Lecture notes for Numerik IVc - Numerics for Stochastic Processes, Wintersemester 2012/2013. Instructor: Prof. Carsten Hartmann

Scribe: H. Lie

Course outline

1. **Probability theory**
   (a) Some basics: stochastic processes, conditional probabilities and expectations, Markov chains
   References: [MS05, Kle06]
2. **Stochastic differential equations**
   (a) Brownian motion: properties of the paths, Strong Markov Property
   References: [MS05, Øks03, Arn73]
   (b) Stochastic integrals: Itô integrals, Itô calculus, Itô isometry
   References: [MS05, Øks03, Arn73]
   (c) SDEs: existence and uniqueness of solutions, numerical discretisation, applications from physics, biolopgy and finance
   References: [Øks03, Arn73, KP92]
   (d) Misc: Kolmogorov forward and backward PDEs, infinitesimal generators, semigroup theory, stopping times, invariant distributions, Markov Chain Monte Carlo methods for PDEs and SDEs
   References: [Øks03, Arn73, KP92]
3. **Filtering theory (if time permits)**
   (a) Linear filtering: conditional expectation, best approximation, Kalman-Bucy filter for SDEs

   Reference: [Jaz07, Øks03]
4. **Approximation of stochastic processes**
   (a) Spectral theory of Markov chains: infinitesimal generator, metastability, aggregation of Markov chains
   References: [HM05, Sar11]
   (b) Markov jump processes: applications from biology, physiocs and finance, Markov decision processes, control theory
   References: [GHL09]

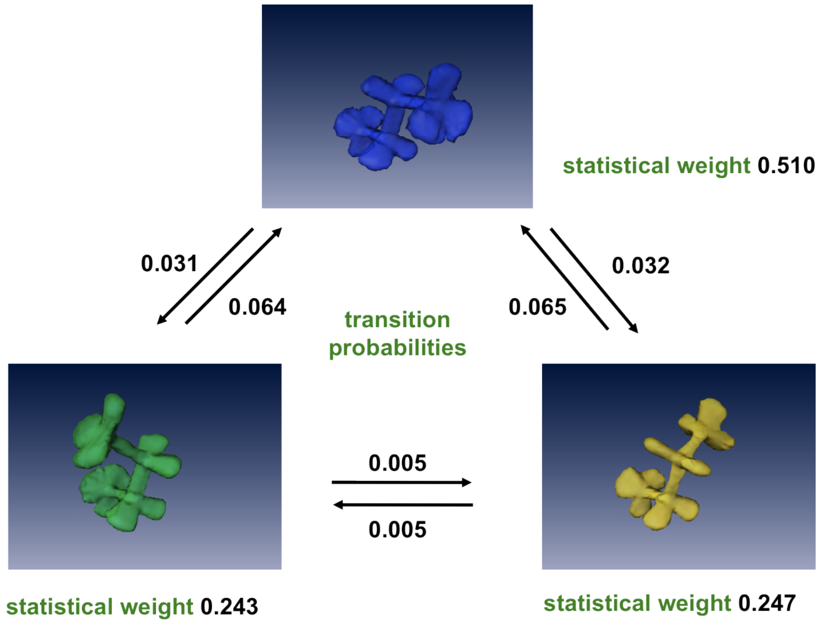

Figure 1: Simulation of a butane molecule and its approximation by a 3-state Markov chain (states in blue, green and yellow; solvent molecules not shown).

1 Day 1, 16.10.2012

1.1 Different levels of modelling

1.1.1 Time-discrete Markov chains

Time index set I is discrete, e.g. $I \subseteq \mathbb{N}$ and state space S is countable or finite, e.g. $S = \{s_1, s_2, s_3\}$ (see Figure 1). Key objects are transition probabilities. For a state space $S = \{1, \ldots, n\}$, the transition probabilities p_{ij} satisfy

$$p_{ij} = \mathbb{P}\left(X_{t+1} = j \mid X_t = i\right)$$

and yield a row-stochastic matrix $P = (p_{ij})_{i,j \in S}$.

1.1.2 Markov jump processes

These are time-continuous, discrete state-space Markov chains. Time index set $I \subseteq \mathbb{R}_+$, S discrete. For a fixed time step $h > 0$, the transition probabilities are given by (see Figure 2)

$$\mathbb{P}\left(X_{t+h} = s_j \mid X_t = s_i\right) = h\ell_{ij} + o(h)$$

where $L = (\ell_{ij})_{i,j \in S}$ and P_h are matrices satisfying $P_h = \exp(hL)$.

Note: the matrix L is row sum zero, i.e. $\sum_j \ell_{ij} = 0$. The waiting times for the Markov chain in any state s_i are exponentially distributed in the sense that

$$\mathbb{P}\left(X_{t+s} = s_i,\ s \in [0, \tau) \mid X_t = s_i\right) = \exp\left(\ell_{ii}\tau\right)$$

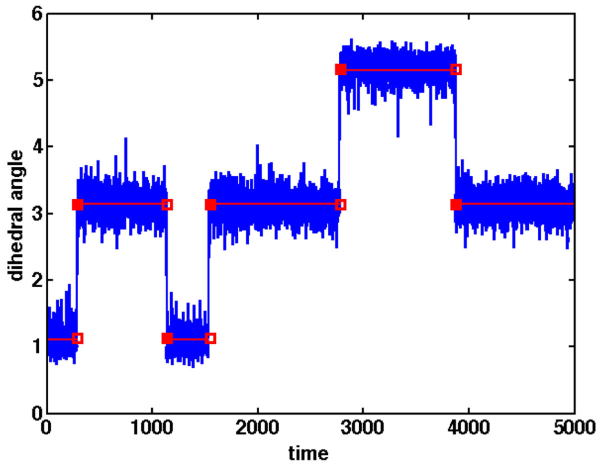

Figure 2: Simulation of butane: typical time series of the central dihedral angle (blue: metastable diffusion process, red: Markov jump process)

and the 'average waiting time' is $-\ell_{ii}$ (by definition of the exponential distribution).

Note: the spectrum of the matrix P_h is contained within the unit disk, i.e. for every eigenvalue λ of P_h, $|\lambda| \leq 1$. This property is a consequence of P_h being row-stochastic, i.e. that $\sum_j P_{h,ij} = 1$. Since $P_h = \exp(hL)$ it follows that

$$\sigma(P_h) \subset D := \left\{x \in \mathbb{R}^2 \mid |x| \leq 1\right\} \Leftrightarrow \sigma(L) \subset \mathbb{C}^- = \{y \in \mathbb{C} \mid \mathrm{Re}(y) \leq 0\}$$

Example 1.1. *Suppose one has a reversible reaction in which one has a large collection of N molecules of the same substance. The molecules can be either in state A or state B and the molecules can change between the two states. Let k^+ denote the rate of the reaction in which molecules change from state A to B and let k^- denote the rate at which molecules change from state B to A.*

For $t > 0$, consider the quantity

$$\mu_i^A(t) := \mathbb{P}\left(\textit{number of molecules in state } A \textit{ at time } t \textit{ is } i\right)$$

where $i = \{0, \ldots, N\}$. One can define quantities $\mu_i^B(t)$ in a similar way, and one can construct balance laws for these quantities, e.g.

$$\frac{d\mu_i^A(t)}{dt} = k^+\mu_{i+1}^A(t) + k^-\mu_{i-1}^A(t) - (k^+ + k^-)\mu_i^A(t).$$

The above balance law can be written in vector notation using a tridiagonal matrix L. By adding an initial condition one can obtain an initial value problem

$$\frac{d\mu^A(t)}{dt} = L^\top \mu^A(t), \quad \mu^A(0) = \mu_0.$$

The solution of the initial value problem above is

$$\mu^A(t) = \mu_0 \exp\left(tL^\top\right).$$

1.1.3 Stochastic differential equations (SDEs)

These are time-continuous, continuous state space Markov chains. SDEs may be considered to be ordinary differential equations (ODEs) with an additional noise term (cf. Figure 2). Let $b : \mathbb{R}^n \to \mathbb{R}^n$ be a smooth vector field and let $x(t)$ be a deterministic dynamical system governed by the vector field $b(\cdot)$. Then $x(t)$ evolves according to

$$\frac{dx}{dt} = b(x), \quad x(0) = x_0. \tag{1}$$

Now let $(B_t)_{t>0}$ be Brownian motion in $\mathbb{R}^d$, and let $(X_t)_{t>0}$ be a dynamical system in $\mathbb{R}^d$ which evolves according to the equation

$$\frac{dX_t}{dt} = b(X_t) + \frac{dB_t}{dt}. \tag{2}$$

The additional term $\frac{dB_t}{dt}$ represents 'noise', or random perturbations from the environment, but is not well-defined because the paths of Brownian motion are nowhere differentiable. Therefore, one sometimes writes

$$dX_t = b(X_t)dt + dB_t,$$

which is shorthand for

$$X_t = X_0 + \int_0^t b(X_t)dt + \int_0^t dB_t.$$

The most common numerical integration method for SDEs is the forward Euler method. If x is a C^1 function of time t, then

$$\left.\frac{dx}{dt}\right|_{t=s} = \lim_{h\to 0} \frac{x(s+h) - x(s)}{h}.$$

The forward Euler method for ODEs of the form (1) is given by

$$X_{t+h} = X_t + hb(X_t)$$

and for SDEs of the form (2) it is given by

$$X_{t+h} = X_t + hb(X_t) + \xi_h$$

where $0 < h \ll 1$ is the integration time step and the noise term ξ in the Euler method for SDEs is modeled by a mean-zero Gaussian random variable.

For stochastic dynamical systems which evolve according to SDEs as in (2), one can consider the probability that a system at some point $x \in \mathbb{R}^d$ will be in a set $A \subset \mathbb{R}^d$ after a short time $h > 0$:

$$\mathbb{P}\left(X_{t+h} \in A \,| X_t = x\right).$$

The associated transition probability density functions of these stochastic dynamical systems are Gaussian because the noise term in (2) is Gaussian.

What has been the generator matrix L in case of a Markov jump process is an infinite-dimensional operator acting on a suitable Banach space. Specifically,

$$Lf(x_0) = \lim_{t\to 0} \frac{\mathbb{E}_{x_0}[f(X_t)] - f(x_0)}{t},$$

provided that the limit exists. Here $f\colon \mathbb{R}^n \to \mathbb{R}$ is any measurable function and $\mathbb{E}_{x_0}[\cdot]$ denotes the expectation over all random paths of X_t satisfying $X_0 = x_0$. L is a second-order differential operator if f is twice differentiable.

2 Day 2, 23.10.2012

Preliminaries from probability theory

Let $(\Omega, \mathcal{E}, \mathbb{P})$ be a probability space, where Ω is a set and $\mathcal{E} \subseteq 2^{\Omega}$ is a σ-field or σ-algebra on Ω, and $\mathbb{P}$ is a probability measure (i.e., $\mathbb{P}$ is a nonnegative, countably additive measure on $(\Omega, \mathcal{E})$ with the property $\mathbb{P}(\Omega) = 1$).

2.1 Conditioning

Let $A \in \mathcal{E}$ be a set of nonzero measure, i.e. $\mathbb{P}(A) > 0$ and define $\mathcal{E}_A$ to be the set of all subsets of A which are elements of $\mathcal{E}$, i.e.

$$\mathcal{E}_A := \{E \subset A \mid E \in \mathcal{E}\}.$$

Definition 2.1 (Conditional probability, part I)**.** *For an event A and an event $E \in \mathcal{E}_A$, the* conditional probability *of E given A is*

$$\mathbb{P}(E|A) := \frac{\mathbb{P}(E \cap A)}{\mathbb{P}(A)}.$$

Remark 2.2. *Think of $\mathbb{P}_A := \mathbb{P}(\cdot\,|A)$ as a probability measure on the measurable space $(A, \mathcal{E}_A)$.*

Given a set $B \in \mathcal{E}$, the *characteristic* or *indicator* function $\chi_B : \Omega \to \{0, 1\}$ satisfies

$$\chi_B(x) = \begin{cases} 1 & x \in B \\ 0 & x \notin B. \end{cases}$$

Definition 2.3 (Conditional expectation, part I)**.** *Let $X : \Omega \to \mathbb{R}$ be a random variable with finite expectation with respect to $\mathbb{P}$. The* conditional expectation *of X given an event A is*

$$\mathbb{E}(X|A) = \frac{\mathbb{E}[X\chi_A]}{\mathbb{P}(A)}.$$

Remark 2.4. *We have*

$$\mathbb{E}(X|A) = \frac{1}{\mathbb{P}(A)} \int_A X d\mathbb{P} = \int X d\mathbb{P}_A.$$

Remark 2.5. *Observe that $\mathbb{P}(E|A) = \mathbb{E}[\chi_E|A]$.*

Up to this point we have only considered the case where A satisfies $\mathbb{P}(A) > 0$. We now consider the general case.

Definition 2.6 (Conditional expectation, part II)**.** *Let $X : \Omega \to \mathbb{R}$ be an integrable random variable with respect to $\mathbb{P}$ and let $\mathcal{F} \subset \mathcal{E}$ be any sub-sigma algebra of $\mathcal{E}$. The* conditional expectation of X given $\mathcal{F}$ *is a random variable $Y := \mathbb{E}[X|\mathcal{F}]$ with the following properties:*

- *Y is measurable with respect to $\mathcal{F}$: $\forall B \in \mathcal{B}(\mathbb{R})$, $Y^{-1}(B) \in \mathcal{F}$.*
- *We have*

$$\int_F X d\mathbb{P} = \int_F Y d\mathbb{P} \quad \forall F \in \mathcal{F}.$$

Remark 2.7. *The second condition in the last definition amounts to the projection property as can be seen by noting that*

$$\mathbb{E}\left[X\chi_F\right] = \int_F X d\mathbb{P} = \int_F Y d\mathbb{P} = \mathbb{E}\left[Y\chi_F\right] = \mathbb{E}\left[\mathbb{E}\left[X|\mathcal{F}\right]\chi_F\right].$$

By the Radon-Nikodym theorem [MS05], the conditional expectation exists and is unique up to $\mathbb{P}$*-null sets.*

Definition 2.8 (Conditional probability, part II)**.** *Define the conditional probability of an event* $E \in \mathcal{E}$ *given* A *by* $\mathbb{P}(E|A) := \mathbb{E}\left[\chi_E|A\right]$

Exercise 2.9. *Let* X*,* $Y : \Omega \to \mathbb{R}$ *and scalars* $a, b \in \mathbb{R}$*. Prove the following properties of the conditional expectation:*

- *(Linearity):*
$$\mathbb{E}\left[aX + bY|A\right] = a\mathbb{E}\left[X|A\right] + b\mathbb{E}\left[Y|A\right].$$
- *(Law of total expectation):*
$$\mathbb{E}\left[X\right] = \mathbb{E}\left[X|A\right] + \mathbb{P}(A) + \mathbb{E}\left[X|A^c\right]\mathbb{P}(A^c)$$
- *(Law of total probability):*
$$\mathbb{P}(B) = \mathbb{P}(B|A)\mathbb{P}(A) + \mathbb{P}(B|A^c)\mathbb{P}(A^c).$$

Example 2.10. *The following is a collection of standard examples.*

- *Gaussian random variables: Let* X_1*,* X_2 *be jointly Gaussian with distribution* $N(\mu, \Sigma)$*, where*
$$\mu = \begin{pmatrix} \mathbb{E}[X_1] \\ \mathbb{E}[X_2] \end{pmatrix}, \quad \Sigma = \begin{pmatrix} a & b \\ b & c \end{pmatrix}$$
such that Σ *is positive definite. The density of the distribution is*
$$\rho(x) = \frac{1}{\sqrt{det(2\pi\Sigma)}} \exp\left[-\frac{1}{2}\left(x - \mu\right)^\top \Sigma \left(x - \mu\right)\right]$$
(Ex.: Compute the distribution of X_1 *given that* $X_2 = a$ *for some* $a \in \mathbb{R}$*.)*
- *(Conditioning as coarse-graining): Let* $Z = \{Z_i\}_{i=1}^M$ *be a partition of* Ω*, i.e.* $\Omega = \cup_{i=1}^M Z_i$ *with* $Z_i \cap Z_j = \emptyset$ *and define*
$$Y(\omega) = \sum_{i=1}^M \mathbb{E}\left[X \ |Z_i\right]\chi_{Z_i}(\omega).$$
Then $Y = \mathbb{E}\left[X|Z\right]$ *is a conditional expectation (cf. Figure 3)*
- *(Exponential waiting times): exponential waiting times are random variables* $T : \Omega \to [0, \infty)$ *with the* memoryless property*:*
$$\mathbb{P}\left(T > s + t \ |T > s\right) = \mathbb{P}\left(T > t\right).$$
This property is equivalent to the statement that T *has an exponential distribution, i.e. that* $\mathbb{P}(T > t) = \exp(-\lambda t)$ *for a parameter value* $\lambda > 0$*.*

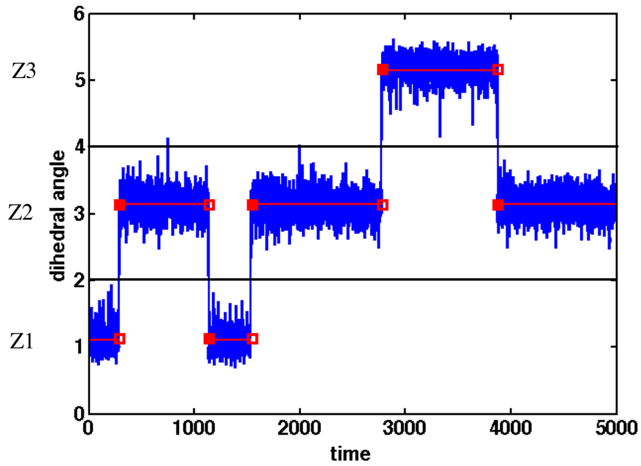

Figure 3: Simulation of butane, coarse-grained into three states Z_1, Z_2, Z_3.

2.2 Stochastic processes

Definition 2.11 (Stochastic process). *A stochastic process $X = \{X_t\}_{t\in I}$ is a collection of random variables on a probability space $(\Omega, \mathcal{E}, \mathbb{P})$ indexed by a parameter $t \in I \subseteq [0, \infty)$. We call X*

- discrete *in time if $I \subseteq \mathbb{N}_0$*
- continuous *in time if $I = [0, T]$ for any $T < \infty$.*

How does one define probabilities for X? We provide a basic argument to illustrate the possible difficulties in defining the probability of a stochastic process in an unambiguous way. By definition of a stochastic process, $X_t = X_t(\omega)$ is measurable for every fixed $t \in I$, but if one has an event of the form

$$E = \{\omega \in \Omega \mid X_t(\omega) \in [a, b] \ \forall t \in I\}$$

how does one define the probability of this event? If t is discrete, the σ-additivity of $\mathbb{P}$ saves us, together with the measurability of X_t for every t. If, however, the process is time-continuous, X_t is defined only almost surely (a.s.) and we are free to change X_t on a set A_t with $\mathbb{P}(A_t) = 0$. By this method we can change X_t on $A = \cup_{t\in I} A_t$, The problem now is that $\mathbb{P}(A)$ need not be equal to zero even though $\mathbb{P}(A_t) = 0 \ \forall t \in I$. Furthermore, $\mathbb{P}(E)$ may not be uniquely defined. So what can we do? The solution to the question of how to define probabilities for stochastic processes is to use finite-dimensional distributions or marginals.

Definition 2.12. *(Finite dimensional distributions): Fix $d \in \mathbb{N}$, $t_1, \ldots, t_d \in I$. The* finite-dimensional distributions *of the stochastic process X for $(t_1, \ldots, t_d)$ are defined as*

$$\mu_{t_1,\ldots,t_d}(B) := \mathbb{P}_{(X_{t_k})_{k=1,\ldots,d}}(B) = \mathbb{P}\left(\{\omega \in \Omega \mid (X_{t_1}(\omega), \ldots, X_{t_d}(\omega)) \in B\}\right)$$

for $B \in \mathcal{B}(\mathbb{R}^d)$.

Here and in the following we use the shorthand notation $\mathbb{P}_Y := \mathbb{P} \circ Y^{-1}$ to denote the *push forward* of $\mathbb{P}$ by the random variable Y.

Theorem 2.13. *(Kolmogorov Extension Theorem): Fix $d \in \mathbb{N}$, $t_1, \ldots, t_d \in I$, and let $\mu_{t_1,\ldots,t_d}$ be a* consistent *family of finite-dimensional distributions, i.e.*

- *for any permutation π of $(1, \ldots, d)$,*

$$\mu_{t_1,\ldots,t_d}(B_1 \times \ldots B_d) = \mu_{(t_{\pi(1)},\ldots,t_{\pi(d)})}(B_{\pi(1)} \times \ldots \times B_{\pi(d)})$$

- *For $t_1, \ldots, t_{d+1} \in I$, we have that*

$$\mu_{t_1,\ldots,t_{d+1}}(B_1 \times \ldots B_d \times \mathbb{R}) = \mu_{t_1,\ldots,t_d}(B_1 \times \ldots \times B_d).$$

Then there exists a stochastic process $X = (X_t)_{t\in I}$ with $\mu_{t_1,\ldots t_d}$ as its finite-dimensional distribution.

Remark 2.14. *The Kolmogorov Extension Theorem does not guarantee uniqueness, not even $\mathbb{P}$-a.s. uniqueness, and, as we will see later on, such a kind of uniqueness would not be a desirable property of a stochastic process.*

Definition 2.15. *(Filtration generated by a stochastic process X): Let $\mathcal{F} = \{\mathcal{F}_t\}_{t\in I}$ with $\mathcal{F}_s \subset \mathcal{F}_t$ for $s < t$ be a filtration generated by $\mathcal{F}_t = \sigma(\{X_s \mid s \leq t\})$ is called the filtration generated by X.*

2.3 Markov processes

Definition 2.16. *A stochastic process X is a* Markov process *if*

$$\mathbb{P}(X_{t+s} \in A \mid \mathcal{F}_s) = \mathbb{P}(X_{t+s} \in A \mid X_s) \tag{3}$$

where

$$\mathbb{P}(\cdot|X_s) := \mathbb{P}(\cdot|\sigma(X_s)),$$
$$\mathbb{P}(E|\sigma(X_s)) := \mathbb{E}[\chi_E \mid \sigma(X_s)]$$

for some event E.

Remark 2.17. *If I is discrete, then X is a Markov process if*

$$\mathbb{P}(X_{n+1} \in A \mid X_0 = x_0, \ldots, X_n = x_n) = \mathbb{P}(X_{n+1} \in A \mid X_n = x_n)$$

Example 2.18. *Consider a Markov Chain $(X_t)_{t\in\mathbb{N}_0}$ on a continuous state space $S \subset \mathbb{R}$ and let $\mathcal{S}$ be a σ-algebra on S. Let the evolution of $(X_t)_{t\in\mathbb{N}_0}$ be described by the* transition kernel $p(\cdot,\cdot) : S \times \mathcal{S} \to [0, 1]$ *which gives the single-step transition probabilities:*

$$p(x, A) := \mathbb{P}(X_{t+1} \in A \mid X_t = x)$$
$$= \int_A q(x, y)dy.$$

In the above, $A \in \mathcal{B}(S)$ and $q = \frac{d\mathbb{P}}{d\lambda}$ is the density of the transition kernel with respect to Lebesgue measure. The transition kernel has the property that

$\forall x \in S$, $p(x, \cdot)$ is a probability measure on $\mathcal{S}$, while for every $A \in \mathcal{S}$, $p(\cdot, A)$ is a measurable function on S.

For a concrete example, consider the Euler-Maruyama discretization of an SDE for a fixed time step Δt,

$$X_{n+1} = X_n + \sqrt{\Delta t}\xi_{n+1}, \quad X_0 = 0,$$

where $(\xi_i)_{i\in\mathbb{N}}$ are independent, identically distributed (i.i.d) Gaussian $N(0,1)$ random variables. The process $(X_i)_{i\in\mathbb{N}}$ is a Markov Chain on $\mathbb{R}$. The transition kernel $p(x, A)$ has the Gaussian transition density

$$q(x, y) = \frac{1}{\sqrt{2\pi\Delta t}} \exp\left[-\frac{1}{2}\frac{|y-x|^2}{\Delta t}\right].$$

Thus, if $X_n = x$, then the probability that $X_{n+1} \in A \subset \mathbb{R}$ is given by

$$\mathbb{P}\left(X_{n+1} \in A | X_n = x\right) = \int_A q(x, y) dy.$$

References

[Arn73] Ludwig Arnold. *Stochastische Differentialgleichungen: Theorie und Anwendung.* Wiley & Sons, 1973.

[GHL09] Xianping Guo and Onésimo Hernández-Lerma. *Continuous-Time Markov Decision Processes: Theory and Applications.* Springer, 2009.

[HM05] Wilhelm Huisinga and Eike Meerbach. *Markov chains for everybody.* Lecture Notes, Fachbereich Mathematik und Informatik, Freie Universität Berlin, 2005.

[Jaz07] Andrew H. Jazwinski. *Stochastic Processes and Filtering Theory.* Dover, 2007.

[Kle06] Achim Klenke. *Wahrscheinlichkeitstheorie.* Springer, 2006.

[KP92] Peter E. Kloeden and Eckhard Platen. *Numerical Solution of Stochastic Differential Equations.* Springer, 1992.

[MS05] David Meintrup and Stefan Schäffler. *Stochastik.* Springer, 2005.

[Øks03] Bernt K. Øksendal. *Stochastic Differential Equations: An Introduction With Applications.* Springer, 2003.

[Sar11] Marco Sarich. *Multiscale Stochastic Processes.* Lecture Notes, Fachbereich Mathematik und Informatik, Freie Universität Berlin, 2011.