Theory of Stochastic Processes
1. Overview

Tomonari Sei
sei@mist.i.u-tokyo.ac.jp

Department of Mathematical Informatics, University of Tokyo

April 6, 2017

http://www.stat.t.u-tokyo.ac.jp/~sei/lec.html

Outline of today's lecture

1. Course plan
2. Introduction of stochastic processes
3. Review of elementary probability theory

Overview

From syllabus:

Course objective Stochastic processes are useful to model random phenomena changing in time. This course is aimed at an introduction to stochastic processes.

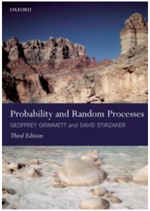
Teaching methods lecture, mainly using black board

Method of evaluation midterm exam 40%, final exam 60%.
(No assignment, but some exercises will be provided every week.)

Notes on taking this course Students are assumed to have basic knowledge of the elementary probability theory that will be reviewed in the first lecture (= today).

Book

- G. Grimmett and D. Stirzaker, *Probability and Random Processes*, 3rd ed., Oxford University Press, 2001.



The book title will be abbreviated as PRP.

- Author's web site:
`http://www.statslab.cam.ac.uk/~grg/books/prp.html`
- Copies of necessary parts will be provided in the class.
- Will be put in the Library (1st floor) possibly in May.

Thanks to Dr. Alfred Kume for his advice on choosing books.

Schedule

We will learn about the following topics week by week.

Apr 6	Overview
Apr 13	Simple random walk
Apr 20	Generating functions
Apr 27	Markov chain
May 11	Continuous-time Markov chain
May 18	Markov chain Monte Carlo
May 25	(midterm exam)
June 8	Stationary processes
June 15	Renewal processes
June 22	Martingales
June 29	Queues
July 6	Diffusion processes
July 13	Review
July 20?	(Final exam)

Note: the order may be changed.

Office hours

Office hours are offered

- Every Tuesday 12:00–14:00 without appointments.
- My office is Room 344 on the 3rd floor of this building.
- Feel free to ask any questions and comments.
- You can also make an appointment at another time by e-mail

  sei@mist.i.u-tokyo.ac.jp

  I am happy to go out for coffee :-)

- Note: On Apr 18 (Tue), it will be reduced to 13:00–14:00.
- Other information will be announced on

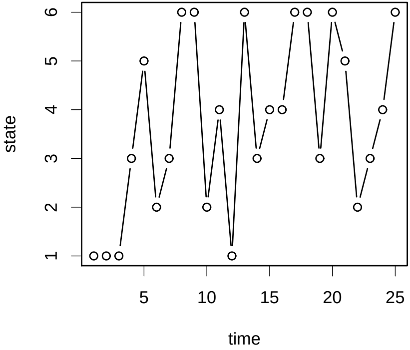
  `http://www.stat.t.u-tokyo.ac.jp/~sei/lec-j.html`

Outline of today's lecture

What is a stochastic process?

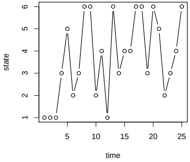
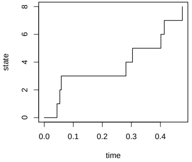
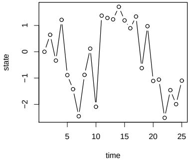
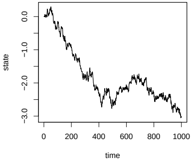
There are two equivalent definitions:

- A set of random variables indexed by time
- A random function of time



An outcome of rolling a dice 25 times.

Classification of stochastic processes

	discrete time	continuous time
discrete state		
continuous state		

Today I introduce only a continuous-time discrete-state process.

An essay written by K. Itô

Go to the next page!

伊藤清「確率論と私」p.57

（上略）しかしどこまで進んでも実在は更に複雑で、科学者の立場からすれば、数学を近似的模型として利用するにすぎない。したがって数学者が苦心して作り上げた厳密な理論などはあまりに顧慮しないで、相当乱暴な数学のつかい方をする。たとえば放射性元素の原子 N 個が時間とともに崩壊して減少して行く状態を

$$\frac{dN(t)}{dt} = -\alpha N(t), \quad N(0) = N$$

という方程式であらわす。ここに $N(t)$ は時間 t の後における原子数で、α は単位時間の崩壊率である。$N(t)$ は整数であるから、「到る所微分不可能な連続関数」の存在すら知っている数学者にとっては、右の方程式は全く我慢のならない代物である。（中略）

1980 年 1 月

An essay written by K. Itô

K. Itô, *Probability theory and I*, p.57. (translated by Sei)

(snip) But reality is always further complicated, and scientists use mathematics just as an approximate model. Therefore, even if mathematicians made great efforts to build a rigorous mathematical theory, it is not much taken into account and is used in a quite rough way. For example, decay of N radioactive elements is described by

$$\frac{dN(t)}{dt} = -\alpha N(t), \quad N(0) = N.$$

Here $N(t)$ is the number of atoms after time t, and α is the decay rate. Since $N(t)$ is an integer, this equation is completely not acceptable for mathematicians, who know "functions that are continuous but nowhere differentiable." (snip)

Jan 1980

Go to the next page!

続き

この問題を数学者が満足するように解くとすれば、次のようになるであろう。各原子が時間 t の後まで生き延びる確率を $p(t)$ とすれば

$$\frac{dp(t)}{dt} = -\alpha p(t), \quad p(0) = 1,$$

これを解いて

$$p(t) = e^{-\alpha t},$$

はじめに与えられた原子に番号をつけて $1, 2, \cdots, N$ とし、原子 n が時間 t の後に生存しているか、否かに応じて

$$X_n(t) = 1 \text{ または } 0$$

とおくと、時間 t の後に生存している総原子数 $N(t)$ は

$$N(t) = \sum_{n=1}^{N} X_n(t)$$

で与えられる確率過程である。

(cont.)

In order to satisfy mathematicians, solve this problem as follows. Let $p(t)$ be the probability that each atom is alive after time t. It satisfies

$$\frac{dp(t)}{dt} = -\alpha p(t), \quad p(0) = 1.$$

The solution is

$$p(t) = e^{-\alpha t}.$$

Let us number the given N atoms from 1 to N, and define

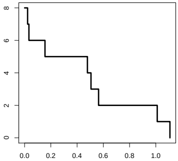
$$X_n(t) = 1 \text{ or } 0$$

according to life or death of n-th atom at time t. Then, the total number $N(t)$ of atoms at time t is a stochastic process

$$N(t) = \sum_{n=1}^{N} X_n(t).$$

The death process

- The above stochastic process $N(t)$ is called the death process.
- A sample path of $N(t)$ is like this:



- The process can be applied to phenomena of decrease.

The world is surrounded by randomness!

Simulation

Let's simulate the death process.

A naive method

For each $1 \leq n \leq N$, let T_n be the time when the n-th atom dies. The distribution of T_n is the exponential distribution:

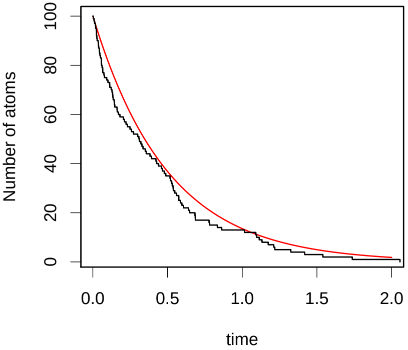
$$P(T_n \geq t) = e^{-\alpha t}.$$

Then, define $X_n(t)$ by

$$X_n(t) = \begin{cases} 1 & \text{if } T_n > t, \\ 0 & \text{otherwise.} \end{cases}$$

This is called the indicator variable of the event $\{T_n > t\}$.

Simulation




```
# In R language
N = 100; alpha = 2
Ts = rexp(N, alpha)
Es = c(0, sort(Ts))
plot(Es, N:0, type="s", xlim=c(0, 2), xlab="time", ylab="Number of atoms")
curve(N * exp(-alpha*x), add=TRUE, col="red")
```

Simulation

Another method of simulation

Let $U_1, \dots, U_N$ be independent random variables with

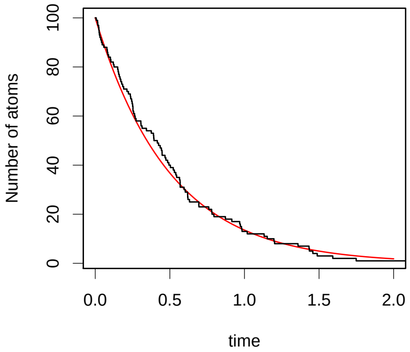
$$P(U_k \geq t) = e^{-(N+1-k)\alpha t}.$$

Then, $\sum_{n=1}^{N} X_n(t)$ has the same distribution as

$$N(t) = N - \max\{n \mid \sum_{k=1}^{n} U_k \leq t\}$$

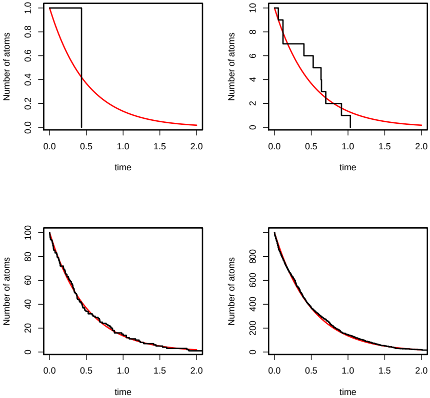
Exercise: Why does this method work?

Simulation




```
Us = rexp(N, (N:1)*alpha)
Es = c(0, cumsum(Us))
plot(Es, N:0, type="s", xlim=c(0, 2), xlab="time", ylab="Number of atoms")
curve(N * exp(-alpha*x), add=TRUE, col="red")
```

Convergence: law of large numbers



$$\lim_{N\to\infty} P\left(\left|\frac{N(t)}{N} - e^{-\alpha t}\right| > \varepsilon\right) = 0 \quad \text{for each } t.$$

We will touch on this kind of limit theorems at a later date.

Questions on the death process

Remark: There are some probabilistic (or analytic) questions:

- Is $N(t)$ constructed above right-continuous?
- Furthermore, is it càdlàg (continue à droite, limite à gauche = right continuous with left limit)?
- Does $N(t)$ converge to 0 as $t \to \infty$ with probability 1?
- Does it hold that

$$\lim_{N\to\infty} P\left(\sup_{t\geq 0}\left|\frac{N(t)}{N} - e^{-\alpha t}\right| > \epsilon\right) = 0?$$

  You may not need to take this course if you can answer this question immediately...

More about the process:

- What is the time complexity of the simulating methods?
- How to obtain the likelihood function with respect to the parameter α if $N(t_1), \ldots, N(t_k)$ are observed?

Remark: The Monte Carlo method

- A concept related to simulation is the Monte Carlo method. This is a quite powerful tool for evaluation of expectations.
- For example, an approximate value (with error estimate) of

$$\mathrm{E}\left[X_1 + \frac{1}{X_2 + \frac{1}{X_3 + \frac{1}{X_4 + \frac{1}{X_5}}}}\right]$$

  for independent uniform random variables $X_1, \ldots, X_5$ on $[1, 2]$ will be easily obtained.
- For more complicated problems, we may need the Markov Chain Monte Carlo (MCMC) methods.
- We will learn about them at a later date.

Outline of today's lecture

1. Course plan
2. Introduction of stochastic processes
3. Review of elementary probability theory

Review of elementary probability theory

The first 3 chapters of the book “PRP” are

- Chapter 1: Events and their probabilities
- Chapter 2: Random variables and their distributions
- Chapter 3: Discrete random variables

Recommended problems:

- §1.8, Problems 20, 22, 28*, 30 and 37.
- §2.7, Problems 4, 12* and 13*.
- §3.11, Problems 5, 6, 7, 16, 17, 33* and 34*.

The asterisk (*) shows difficulty.

Work out!

Review of Chapter 1

Definition

Let Ω be a set. A collection $\mathcal{F}$ of subsets of Ω is called a σ-field if it satisfies

1. $\emptyset \in \mathcal{F}$;
2. $A_1, A_2, \ldots \in \mathcal{F} \Rightarrow \cup_{i=1}^{\infty} A_i \in \mathcal{F}$;
3. $A \in \mathcal{F} \Rightarrow A^{\mathrm{c}} \in \mathcal{F}$.

The set Ω is called the sample space. Each $A \in \mathcal{F}$ is called an event.

Definition

A probability measure P on $(\Omega, \mathcal{F})$ is a function $P : \mathcal{F} \to [0, 1]$ such that

- $P(\Omega) = 1$;
- $A_1, A_2, \ldots \in \mathcal{F}$ and $A_i \cap A_j = \emptyset$ for any $i \neq j$
  $\Rightarrow P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$.

The triplet $(\Omega, \mathcal{F}, P)$ is called a probability space.

Definition

Let A and B be events. If $P(B) > 0$, then the conditional probability that A occurs given that B occurs is defined to be

$$P(A|B) = \frac{P(A \cap B)}{P(B)}.$$

Definition

Events A and B are called independent if

$$P(A \cap B) = P(A)P(B).$$

More generally, events $A_1, \ldots, A_n$ are called independent if

$$P(\cap_{i \in J} A_i) = \prod_{i \in J} P(A_i)$$

for any subset $J \subset \{1, \ldots, n\}$. (Note: we interpret $\subset$ as $\subseteq$.)

Review of Chapter 2

Definition

A random variable X on a probability space $(\Omega, \mathcal{F}, P)$ is a function $X : \Omega \to \mathbb{R}$ with the property that $\{X \leq a\} = \{\omega \in \Omega \mid X(\omega) \leq a\}$ is a member of $\mathcal{F}$ for any $a \in \mathbb{R}$.

Keep in mind that a random variable is a function!

But, you can sometimes forget about it
when you use the distribution functions:

Definition

The (cumulative) distribution function of a random variable X is defined by $F(x) = F_X(x) = P(X \leq x)$.

Definition

A random variable X is discrete if it takes values only in some countable subset of $\mathbb{R}$. For the discrete random variable X, the (probability) mass function is defined by $f(x) = P(X = x)$.

Remark: $P(X \in A) = \sum_{x \in A} P(X = x)$.

Definition

A random variable X is continuous if its distribution function is written as

$$F(x) = \int_{-\infty}^{x} f(u)du$$

for some function $f(x)$ called the (probability) density function.

Remark: $P(X \in A) = \int_A f(x)dx$.

Definition

The joint distribution function of a random vector $\mathbf{X} = (X_1, \ldots, X_n)$ on the probability space $(\Omega, \mathcal{F}, P)$ is the function

$$F(\mathbf{x}) = F_{\mathbf{X}}(\mathbf{x}) = P(X_1 \le x_1, \ldots, X_n \le x_n).$$

For $n = 2$, the joint mass function of a discrete random vector (X, Y) is

$$f(x, y) = P(X = x, Y = y).$$

The joint density function $f(x, y)$ of a continuous random vector (X, Y) is defined by

$$F(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(u, v) du dv.$$

The functions $F_X(x)$ and $F_Y(y)$ are called the marginal distribution functions of $F_{X,Y}(x,y)$.

Definition

Random variables X and Y are called independent if

$$F_{X,Y}(x,y) = F_X(x)F_Y(y)$$

for any x and y in $\mathbb{R}$.

Lemma: If (X, Y) has the joint mass (or density) function $f_{X,Y}(x,y)$, then X and Y are independent if and only if $f_{X,Y}(x,y) = f_X(x)f_Y(y)$.

Review of Chapter 3

Definition

If a discrete random variable X has a mass function $f(x)$, then the expectation of X is defined by

$$E[X] = \sum_x x f(x).$$

Lemma: For $g : \mathbb{R} \to \mathbb{R}$, the expectation of $g(X)$ is

$$E[g(X)] = \sum_x g(x) f(x).$$

Definition

The variance of X is defined by $\mathrm{Var}[X] = E[(X - E[X])^2]$.

Lemma: For $h : \mathbb{R}^2 \to \mathbb{R}$, the expectation of $h(X, Y)$ is

$$\mathrm{E}[h(X, Y)] = \sum_x \sum_y h(x, y) f(x, y).$$

Definition

The covariance of random variables X and Y is

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

The correlation (coefficient) of X and Y is

$$\rho[X, Y] = \frac{\mathrm{Cov}[X, Y]}{\sqrt{\mathrm{Var}[X]\mathrm{Var}[Y]}}$$

Lemma: $|\rho[X, Y]| \leq 1$.

Lemma: If X and Y are independent, then X and Y are uncorrelated, meaning $\rho[X, Y] = 0$. But the converse is not true.

Definition

The conditional mass function of Y given $X = x$ is

$$f(y|x) = P(Y = y|X = x).$$

The conditional expectation of Y given $X = x$ is

$$\mathrm{E}[Y|X = x] = \sum_y y f(y|x).$$

Let $\psi(x) = \mathrm{E}[Y|X = x]$. Then $\psi(X)$ is denoted by $\mathrm{E}[Y|X]$. In particular, $\mathrm{E}[Y|X]$ is a random variable.

Lemma (tower property): $\mathrm{E}[\mathrm{E}[Y|X]] = \mathrm{E}[Y]$.