Lecture Notes: Modelling of Stochastic Processes
Application in Multi-Agent Systems

Olaf Bochmann

draft: June 1, 2006

Preface: Stochastic Processes in Multi-Agent Applications

This is intended to be a course in advanced probability of stochastic processes. The course provides the theoretical foundation for multi-agent modeling and is targeting graduate students in engineering and computer-science. This course is based on a lecture on advances probability by Shalizi [2006] as it is taught at Carnegie Mellon University. It is assumed that you have already experiences in building probability models. That you have studied a range of important stochastic processes (including Markov Processes, Martingales and diffusions Processes).

LectureNotes.tex 69 2006-06-01 13:13:41Z olaf

Contents

LectureNotes.tex 69 2006-06-01 13:13:41Z olaf

List of Figures

List of Tables

Chapter 1

Basic Definitions

This chapter defines basic terms of the measure theoretic and probability framework and uses them to introduce the stochastic process as a collection of random variables and as random functions.

1.1 Measure Theory

We generally think of a σ-algebra as representing available information. For any event $f \in \mathcal{F}$, we can answer the question "did f happen?"

def:sigmaAlg **Definition 1.1.1** (σ-Algebra)**.** The σ-algebra over a set Ω is a nonempty collection of subsets Σ of Ω such that the following hold:

1. the empty set $\emptyset$ is in Σ
2. if E is in Σ, then the complement of $\Omega \setminus E$ is in Σ
3. if F is a countable collection of sets in Σ, then the union of all the sets $\bigcup F_i$ is also in Σ

From 1 and 2 it follows that Ω is in Σ; from 2 and 3 it follows that the σ-algebra is also closed under countable intersections.

Example 1.1.1. Let $\Omega = \{1, 2, 3\}$, then $(\Omega, \Sigma) = \{\emptyset, \Omega\}$ is the most trivial σ-algebra; the power-set $\mathcal{P}(\Omega)$ is the most profound σ-algebra over a countable set Ω: $(\Omega, \Sigma) = \{\emptyset, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{2, 3\}, \{1, 3\}, \Omega\}$.

def:sigmaOp **Definition 1.1.2** (σ-Operator)**.** For an arbitrary subset C of the power-set $\mathcal{P}(\Omega)$ a σ-operator is defined such that $\sigma(C) := \bigcap_{\Sigma \in \mathcal{F}(C)} \Sigma$, with $\mathcal{F}(C) = \{\Sigma \subseteq \mathcal{P}(\Omega) : C \subseteq \Sigma\}$. $\sigma(C)$ is the smallest σ-algebra[1] generated by C.

1@@vr [1] Definition def:sigmaAlg 1.1.1

Example 1.1.2. Let $\Omega = \{1, 2, 3\}$ and $C = \{1\}$ be an arbitrary family of subsets of Ω, then $\sigma(C) = \{\emptyset, \{1\}, \{2, 3\}, \Omega\}$ is the σ-algebra generated by C.

Question. Generate a σ-algebra over an open set (e.g. a topological space, Euclidean space)

def:measurableSet **Definition 1.1.3** (Measurable Sets)**.** The elements of the σ-algebra[2] are called measurable sets E.

def:measurableSp **Definition 1.1.4** (Measurable Space)**.** An ordered pair (Ω, Σ), where Ω is a set and Σ is a σ-algebra[3] over Ω, is called a measurable space.

def:measurableFct **Definition 1.1.5** (Measurable Function)**.** A function f between a measurable space[4] (Ω_1, Σ_1) and a measurable space (Ω_2, Σ_2) is called measurable if the inverse image of every measurable set is measurable. Those $f : \Omega_1 \mapsto \Omega_2$, if $\forall E \in \Sigma_2 : f^{-1}(E) \in \Sigma_1$.

Example 1.1.3. A stochastic process is stationary if the domain of the sample functions is a time interval and all the time-shift transformations are stationary. If given a stationary or ergodic transformation f^t for all time t where $(\forall t, t_1)(f^t = f^{t_1} \circ f^{t-t_1})$, a stationary or stationary ergodic process X can be constructed by defining a measurable function X_0, composing it with f^t. i.e. $X(\omega) := X_0(f^t(\omega)) \forall \omega \in \Omega$ and so X maps the sample space to a functional space with domain t.

def:measure **Definition 1.1.6** (Measure)**.** A measure μ is a function from a σ-algebra[5] Σ over a set Ω to a value in the interval $[0, \infty]$, e,g, a "size", "volume", or "probability". Properties are:

- $\mu(\emptyset) = 0$;
- $\mu\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} \mu(E_i)$ if $E_1, E_2, E_3 \ldots$ are disjoint sets in Σ.

def:measureSp **Definition 1.1.7** (Measure Space)**.** The triple (Ω, Σ, μ) is called a measure space.

def:borelAlg **Definition 1.1.8** (Borel Algebra)**.** The Borel algebra is a σ-algebra[6] over a set Ω associated to the topology of Ω. For a given topology, the Borel

2@@vr [2]Definition def:sigmaAlg 1.1.1 on the previous page
3@@vr [3]Definition def:sigmaAlg 1.1.1 on the preceding page
4@@vr [4]Definition def:measurableSp 1.1.4
5@@vr [5]Definition def:sigmaAlg 1.1.1 on the preceding page
6@@vr [6]Definition def:sigmaAlg 1.1.1 on the previous page

algebra is the σ-operator[7] $\sigma(C)$ - the smallest σ-algebra generated by the set of the closed intervals in Ω: $C = \{[c,d] : [c,d] \in \Omega\}$, if Ω is an arbitrary interval $\Omega = \langle a, b\rangle$.

def:borelSet **Definition 1.1.9** (Borel Sets)**.** The elements of the Borel algebra[8] are called Borel sets.

Example 1.1.4. Countable unions and countable cuts of sets are Borel sets.

def:pDimBorelSet **Definition 1.1.10** (p-Dimensional Borel Set)**.** If Ω is an undefined interval in a p-dimensional space $\Omega = \langle a_1, b_1\rangle \times \langle a_2, b_2\rangle \times \cdots \times \langle a_p, b_p\rangle \subset \mathbb{R}^p$ and C is the set of closed intervalls in Ω: $C = \{[c_1, d_1] \times \cdots \times [c_p, d_p] : \cdots \subset \Omega\}$, then the σ-operator[9] $\sigma(C)$ is the p-dimensional Borel set.

def:cylinderSet **Definition 1.1.11** (Cylinder Set)**.** If Ω_0 is a finite set of q elements (e.g. an alphabet) $\Omega_0 = \{x_1, \dots, x_q\}$, Ω is a sequence of this elements (e.g. words) $\Omega = \Omega_0^N$ and $(\omega_1 \dots \omega_n) \in \Omega_0^N$ is a n element long prefix, then the cartesian product of that prefix and the rest of the sequence $Z(\omega_1 \dots \omega_n) = \{(\omega_1 \dots \omega_n)\} \times \Omega_0^{N \setminus \{1,\dots,n\}}$ is the cylinder set.

Example 1.1.5. If $\Omega_0 = \{a, b, \dots, z\}$, $n = 4$, $(\omega_1 \dots \omega_4) = (l, o, v, e)$, then $Z(\omega_1 \dots \omega_4) = \{\omega : \omega = (l, o, v, e, \omega_5, \omega_6, \dots)\}$ is the cylinder set (the set of all word in the sequence with the prefix "love")[10].

def:cylinderAlg **Definition 1.1.12** (Cylinder Algebra)**.** If C is the cylinder set[11] $C = Z(\omega_1 \dots \omega_n)$, then the σ-operator[12] $\sigma(C)$ is the cylinder algebra.

A filtration is a way of representing our information about a system growing over time. To understand right-continuity, imagine we fail to answer the question "did f happen?" This would be the case if $\mathcal{F}_t \subset \bigcap_{s>t} \mathcal{F}_s$. Then there would have to be events which were detectable at all times after t, but not at t itself — a sudden jump in our information. This is what right-continuity rules out. Therefore filtration is a growing collection of σ-fields:

def:filtration **Definition 1.1.13** (Filtration)**.** Let T be an ordered index set. A collection $\mathcal{F}_t \in T$ of σ-algebras is a filtration (with respect to its order) if it is non-decreasing, i.e., $f \in \mathcal{F}_t$ implies $f \in \mathcal{F}_s$ for all $s > t$ (filtration by $\mathcal{F}$).

In contrast $\mathcal{F}$ is right-continuous if $\mathcal{F} = \mathcal{F}^+$ and $\mathcal{F}_t^+ = \bigcap_{s>t} \mathcal{F}_s$

[7] Definition 1.1.2 on page 1 (def:sigmaOp)
[8] Definition 1.1.8 on the preceding page (def:borelAlg)
[9] Definition 1.1.2 on page 1 (def:sigmaOp)
[10] You may need this to calculate if the average success rate of an experiment series ($\Omega_0 = \{$"*success*", "*failure*"$\}$) converges to the experiment probability.
[11] Definition 1.1.11 (def:cylinderSet)
[12] Definition 1.1.2 on page 1 (def:sigmaOp)

1.2 Probability Theory

def:probMe **Definition 1.2.1** (Probability Measure)**.** The probability measure P is a certain kind of measure[13]. It is a positive measure on the measurable space[14] (Ω, Σ) such that $P(\Omega) = 1$.

def:probSp **Definition 1.2.2** (Probability Space)**.** The probability space is a certain kind of measure space[15]. It is a triple (Ω, Σ, P); a set Ω, together with a σ-algebra[16] Σ on Ω and a probability measure[17] P on that σ-algebra.

def:event **Definition 1.2.3** (Event)**.** The measurable subsets of Ω, i.e., the sets belonging to the σ-algebra[18] Σ, are called events.

def:prob **Definition 1.2.4** (Probability)**.** $P(E)$ is the probability of the event[19] E.

def:outcome **Definition 1.2.5** (Outcome)**.** If $\omega \in \Omega$, ω is an outcome (of an experiment).

def:sampleSp **Definition 1.2.6** (Sample Space)**.** The set S, mapped from Ω using a measurable function[20] Ω, is called sample space.

def:randV **Definition 1.2.7** (Random Variable)**.** A random variable is a measurable function[21] X defined on sample space[22].

Example 1.2.1. -The people of earth-

Probability space: $\Omega = \{\text{african, american, asian, european}, \dots\}$

$$\text{Random Variable: } X(\omega) = \begin{cases} \omega = \text{african} & N(\mu_1, \sigma^2) \\ \omega = \text{american} & N(\mu_2, \sigma^2) \\ \omega = \text{asian} & N(\mu_3, \sigma^2) \\ \omega = \text{european} & N(\mu_2, \sigma^2) \\ \dots & \dots \end{cases}$$

Sample space: $S = \{\text{height vector}\}$

[13] Definition 1.1.6 on page 2 def:measure
[14] Definition 1.1.4 on page 2 def:measurableSp
[15] Definition 1.1.7 on page 2 def:measureSp
[16] Definition 1.1.1 on page 1 def:sigmaAlg
[17] Definition 1.2.1 def:probMe
[18] Definition 1.1.1 on page 1 def:sigmaAlg
[19] Definition 1.2.3 def:event
[20] Definition 1.1.5 on page 2 def:measurableFct
[21] Definition 1.1.5 on page 2 def:measurableFct
[22] Definition 1.2.6 def:sampleSp

1.3 Stochastic Process

A stochastic process is one whose behavior is non-deterministic in that the next state of the process is not fully determined by the previous state of the process.
Two ways of thinking are used to define a stochastic process:

1. as indexed collection of random variables (RV),
2. and as random functions.

def:stochProc1 **Definition 1.3.1** (Stochastic Process: Collection of RV)**.** A stochastic process $\{\Omega_t\}_{t\in T}$ is a collection of random variables[23] Ω_t, taking values in a common measure space[24] (Ξ, χ), indexed by a set T.

Remark. This means, for each $t \in T$, $\Omega_t(\omega)$ is a measurable function (RV) that maps the probability space Ω into the sample space Ξ, which induces a probability measure $\Pr(\Xi)$ on Ξ.

$$\Omega \longmapsto^{\Omega(\omega)} \Xi$$

Example 1.3.1. Typical examples of stochastic processes are given in table tab:stochProc 1.1. The matrix classifies the examples by the nature of index T and measure space Ξ.

exe:empProc **Example 1.3.2** (Empirical Process)**.** Z_i for $i = 1, 2, \dots$ are independent, identical-distributed real-valued random variables (one-sided real-valued random sequence). For each Borel set B on the reals and each n, there is a inferred probability distribution

$$\hat{P}_n(B) = \frac{1}{n}\sum_{i=1}^{n} 1_B(Z_i),$$

e.g. the fraction of the samples up to time n which fall into the set B. This is the empirical measure. $\hat{P}_n(B)$ is a one-sided random sequence of set functions with $T = B \times N$ and $\Xi = \mathbb{R}$ (probability measures). One would be interested in showing that it converges to the common distribution of the Z_i.

22@@vr [23] Definition def:randV 1.2.7 on the preceding page.
23@@vr [24] Definition def:measureSp 1.1.7 on page 2

Table 1.1: examples of stochastic processes

	$\Xi = \{1, 2, \dots, n\}$	$\Xi = \mathbb{R}$	$\Xi = \mathbb{C}$	$\Xi = \mathbb{R}^k$
$T = \{1\}$	discrete	cont.		
	random variable			
$T = \{1, 2, \dots, k\}$	n^k	$\mathbb{R}^k$		
	random vector			
$T = \{1, 2, \dots\}$	discrete	real	complex	vector
	-valued one-sided random sequence[25]			
$T = \mathbb{Z}$	discrete	real	complex	vector
	-valued two-sided random sequence			
$T = \mathbb{Z}^d$	discrete	real	complex	vector
	d-dimensional spatially-discrete random field			
$T = \mathbb{R}$	discrete	real	complex	vector
	continuous-time random process/signal/motion			

tab:stochProc

1.4 Random Functions

Why would we want to analyze stochastic processes as random elements of spaces of functions? It is because of convenience: geometric intuitions about distance carry over to those abstract spaces as they do for approximation, convergence, orthogonality or any other ideas learned form the study of euclidean spaces Pollard:90 [Pollard, 1990].

We have defined a stochastic process[26] as a collection of random variables $\{\Omega_t\}_{t \in T}$. The index T may be an interval on the real line, but also it may be something more fancy like a subset of a higher-dimensional euclidean space or a collection of functions. It can be more convenient to write $\Omega(t)$ in place of Ω_t. Then $\Omega(t, \omega)$ has two arguments t and ω. For each value of t, $\Omega_t(\omega)$ is a RV. However, for each fixed value of ω, $\Omega_\omega(t)$ is called a *sample path* of the stochastic process. It is a random function that maps from the index set T into the sample space Ξ:

$$T \stackrel{\Omega(t)}{\longmapsto} \Xi.$$

This is just a different perspective with the advantage that it lets us consider the realizations of stochastic processes as single objects, rather than large collections. The advantage is — apart from the implementation issue —

2002vr

[26] Definition def:stochProc1 1.3.1 on the previous page

that it it looks for relations among the variables in the collection or there realizations, rather than just properties of individual variables. In example exe:empProc 1.3.2 we would like to have random probability measures, rather then just random set functions. This would imply dependencies among the random variables in the collection, e.g. a measure must respect finite additivity over disjoint Borel sets $\hat{P}_n(A \cup B) = \hat{P}_n(A) + \hat{P}_n(B)$. To explicitly work out all this dependencies in a collection is tedious. The random function notation offers an effective way to define acceptable realizations of the whole stochastic process.

def:stochProc2 **Definition 1.4.1** (Stochastic Process: as Random Function)**.** A Ξ-valued stochastic process on T with path in U, $U \subseteq \Xi^T$, is a random function $X : \Omega \mapsto U$ which is $F/U \cup \Xi^T$-measurable.

1.5 Notation

- Ω, Ξ be the set
- Σ, χ be the collection of subsets
- E are the elements of Σ (measurable sets)
- μ is the measure
- (Ω, Σ), (Ξ, χ) is the measurable space (sigma algebra)
- (Ω, Σ, μ) is the measure space.

Chapter 2

One-Parameter Processes, a process as function of time, distance or sample size

sec:oneParProc
Section 2.1 defines one-parameter processes, and their variations.

There are three application areas for stochastic processes: dynamics (systems changing over time) spacious (systems changing over space) and inference (conclusions changing as more and more data become available). All of these can be considered as "one-parameter" processes, where the parameter is time, distance or sample size respectively.

2.1 One-Parameter Process

sec:oneParProc

The index set T has in general some topological or geometrical structure. The number of dimensions in this structure corresponds to the number of parameters of the process.

def:oneParProc

Definition 2.1.1 (One-Parameter Process)**.** A process whose index set T has one dimension is a one-parameter process (otherwise it is a multi-parameter process). The process is discrete or continuos depending on weather its index set is countable or not. The process is single-sided if the index set has a minimum (otherwise the process is two-sided).

The best known one-parameter process is inference — repeating an experiment under the same conditions ever and ever again. It is assumed that the outcome can be only "success" or "failure", e.g. drawing a ball from a bag with different colored balls — only drawing a read ball is a "success".

In order to repeat the experiment under the same condition the ball has to be returned into the bag. This is what is called a Bernoulli process.

exe:bernoulliProc **Example 2.1.1** (Bernoulli Process)**.** The Bernoulli process is a one-sided infinite sequence of IID binary variables (e.g. experiments), where $X_t = 1$ with probability p (e.g. "success"), $X_t = 0$ with probability $q = p - 1$ (e.g. "failure"), for all t.

Instances of stochastic processes in dynamics are position and velocity of planets; position and velocity of molecules in gas; pressure, temperature and volume of gas. They are Markov processes.

exe:markovProc **Example 2.1.2** (Markov Process)**.** Markov chains, Markov models of order k and hidden Markov models are discrete-parameter stochastic processes. Continuous-time Markov processes are continuous parameter stochastic processes. They all may be one-sided or two-sided This is an example of dynamics or spacious extension.

exe:whiteNoiseProc **Example 2.1.3** (White Noise Process)**.** For each $t \in \mathbb{R}^+$, let $X_t \sim \mathcal{N}(0, 1)$, all mutually independent of each other. This process has a one-sided continuous parameter.

The next example is a process that plays a similar role in the theory of stochastic processes analogous to Gaussian distribution in elementary probability (limit theorem).

exe:wienerProc **Example 2.1.4** (Wiener Process)**.** The Wiener process is a continuous-parameter process with $T = \mathbb{R}^+$ and $\Xi = \mathbb{R}$ with the following properties:

1. $W(0) = 0$;
2. for any times $t_1 < t_2 < t_3$, $W(t_2) - W(t_1)$ and $W(t_3) - W(t_2)$ are independent;
3. $W(t_2) - W(t_1) \sim \mathcal{N}(0, t_2 - t_1)$;
4. $W(t, \omega)$ is a continuous function of $t$for almost all ω.

There is a large class of deterministic processes that have stochastic properties. One example is the logistic map:

exe:logisticMap **Example 2.1.5** (Logistic Map)**.** The Logistic map is a one-sided discrete process with $T = \mathbb{N}$, $\Xi = [0, 1]$, $X(0) \sim \mathcal{U}(0, 1)$ and $X(t + 1) = aX(t)(1 - X(t))$, $a \in [0, 4]$. All the randomness is in the initial condition. Given the initial condition the process is deterministic.

It can be shown that for certain values of a the process satisfies the laws of large numbers and the central limit theorem. Lets partition the original process of the logistic map $X(t)$ and let $S(t) = 0$ if $X(t) \in [0, 0.5)$ and $S(t) = 1$ if $X(t) \in [0.5, 1]$. Even if $X(t)$ is Markovian, $S(t)$ is not necessarily. We want to know when functions of Markovian processes are themselves Markovian. Another interesting aspect is that the partition $S(t)$ is exactly as informative as the original process $X(t)$ with continuous states. Finally, for a logistic map with $a = 4$ (and $X(0) \neq 0.5$) the symbol sequence of the partition is a Bernoulli process — the deterministic function of a deterministic process provides a model for IID randomness.

2.2 Operator Representation

sec:OneParPrOp

Instead of relating the history of a discrete-parameter process X_t to its preceding values, it may be less complicated to represent the dynamic part of any process as a semi-group of operators.

def:shiftOp

Definition 2.2.1 (Shift Operators)**.** Consider Ξ^t, $T = \mathbb{N}$, $T = \mathbb{Z}$, $T = \mathbb{R}^+$ or $T = \mathbb{R}$. The shift-by-τ operator Σ_τ, $\tau \geq 0$, maps Ξ^t into itself by shifting forward in time: $(\Sigma_\tau)x(t) = x(t + \tau)$. The collection of all shift-operators is the shift semi-group or time-evolution semi-group.

Chapter 3

Markov Processes

Section 3.1 formally defines the Markov property for a one-parameter process. It explains the concepts of complete determinism and complete statistical independence.

Section 3.2 introduces transition probabilities of Markov processes.

The intention is to fix notation and to clarify that the velocity (momentum) plays a crucial role.

3.1 Markov Property

sec:MarkovProp

The Markov property is the independence of the future from the past.

def:MarkovProp **Definition 3.1.1** (Markov Property)**.** A one-parameter process[1] X is a Markov process with respect to filtration[2] $\mathcal{F}$ when X_t is adapted to the filtration, and, for any $s > t$, X_s is independent of $\mathcal{F}_t$ given X_t, $X_s \perp\!\!\!\perp \mathcal{F}_t \mid X_t$.

There are two ways of explaining the Markov property. One, followed by Markov himself, is based on the desire to weaken the assumption of strict statistical independence between variables to mere conditional independence. His motivation was to show that all the key limit theorems of probability[3] hold for Markov processes, as well as for IID variables. The other way to explain the Markov property begins with completely deterministic systems. There is some variable in a dynamic system, called state, that fixes the values of all present and future observables — the present state determines the

[1] Definition ?? on page ??

[2] Definition 1.1.13 on page 3

[3] the weak and the strong laws of large numbers, the central limit theorem, etc.

state of all future times or as generalization the present state determines the distribution of all future states.

It is possible to represent any one-parameter stochastic process X as a noisy function of a Markov process Z. The shift operators[4] give a trivial way of doing this, where the Z process is not just homogeneous but actually fully deterministic[5]. Another approach is to set $Z_t = X_t^-$, the complete past up to an including time t. It can be shown that there is a unique representation where $Z_t = \epsilon(X_t^-)$ for some function ϵ, Z_t is a homogeneous Markov process, and $X_u \perp\!\!\!\perp \sigma(\{X_t, t \leq u\}) \mid Z_t$ Knight-PredTimeProc:75 [Knight, 1975].

3.2 Transition Probability

sec:TransProb

2702vr

[4] Definition def:shiftOp 2.2.1 on page def:shiftOp 11

[5] CSSR is following this approach.

Bibliography

t-PredTimeProc:75 F. B. Knight. A predictive view of continuous time processes. *The Annals of Probability*, (3):573–596, 1975. 3.1

Pollard:90 D. E. Pollard. *Convergence of Stochastic Processes*. Springer-Verlag New York, LLC, New York, n/a edition, Nov. 1990. ISBN 0387909907. URL `http://ameliabedelia.library.yale.edu/dbases/pollard1984.pdf`. 1.4

Shalizi:06 C. Shalizi. Stochastic processes. Lecture Notes on Stochastic Processes (Advanced Probability II), 2006. URL `http://www.stat.cmu.edu/~cshalizi/754/notes/all.pdf`. (document)