + a_n(1 + kg(x_n)\Delta t)h^n(x_{n+1}). \tag{5.25}$$

The sequence $(a_n)_{n\geq 1}$ is called the annealing or learning sequence and serves as a smoothing parameter that filters noisy updates by controlling the rate at which h^{n+1} is updated. The idea, as can be seen from (5.24), is that the iteration "forgets" the current estimate h^n at a rate of $(1 - a_n)$ and updates with $Q_{\Delta t}^k h^n$ at a rate of a_n.

To ensure convergence, it is known that a_n must be chosen such that

$$\sum_{n\geq 1} a_n = \infty \quad \text{and} \quad \sum_{n\geq 1} a_n^2 < \infty. \tag{5.26}$$

For example, we can choose $a_n = 1/n$ or, more generally,

$$a_n = n^{-1/\nu} \tag{5.27}$$

with ν chosen so that (5.26) holds.

In simulations or in experiments, we need of course to approximate the continuous function $h^n(x)$ onto a set of points, components, or a function basis, which we take here to be a mesh of equidistant points on $\mathcal{X}$. In particular, for $\mathcal{X} = \mathbb{R}$, we re-write (5.23) as

$$h^{n+1}(i) = [1 + kg(i)\Delta t]h^n(j), \tag{5.28}$$

while the annealed version of this iteration (5.25) becomes

$$h^{n+1}(i) = (1 - a_n)h^n(i) + a_n Q_{\Delta t}^k h^n(i) = (1 - a_n)h^n(i) + a_n(1 + kg(i)\Delta t)h^n(j). \tag{5.29}$$

In these expressions, i and j represent the discretized indices of the positions x_n and x_{n+1}, respectively, projected onto the mesh. In this way, $h^n(x)$ is approximated spatially as a discrete vector, so that (5.29) is basically a stochastic version of the power method applied to vectors. As is common in that method, we can include a normalization step between iterates, that is, $h^n \to h^n/h^n(i)$ for some component i. Hence, $h^n(i) = 1$ and this ensures that h^n is normalized with respect to the expected value of the stationary distribution as $n \to \infty$.

Now that we have defined an estimator for r_k in terms of the iterate h^n, we can define our estimator for the Poisson solution. To this end, we replace r_k with h^{n+1} in (4.20) so as to define the following estimator for $\phi(x)$:

$$q^{n+1}(x_n) = \lim_{k\to 0} \frac{h^{n+1}(x_n) - 1}{k} = \lim_{k\to 0} \frac{(1 + kg(x_n)\Delta t)h^n(x_{n+1}) - 1}{k}, \tag{5.30}$$

which becomes

$$q^{n+1}(x_n) = \lim_{k\to 0} \frac{h^n(x_{n+1}) - 1}{k} + g(x_n)\Delta t h^n(x_{n+1}) \approx q^n(x_{n+1}) + g(x_n)\Delta t \tag{5.31}$$

after some manipulations and using the fact that $h^n \to 1$ for $k = 0$. As we did for h^n, we can also apply the annealing scheme here to obtain the following iteration scheme for the Poisson solution:

$$q^{n+1}(x_n) = q^n(x_n) + a_n[Q^k_{\Delta t} q^n(x_n) - q^n(x_n)] \tag{5.32}$$

$$= (1 - a_n)q^n(x_n) + a_n[q^n(x_{n+1}) + g(x_n)\Delta t]. \tag{5.33}$$

Note that this iteration does not depend on k, although the iteration for h^n does. When discretized on a mesh grid, we thus obtain

$$q^{n+1}(i) = (1 - a_n)q^n(i) + a_n[q^n(j) + g(i)\Delta t] \tag{5.34}$$

as our stochastic approximation of ϕ. This assumes that g is centered; if this is not the case, we must recenter it using an estimate of the expectation of S_T.

With our estimate of ϕ, we finally turn our attention to the estimation of the asymptotic variance using the main result (4.11) presented in Sec. 4.1. Since this result involves an expectation over the stationary density π, it is natural to first use the time average

$$\hat{\sigma}^2_{A,T} = \lim_{T\to\infty} \frac{2}{T} \int_0^T g(X_t)\hat{\phi}(X_t)dt \tag{5.35}$$

as our estimator of σ^2_A. Discretizing the time integral and plugging in our estimator of the Poisson solution, we then write

$$\hat{\sigma}^2_{A,N} = \frac{2}{N\Delta t} \sum_{i=1}^{N} g(x_i)q^i(x_i), \tag{5.36}$$

where $N = T/\Delta t$ and q^i is the i-th iterate approximation of ϕ. Note that q^i is updated concurrently with the iterate x_i of the process, which allows us to estimate the asymptotic variance using only the information available at each time step of a single run. The estimator $\hat{\sigma}^2_{A,N}$ converges to σ^2_A from the fact that $q^n \to \phi$ and the ergodic theorem, which guarantees that the running sum converges to the ergodic expectation with respect to π.

For the derivation of the estimator, we assumed that g is centered with respect to the stationary distribution, i.e. $\mathbb{E}_\pi[g] = 0$. If the validity of this property is unknown or g is not centered, the adjusted estimator has the form

$$\hat{\sigma}^2_{A,N} = \frac{2}{N\Delta t} \sum_{i=1}^{N} [g(x_i) - A_i]\, q^i(x_i), \tag{5.37}$$

where

$$A_i = \frac{1}{i} \sum_{j=1}^{i} g(x_j) \tag{5.38}$$

is the estimate of the expectation of S_T, discretized in time, up to the current time index i.

5.4 Applications

In this section we implement our proposed online estimator of the asymptotic variance (5.36) on three test cases already seen throughout the thesis, namely, the two-state Markov chain with a linear observable and the Ornstein-Uhlenbeck process with a linear and quadratic observables, respectively. The accuracy of these estimates is analysed and compared to the batch means method, introduced in Sec. 3.3.

5.4.1 Two-state Markov Chain

The first example that we consider is the ergodic two-state Markov chain with a linear observable, as introduced in Sec. 2.3.1. Following (5.36), we repetitively determine $\hat{\sigma}^2_{A,N}$ while increasing the number of iterations of a two-state Markov chain with parameters $\alpha = 0.3$ and $\beta = 0.8$. For this discrete-time model, we obviously have $\Delta t = 1$, and we use the annealing sequence shown in (5.27) with $\nu = 0.9$. As the theoretical variance is known from (2.33), we can repeat this procedure to calculate the error between the estimated and exact asymptotic variance. The relative error given in (3.12) depends on the simulation and in this case produces varying results as the online estimator only considers a single trajectory. Therefore, we consider the errors resulting from multiple simulations by calculating the average relative error (ARE_N) at respective number of observations N as

$$\text{ARE}_N = \frac{1}{n_{\text{ARE}}} \sum_{i=1}^{n_{ARE}} \frac{\left|\hat{\sigma}^2_{A,N} - \sigma^2_A\right|}{\sigma^2_A}, \tag{5.39}$$

where n_{ARE} is the number of repeated simulations. The results are shown in Fig. 5.3 (left), for $n_{\text{ARE}} = 40$, and show that the estimator (5.36) converges to σ^2_A as the number N of iterations increase.

In order to gain insight into the estimator's performance for this example, we use (5.36) to estimate a range of different asymptotic variances. We determine the values of α and β required to produce the desired range of variances by solving (2.33) with respect to each predefined variance. We use the calculated set of $\alpha = [0.2, 0.1, 0.05, 0.05, 0.05, 0.01, 0.01, 0.01, 0.01, 0.01]$ and set of $\beta = [0.2000, 0.1205, 0.0915, 0.0656, 0.0450, 0.0411, 0.0368, 0.0334, 0.0305, 0.0280]$ values to simulate separate two-state Markov chains for $N = 10^6$. The results are shown in Fig. 5.3 (right) and compared to the theoretical asymptotic variance (dashed blue). We see that the proposed estimator accurately estimates the considered range of variances.

Lastly, we compare the convergence of the proposed estimator to that of the batch means method for a two-state Markov chain with parameters $\alpha = 0.3$ and $\beta = 0.8$. Both techniques have hyperparameters that influence the accuracy of the estimation. Hence, we firstly determine the annealing constant and batch size that give the smallest average relative error. This is done by calculating the average relative error, with $n_{\text{ARE}} = 40$, over a range of annealing constants and batch sizes for each considered N and by selecting the constants corresponding to the smallest error. Secondly, we use these optimal annealing constants and optimal batch sizes to calculate the average relative error for an increasing number N of observations. The results are shown in Fig. 5.4 and show that the batch

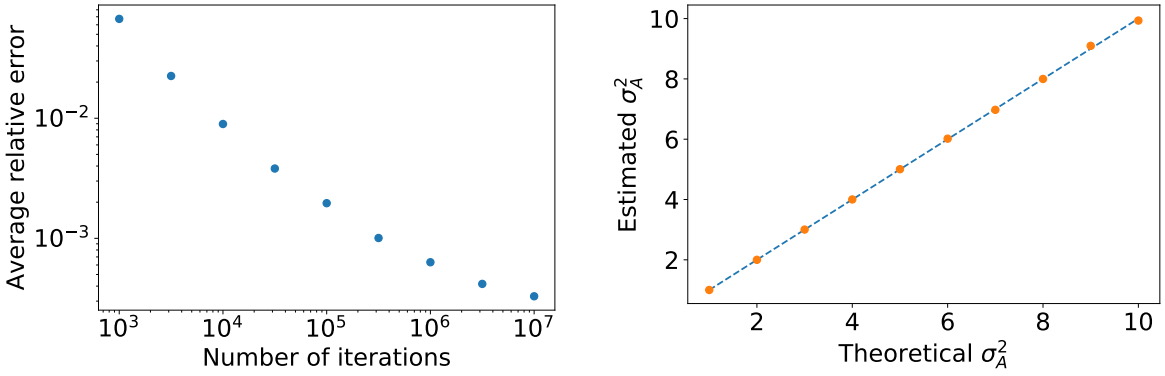


Figure 5.3: (Left) Convergence of the proposed online estimator as a function of N for the two-state Markov process with parameters $\alpha = 0.3$, $\beta = 0.8$, $\nu = 0.9$ and $n_{\text{ARE}} = 40$. (Right) Proposed online estimator approximating a range of different theoretical σ_A^2 values of a two-state Markov chain with $N = 10^6$. Parameters: $\alpha = [0.2, 0.1, 0.05, 0.05, 0.05, 0.01, 0.01, 0.01, 0.01, 0.01]$ and $\beta = [0.2000, 0.1205, 0.0915, 0.0656, 0.0450, 0.0411, 0.0368, 0.0334, 0.0305, 0.0280]$

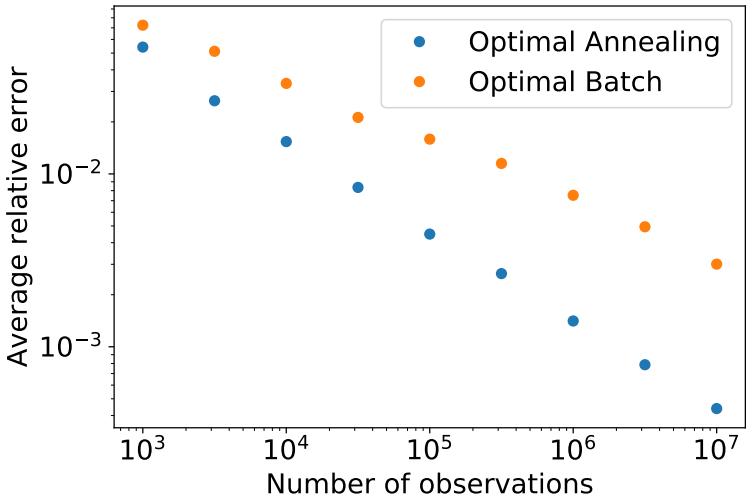


Figure 5.4: Comparison of the convergence rates of the batch means method and proposed estimator as a function of N, with optimal batch size and optimal annealing imposed on the two-state Markov chain with parameters $\alpha = 0.3$ and $\beta = 0.8$.

means method (orange) converges to the theoretical asymptotic variance at a rate of approximately 0.35, whereas the proposed method (blue) converges at approximately 0.52. Therefore, the proposed method has a faster convergence rate than the batch means method for this example.

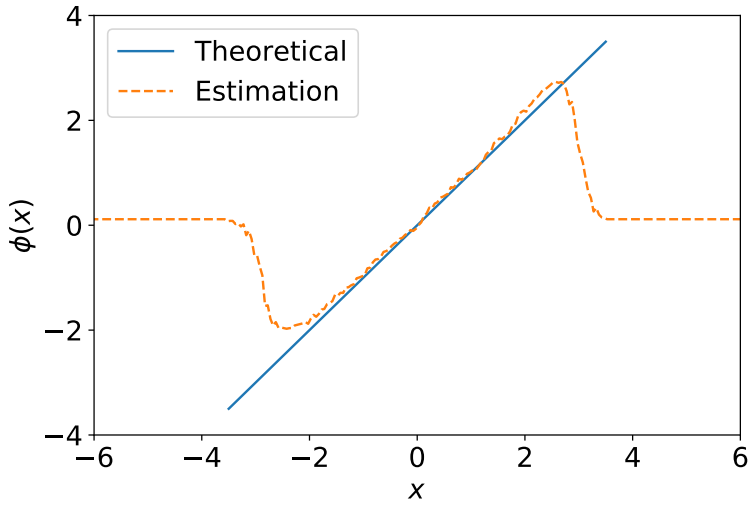


Figure 5.5: Stochastic approximation of the Poisson solution associated with the Ornstein-Uhlenbeck process and linear observable. Parameters: $\gamma = 1$, $\sigma = 2$, $T = 100000$, $\Delta t = 0.01$, $\nu = 9$ and $q^0 = 0$.

5.4.2 Ornstein-Uhlenbeck process with linear observable

The second test case that we consider involves the Ornstein-Uhlenbeck process with a linear observable, as introduced in Sec. 2.8.2, with known asymptotic variance given in (2.117).

The proposed estimator (5.36) depends on the stochastic approximation of the Poisson solution q, given in (5.34). We calculate q for the process with parameters $\gamma = 1$ and $\sigma = 2$. We use one trajectory of length $T = 100000$, $\Delta t = 0.01$, as well as an annealing sequence following (5.27) with $\nu = 9$ and $q^0 = 0$. The results are shown in Fig. 5.5 and compared with the theoretical Poisson solution (blue). The limitation of the stochastic approximation is similar to that of the single trajectory method, discussed in Sec. 5.2, in that ϕ is estimated using only the states visited by the trajectory. Thus, the estimated Poisson solution (dashed orange) is more accurate in the region that is most visited by the process and less accurate further away from this region. Fortunately, this inaccuracy does not influence $\hat{\sigma}^2_{A,N}$ as q^n is weighted again over π. The inaccuracy of q^n is expressed as a horizontal line resulting from the linear initial guess q^0 not being updated throughout the simulation.

Next, we illustrate the importance of the annealing sequence by estimating the asymptotic variance using different sequences. Following (5.36), we estimate σ^2_A according to an increasing sequence of possible ν-values, for a fixed $T = 100000$. We use $\gamma = 1$, $\Delta t = 0.01$ and $\sigma = [1, 1.5, 2]$ to calculate the average relative error with $n_{\text{ARE}} = 100$ for each ν-value. We normalize these results in order to compare the shapes of the obtained error functions (Fig. 5.6). Firstly, we consider the case where $\sigma = 1.5$ (orange) and note that the optimal annealing value is $\nu = 9$. When $\nu > 9$, a_n decreases too slowly and the stochastic nature of the process contributes a too large portion of the estimator leading to an inaccurate

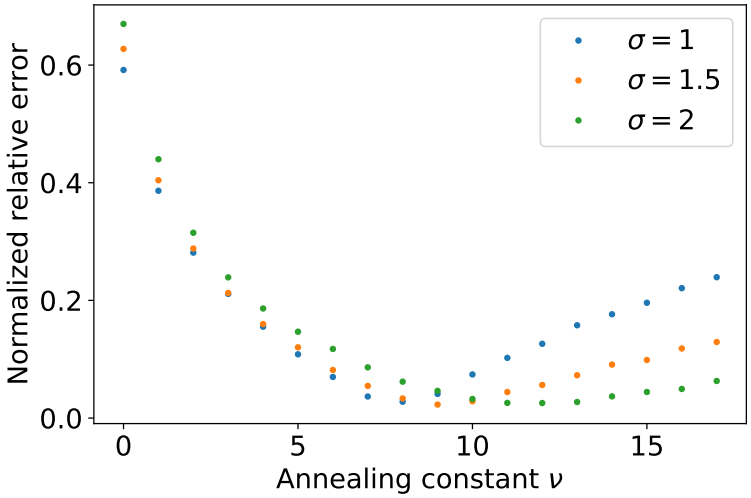


Figure 5.6: The influence of the annealing scheme on the estimated asymptotic variance of the Ornstein-Uhlenbeck process with linear observable. Parameters: $\gamma = 1$, $T = 100000$, $\Delta t = 0.01$ and $n_{\text{ARE}} = 100$.

result. On the other hand, if $\nu < 9$, a_n decreases too quickly and the estimator does not retain enough information from the updates to accurately estimate σ_A^2. Therefore it is important that a_n decreases at an optimal rate. Secondly, although the shapes between the results of $\sigma = 1$ (blue) and $\sigma = 2$ (green) are similar to that of $\sigma = 1.5$, none have the same optimal annealing value. Hence, the optimal annealing value should be calculated for each individual process.

To study the convergence of the method with respect to the number of observations, we calculate the average relative error (5.39) of the proposed estimator for the Ornstein-Uhlenbeck process with $\gamma = 1$ and $\sigma = 1$ at distinct iterations. We use $n_{\text{ARE}} = 40$ replications, $\Delta t = 0.01$ and the annealing scheme (5.27) with $\nu = 8$. The results are shown in Fig. 5.7 (left) and we see that $\hat{\sigma}_{A,N}^2$ converges to the theoretical asymptotic variance (2.117) as the number N of iterations increases. We note that the error does not converge at a constant rate. This is due to the suboptimal choice of ν at each instance of the considered iterations.

To evaluate the influence that the variability of a process has on the proposed online estimator, we estimate a range of different asymptotic variances for the Ornstein-Uhlenbeck process. The γ and σ pairs required to generate an Ornstein-Uhlenbeck process which has a desired $\sigma_{A,N}^2$ is determined by solving (2.115). We use $\gamma = [1, 1, 1, 1, 1, 1, 1, 1, 1, 1]$ and $\sigma = [1.0000, 1.4142, 1.7321, 2.0000, 2.2361, 2.4495, 2.6458, 2.8284, 3.0000, 3.1623]$ to simulate separate trajectories of $T = 100000$, where $\Delta t = 0.01$ and $\nu = 8$. From Fig. 5.7 (right), we see that the proposed estimator accurately estimates the range of variances.

Finally, we compare the convergence between the proposed estimator (blue) and the batch means method (orange) for the Ornstein-Uhlenbeck process with $\gamma = 1$ and $\sigma = 1$. As before, we calculate the optimal annealing constant and optimal batch size corresponding to the number of iterations. We use the optimal hyperparameters to calculate the av-

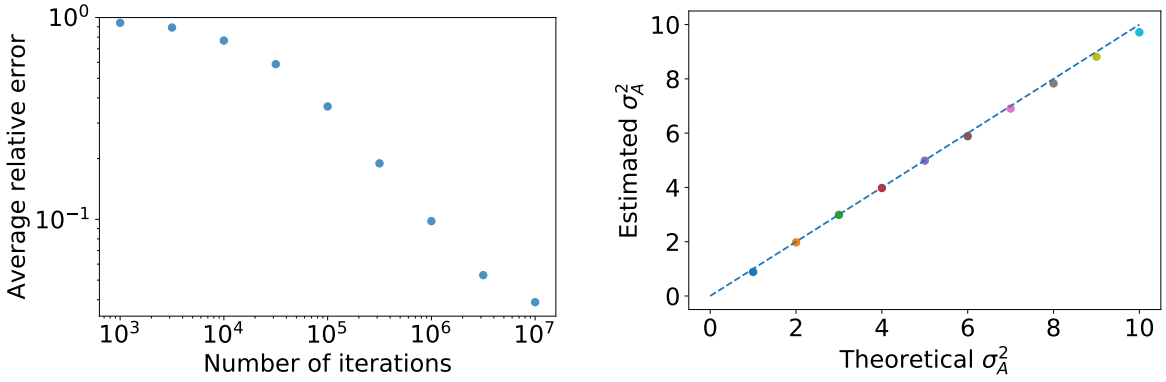


Figure 5.7: (Left) Convergence of $\hat{\sigma}^2_{A,N}$ as a function of the number N of iterates for the Ornstein-Uhlenbeck process with linear observable. Parameters: $\gamma = 1$, $\sigma = 1$, $\Delta t = 0.01$ and $\nu = 8$. (Right) Range of estimated asymptotic variances for the Ornstein-Uhlenbeck process with $\gamma = [1, 1, 1, 1, 1, 1, 1, 1, 1, 1]$, $\sigma = [1.0000, 1.4142, 1.7321, 2.0000, 2.2361, 2.4495, 2.6458, 2.8284, 3.0000, 3.1623]$, $T = 100000$, $\Delta t = 0.01$ and $\nu = 8$.

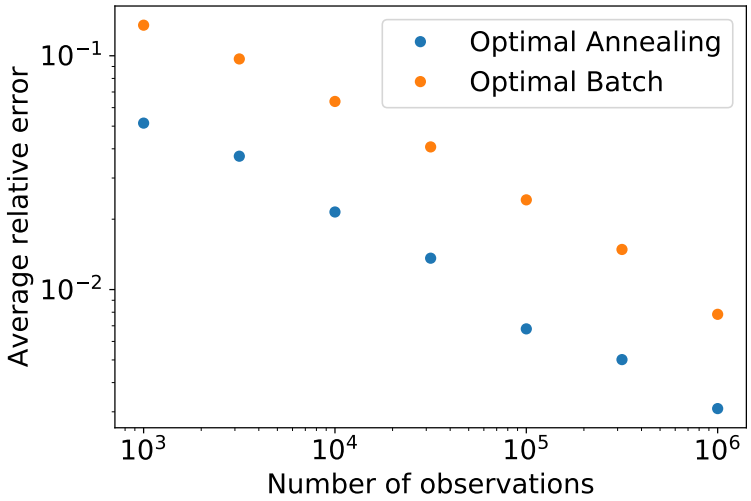


Figure 5.8: Convergence of the batch means method with optimal batch size and the proposed online estimator with optimal annealing as a function of N for Ornstein-Uhlenbeck process with a linear observable. Parameters: $\gamma = 1$, $\sigma = 1$ and $\Delta t = 0.01$

erage relative error of both methods for an increasing number N of iterations (Fig. 5.8). The results in Fig. 5.8 show that both methods converge to σ_A^2 at a rate of approximately 0.4. However, the proposed estimator is more accurate than the batch means method for the considered case, as it starts with a lower average relative error.

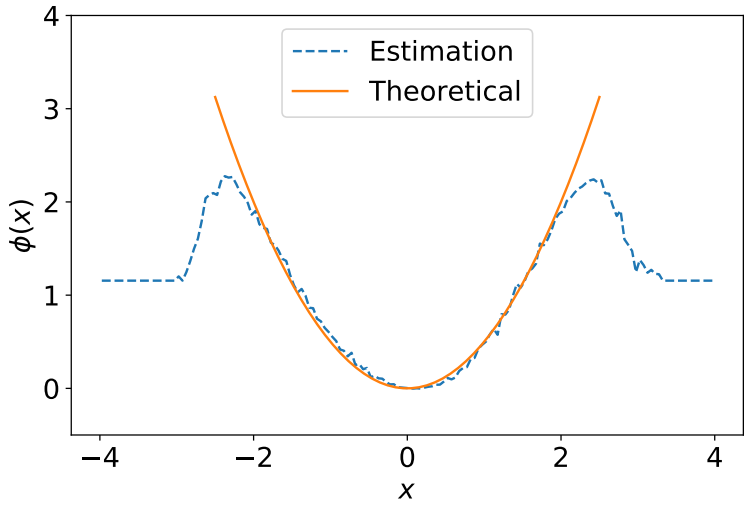


Figure 5.9: Stochastic approximation of the Poisson solution associated with the Ornstein-Uhlenbeck process and quadratic observable. Parameters: $\gamma = 1$, $\sigma = 1$, $T = 100000$, $\Delta t = 0.01$, $q^0 = 0$ and $\nu = 8$.

5.4.3 Ornstein-Uhlenbeck process with quadratic observable

As a last example, we revisit the Ornstein-Uhlenbeck process with a quadratic observable, as introduced in Sec. 2.8.3. Similar to the previous sections, we use the proposed estimator (5.36) to approximate the known asymptotic variance (2.129).

As mentioned previously, the estimator depends on the stochastic approximation q of the Poisson solution ϕ. We calculate this approximation for the Ornstein-Uhlenbeck process with $\gamma = 1$ and $\sigma = 1$. We use a trajectory of length $T = 100000$ and $\Delta t = 0.01$ to estimate ϕ starting with an initial guess of $q^0 = 0$ and annealing sequence (5.27) $\nu = 8$. The results are shown in Fig. 5.9 and compared with the exact solution (orange). The Poisson solution is estimated using the states visited by the trajectory. Hence, from the results we again see that the estimator accurately represents the known Poisson solution (4.41) in the region visited by the process. The inaccuracy of the estimation is shown by the horizontal line resulting from the linear initial guess not being updated. We note that the horizontal line does not correspond to the initial value $q^0 = 0$. This displacement is due to the continuous centering of q^n with respect to the stationary distribution between iterations. Fortunately, the inaccurate representation of $\phi(x)$ has a negligible contribution on the final $\hat{\sigma}^2_{A,N}$.

We repetitively calculate $\hat{\sigma}^2_{A,N}$ for a increasing number N of observations of the Ornstein-Uhlenbeck process with $\gamma = 1$ and $\sigma = 1$. We use $\Delta t = 0.01$, $q^0 = 0$, $n_{\text{ARE}} = 40$ and $\nu = 8$ for the annealing sequence (5.27). From the results shown in Fig. 5.10 (left), we see that the estimator converges to the theoretical asymptotic variance as the number of iterations increase. However, the convergence is slow for the initial iterations, but increase as the number of iterations increases. The inconsistent convergence rate is due to the suboptimal choice of ν for the considered iterations.

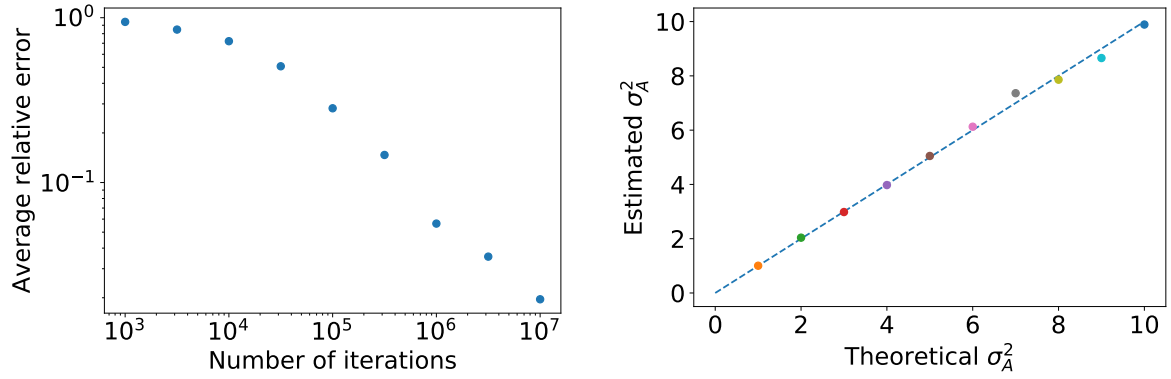


Figure 5.10: (Left) Convergence of the proposed estimator as a function of the number N of iterates for the Ornstein-Uhlenbeck process with quadratic observable. Parameters: $\gamma = 1$, $\sigma = 1$, $\Delta t = 0.01$, $n_{\text{ARE}} = 40$ and $\nu = 8$. (Right) Proposed estimator approximating a range σ_A^2 values of an Ornstein-Uhlenbeck process with quadratic observable. Parameters: $\gamma = [1, 1, 1, 1, 1, 1, 1, 1, 1, 1]$, $\sigma = [1.1892, 1.4142, 1.5651, 1.6818, 1.7783, 1.8612, 1.9343, 2.0000, 2.0598, 2.1147]$, $T = 100000$, $\Delta t = 0.01$ and $\nu = 8$

As before, we generalize this simulation and estimate a range of σ_A^2 values of the Ornstein-Uhlenbeck process. From (2.129), we determine a combination of γ and σ values that result in the desired theoretical asymptotic variance of an Ornstein-Uhlenbeck process and estimate it using our proposed estimator. We use $\gamma = [1, 1, 1, 1, 1, 1, 1, 1, 1, 1]$, $\sigma = [1.1892, 1.4142, 1.5651, 1.6818, 1.7783, 1.8612, 1.9343, 2.0000, 2.0598, 2.1147]$, $T = 100000$, $\Delta t = 0.01$, $q^0 = 0$, $\nu = 8$ (Fig. 5.10 (right)). These results show that $\hat{\sigma}_{A,N}^2$ accurately approximates the smaller asymptotic variances of the tested range. There are some inaccuracies for the larger asymptotic variances, from $\sigma_A^2 = 7$ and upwards. However, increasing the number of observations will increase the accuracy of these estimates.

To conclude the chapter, we compare the convergence of the proposed estimator and the batch means method. We use $\gamma = 1$, $\sigma = 1$, $\Delta t = 0.01$ and $n_{\text{ARE}} = 50$. As before, we need to find the optimal annealing constant and batch size for each considered number of observations. From the results shown in Fig. 5.11, we see that both estimators converge to the asymptotic variance at the same rate of approximately 0.4. However, here we see that the batch means method is slightly more accurate than our proposed method for the considered iteration span.

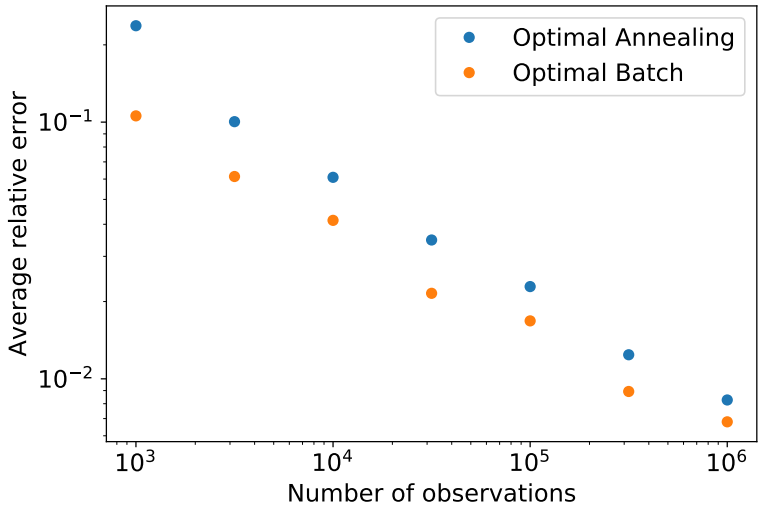


Figure 5.11: Convergence as a function of N of the optimal batch means method and proposed estimator with optimal annealing for the Ornstein-Uhlenbeck process with a quadratic observable. Parameters: $\gamma = 1$, $\sigma = 1$, $\Delta t = 0.01$ and $n_{\text{ARE}} = 50$.

Chapter 6

Conclusions and future research

In this thesis we studied the asymptotic variance of time-integrated functionals of Markov processes. Our primary focus was to propose new estimators of the asymptotic variance that could be implemented in parallel with the evolution of the process.

In Chapter 2 we presented a general introduction to the two types of Markov processes used in the thesis, namely, Markov chains evolving in discrete time and Markov diffusions evolving in continuous time defined by stochastic differential equations. We theoretically determined the asymptotic variance of additive functionals of these processes using the covariance function and validated the results using concepts based on large deviation theory.

In Chapter 3 we introduced and implemented conventional estimation techniques currently used in practice to estimate the asymptotic variance of additive functionals. In particular, we introduced the batch means method, which served as the benchmark for the newly proposed estimators.

Chapter 4 dealt with the theory regarding the Poisson equation associated with a Markov process and its observable. We presented the link between the asymptotic variance and the Poisson solution, which allowed us to express the former in terms of the latter. We also discussed two representations of the Poisson solution that formed the theoretical basis of our proposed estimators.

In Chapter 5 we proposed and evaluated three methods for numerically estimating the asymptotic variance. The first two estimators rely on the fact that the solution to the Poisson equation can be represented as a conditional expectation. For the first of these estimators we simulated many copies or replications of trajectories in order to estimate the conditional expectation. The second method only considers one, long trajectory, from which we extracted the relevant segments to compute the conditional expectation. The third estimator is based on a stochastic approximation of the Poisson solution, which describes the solution as an eigenfunction that can be iteratively estimated in an 'online' way as a simulation unfolds.

We compared these three estimators with the batch means method for simple Markov processes, including the two-state Markov chain and the Ornstein–Uhlenbeck process, for different observables in each case. The results showed that the 'online' estimator produced more accurate estimates of the asymptotic variance for the two-state process

and Ornstein-Uhlenbeck process each with a linear observable. However, we saw that the batch means method was slightly more accurate than the 'online' estimator for the Ornstein-Uhlenbeck process with a quadratic observable.

Based on these results, we recommend in the end the use of the online estimator, based on the stochastic approximation of the Poisson solution, for two reasons. First, it relies on having only one trajectory or stream of data, which is the common situation in practice and second, it is online and so does not require that we scan the trajectory or stream with different time windows or lags.

To conclude this thesis, we suggest possible directions for further research:

- Applications to more realistic and possibly higher-dimensional systems: We tested our estimators on simple processes as a matter of practicality to focus on the construction of these estimators and their relationship with the Poisson equation. In future work, it would be interesting to apply them to more realistic applications involving higher-dimensional processes, as well as processes involving many components or particles. A particularly interesting application is Monte Carlo simulations based on the Metropolis scheme, which is applied in statistical physics to sample the configuration space of many-particle systems.

- Representation of the Poisson solution: We have represented here the Poisson solution on a simple mesh in $\mathbb{R}$, owing to the simple applications that we have used. Other function bases can also be used, e.g., using a Fourier basis or a basis of Hermite polynomials, as described in [47]. One could also represent the Poisson solution as a neural network which iteratively learns from the simulated observations according to some cost function. The choice of representation is particularly important for applications involving high-dimensional state spaces and needs to be balanced in terms of complexity and computational storage.

- Convergence: We have shown with basic examples that the estimators that we have presented converge as the number of data (or replications) is increased. However, we have not presented an exhaustive study of that convergence with the number of observations and of the effect of the initial state on the estimation, that is, the choice of initial condition for the process. In this thesis, we simply considered a random initial state. More work on convergence is needed, especially for the online estimator, as applied to more realistic systems.

- Other types of observables: Most of the literature on asymptotic variance focuses, as we have done here, on sample means, which are additive functionals (in time) of Markov processes. Other types of estimators or observables could be considered. In physics, for example, one often considers physical quantities related not to the state of a process but its jumps or increments in time. These quantities are related to particle or energy currents in physics and can be defined mathematically either as

$$S_n = \frac{1}{n}\sum_{i=1}^{n-1} g(X_i, X_{i+1}) \tag{6.1}$$

if we are dealing with a discrete-time Markov chain or as

$$S_T = \frac{1}{T}\int_0^T g(X_t)dX_t \tag{6.2}$$

if we are dealing with a continuous-time process [26]. In the first case, the observable depends via the function g on the jumps of the process between states X_i and X_{i+1}, whereas in the second case, the observable depends on the increments dX_t of the process X_t. The large deviation techniques used to study these observables are slightly different from those used for additive functionals (see [26] for details) and so are the techniques, we expect, for estimating their asymptotic variance.

Bibliography

[1] L. Wasserman. *All of Statistics*. Springer New York, 2004.

[2] G. Grimmett and D. Stirzaker. *Probability and Random Processes*. Oxford University Press, Oxford, 1982.

[3] M. G. Bulmer. *Principles of Statistics*. Dover Publications, 3rd edition, 1979.

[4] S. Ross. *First Course in Probability*. Pearson, Upper Saddle River, N.J, 9th edition, 2014.

[5] Y. Dai, D. Han, and W. Dai. Modeling and computing of stock index forecasting based on neural network and Markov chain. *The Scientific World Journal*, 2014:1–9, 2014.

[6] S. E. Shreve. *Stochastic Calculus for Finance II: Continuous-time Models*, volume 11. Springer, 2004.

[7] J.T. Schoof, A. Arguez, J. Brolley, and J.J. O'Brien. A new weather generator based on spectral properties of surface air temperatures. *Agricultural and Forest Meteorology*, 135(1-4):241–251, 2005.

[8] D. Rayner, C. Achberger, and D. Chen. A multi-state weather generator for daily precipitation for the Torne River basin, northern Sweden/western Finland. *Advances in Climate Change Research*, 7(1-2):70–81, 2016.

[9] A. Carpinone, M. Giorgio, R. Langella, and A. Testa. Markov chain modeling for very-short-term wind power forecasting. *Electric Power Systems Research*, 122:152–158, 2015.

[10] R. H. M. Zargar and M. H. Y. Moghaddam. Development of a Markov-chain-based solar generation model for smart microgrid energy management system. *IEEE Transactions on Sustainable Energy*, 11(2):736–745, 2020.

[11] G. L. Jones. On the Markov chain central limit theorem. *Probability Surveys*, 1:299–320, 2004.

[12] D. Chauveau and J. Diebolt. Estimation of the asymptotic variance in the CLT for Markov chains. *Stochastic Models*, 19(4):449–465, 2003.

[13] P. Cattiaux, D. Chafai, and A. Guillin. Central limit theorems for additive functionals of ergodic Markov diffusions processes. *Alea*, 9:337–382, 2011.

[14] R. H. Jones. Estimating the variance of time averages. *Journal of Applied Meteorology*, 14(2):159–163, 1975.

[15] C. Alexopoulos. A comprehensive review of methods for simulation output analysis. In *Proceedings of the 2006 Winter Simulation Conference*. IEEE, 2006.

[16] P. W. Glynn and W. Whitt. Estimating the asymptotic variance with batch means. *Operations Research Letters*, 10(8):431–435, 1991.

[17] C. Alexopoulos, N. T. Argon, D. Goldsman, G. Tokol, and J. R. Wilson. Overlapping variance estimators for simulation. *Operations Research*, 55(6):1090–1103, 2007.

[18] P. W. Glynn and D. L. Iglehart. Simulation output analysis using standardized time series. *Mathematics of Operations Research*, 15(1):1–16, 1990.

[19] P. Bratley, B. L. Fox, and L. E. Schrage. *A Guide to Simulation*. Springer New York, 1987.

[20] K. De Turck, S. De Clercq, S. Wittevrongel, H. Bruneel, and D. Fiems. Transform-domain solutions of Poisson's equation with applications to the asymptotic variance. In *Analytical and Stochastic Modeling Techniques and Applications*, pages 227–239. Springer Berlin Heidelberg, 2012.

[21] E. B. Dynkin. *Markov Processes*, volume 2 of *Grundlehren der Mathematischen Wissenschaften ; 121-122*. Springer, Berlin, 1965.

[22] E. Pardoux and A. Y. Veretennikov. On the Poisson equation and diffusion approximation. I. *The Annals of Probability*, 29(3):1061–1085, 2001.

[23] S. Brooks, A. Gelman, G. Jones, and X. Meng, editors. *Handbook of Markov Chain Monte Carlo*. Chapman and Hall/CRC, 2011.

[24] G. Roberts and J. Rosenthal. Geometric ergodicity and hybrid Markov chains. *Electronic Communications in Probability*, 2(0):13–25, 1997.

[25] H. Touchette. The large deviation approach to statistical mechanics. *Physics Reports*, 478(1-3):1–69, 2009.

[26] H. Touchette. Introduction to dynamical large deviations of Markov processes. *Physica A: Statistical Mechanics and its Applications*, 504:5–19, 2018.

[27] A. Dembo and O. Zeitouni. *Large Deviations Techniques and Applications*. Springer Berlin Heidelberg, 2010.

[28] S. P. Meyn, R. L. Tweedie, and P. W. Glynn. *Markov Chains and Stochastic Stability*. Cambridge University Press, 2nd edition, 2009.

[29] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21(6):1087–1092, 1953.

[30] C. Sherlock, P. Fearnhead, and G. O. Roberts. The random walk Metropolis: linking theory and practice through a case study. *Statistical Science*, 25(2):172–190, 2010.

[31] H. Risken. *The Fokker-Planck Equation*. Springer Berlin Heidelberg, 1989.

[32] M.B. Priestley. *Spectral Analysis and Time Series*. Number 1 in Probability and mathematical statistics : A series of monographs and textbooks. Academic Press, 1981.

[33] H. Touchette. A basic introduction to large deviations: Theory, applications, simulations, 2011.

[34] D. J. Griffiths and D. F. Schroeter. *Introduction to Quantum Mechanics*. Cambridge University Press, 2018.

[35] W. Whitt. Asymptotic formulas for Markov processes with applications to simulation. *Operations Research*, 40(2):279–291, 1992.

[36] C. J. Geyer. Practical Markov Chain Monte Carlo. *Statistical Science*, 7(4):473–483, 1992.

[37] M. S. Meketon and B. Schmeiser. Overlapping batch means: something for nothing? In *Proceedings of the 16th Conference on Winter Simulation - WSC '84*. IEEE, 1984.

[38] C. Alexopoulos. Statistical analysis of simulation output: State of the art. In *2007 Winter Simulation Conference*. IEEE, 2007.

[39] D. Goldsman and B. W. Schmeiser. Computational efficiency of batching methods. In *Proceedings of the 29th Conference on Winter Simulation - WSC '97*. ACM Press, 1997.

[40] P. D. Welch. On the relationship between batch means, overlapping means and spectral estimation. In *Proceedings of the 19th Conference on Winter Simulation - WSC '87*. ACM Press, 1987.

[41] B. L. Fox, D. Goldsman, and J. J. Swain. Spaced batch means. *Operations Research Letters*, 10(5):255–263, 1991.

[42] P. W. Glynn and S. P. Meyn. A Liapounov bound for solutions of the Poisson equation. *The Annals of Probability*, 24(2):916–931, 1996.

[43] P. W. Glynn. Poisson's equation for the recurrent M/G/1 queue. *Advances in Applied Probability*, 26(4):1044–1062, 1994.

[44] S. P. Meyn. Large deviation asymptotics and control variates for simulating large functions. *The Annals of Applied Probability*, 16(1):310–339, 2006.

[45] I. Kontoyiannis and S. P. Meyn. Spectral theory and limit theorems for geometrically ergodic Markov processes. *The Annals of Applied Probability*, 13(1):304–362, 2003.

[46] I. Kontoyiannis and S. P. Meyn. Large deviations asymptotics and the spectral theory of multiplicatively regular Markov processes. *Electronic Journal of Probability*, Oct(0):61–123, 2005.

[47] Grégoire Ferré and Hugo Touchette. Adaptive sampling of large deviations. *Journal of Statistical Physics*, 172(6):1525–1544, 2018.

[48] R. Chetrite and H. Touchette. Nonequilibrium Markov processes conditioned on large deviations. *Annales Henri Poincaré*, 16(9):2005–2057, 2014.