MA933 - Networks and Random Processes

MSc in Mathematics of Systems

Stefan Grosskinsky

Warwick, 2016

These notes and other information about the course are available on
`www2.warwick.ac.uk/fac/sci/mathsys/courses/msc/ma933/`

Contents

References

- G. Grimmett, D. Stirzaker: Probability and Random Processes (3rd edition), Oxford 2001
- C.W. Gardiner: Handbook of Stochastic Methods (3rd edition), Springer 2004
- G. Grimmett: Probability on Graphs, CUP 2010
  `http://www.statslab.cam.ac.uk/~grg/books/pgs.html`
- M.E.J. Newman: Networks: An Introduction, OUP 2010

1. Probability

- **probability space** Ω $\big($ e.g. $\{H, T\}$, $\{$paths of a stoch. process$\}\big)$
- **events** $A \subseteq \Omega$ (measurable) subsets (e.g. odd numbers on a die)
  $\mathcal{F} \subseteq \mathcal{P}(\Omega)$ is the set of all events (subset of the powerset)

Definition 1.1

A **probability distribution** $\mathbb{P}$ on $(\Omega, \mathcal{F})$ is a function $\mathbb{P} : \mathcal{F} \to [0, 1]$ which is

(i) normalized, i.e. $\mathbb{P}[\emptyset] = 0$ and $\mathbb{P}[\Omega] = 1$

(ii) additive, i.e. $\mathbb{P}\big[\cup_i A_i\big] = \sum_i \mathbb{P}[A_i]$,
where $A_1, A_2, \ldots$ is a collection of disjoint events, i.e. $A_i \cap A_j = \emptyset$ for all i, j.

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

- For **discrete** Ω: $\mathcal{F} = \mathcal{P}(\Omega)$ and $\mathbb{P}[A] = \sum_{\omega \in A} \mathbb{P}[\omega]$
  e.g. $\mathbb{P}[\text{even number on a die}] = \mathbb{P}[2] + \mathbb{P}[4] + \mathbb{P}[6] = 1/2$
- For **continuous** Ω (e.g. $[0, 1]$): $\mathcal{F} \subsetneq \mathcal{P}(\Omega)$

1. Independence and conditional probability

- Two events $A, B \subseteq \Omega$ are called **independent** if $\mathbb{P}[A \cap B] = \mathbb{P}[A]\mathbb{P}[B]$.
  Example. rolling a die repeatedly
- If $\mathbb{P}[B] > 0$ then the **conditional probability** of A given B is

$$\mathbb{P}[A|B] := \mathbb{P}[A \cap B]/\mathbb{P}[B] \ .$$

  If A and B are independent, then $\mathbb{P}[A|B] = \mathbb{P}[A]$.

Lemma 1.1 (Law of total probability)

Let $B_1, \ldots, B_n$ be a **partition** of Ω such that $\mathbb{P}[B_i] > 0$ for all i. Then

$$\mathbb{P}[A] = \sum_{i=1}^{n} \mathbb{P}[A \cap B_i] = \sum_{i=1}^{n} \mathbb{P}[A|B_i]\, \mathbb{P}[B_i] \ .$$

Note that also $\mathbb{P}[A|C] = \sum_{i=1}^{n} \mathbb{P}[A|C \cap B_i]\, \mathbb{P}[B_i|C]$.

1. Random variables

Definition 1.2

A **random variable** X is a (measurable) function $X : \Omega \to \mathbb{R}$.
The **distribution function** of the random variable is

$$F(x) = \mathbb{P}[X \leq x] = \mathbb{P}\big[\{\omega : X(\omega) \leq x\}\big] \ .$$

X is called **discrete**, if it only takes values in a countable subset $\{x_1, x_2, \ldots\}$ of $\mathbb{R}$, and its distribution is characterized by the **probability mass function**

$$\pi(x_k) := \mathbb{P}[X = x_k] \ , \quad k = 1, 2, \ldots \ .$$

X is called **continuous**, if its distribution function is

$$F(x) = \int_{-\infty}^{x} f(y)\, dy \quad \text{for all } x \in \mathbb{R} \ ,$$

where $f : \mathbb{R} \to [0, \infty)$ is the **probability density function (PDF)** of X.

1. Random variables

- In general, $f = F'$ is given by the derivative (exists for cont. rv's).
  For discrete rv's, F is a step function with 'PDF'

$$f(x) = F'(x) = \sum_k \pi(x_k)\delta(x - x_k) \ .$$

- The **expected value** of X is given by

$$\mathbb{E}[X] = \begin{cases} \sum_\omega X(\omega)\, \mathbb{P}[\omega] = \sum_k x_k \pi(x_k) \\ \int_\Omega X(\omega) d\mathbb{P}(\omega) = \int_\mathbb{R} x f(x)\, dx \end{cases}$$

- The **variance** is given by $\mathrm{var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$
- Two random variables X, Y are independent if the events $\{X \leq x\}$ and $\{Y \leq y\}$ are independent for all $x, y \in \mathbb{R}$. This implies for **joint distributions**

$$f(x, y) = f_X(x) f_Y(y) \quad \text{or} \quad \pi(x_k, x_l) = \pi^X(x_k)\, \pi^Y(x_l)$$

  where $f_X(x) = \int_\mathbb{R} f(x, y)\, dy$ and $\pi^X(x_k) = \sum_l \pi(x_k, x_l)$ are the **marginals**.

Example. (successive) coin tosses with $\Omega = \{H, T\}$ and $X(H) = -1$, $X(T) = 1$

1. Simple random walk

Definition 1.3

Let $X_1, X_2, \ldots \in \{-1, 1\}$ be a sequence of independent, identically distributed random variables **(iidrv's)** with

$$p = \mathbb{P}[X_i = 1] \quad \text{and} \quad q = \mathbb{P}[X_i = -1] = 1 - p\,.$$

The sequence $Y_0, Y_1, \ldots$ defined as $\quad Y_0 = 0 \quad$ and $\quad Y_n = \sum_{k=1}^n X_k$
is called the **simple random walk (SRW)** on $\mathbb{Z}$.

- for a single **increment** X_k we have

$$\mathbb{E}[X_k] = p - q = 2p - 1\,, \quad \mathrm{var}[X_k] = p + q - (p-q)^2 = 4p(1-p)$$

- $\mathbb{E}[Y_n] = \mathbb{E}\Big[\sum_{k=1}^n X_k\Big] = \sum_{k=1}^n \mathbb{E}[X_k] = n(2p-1)$
  (expectation is a linear operation)
- $\mathrm{var}[Y_n] = \mathrm{var}\Big[\sum_{k=1}^n X_k\Big] = \sum_{k=1}^n \mathrm{var}[X_k] = 4np(1-p)$
  (for a sum of **independent** rv's the variance is additive)

1. LLN and CLT

Theorem 1.2 (Weak law of large numbers (LLN))

Let $X_1, X_2, \ldots \in \mathbb{R}$ be a sequence of iidrv's with $\mu := \mathbb{E}[X_k] < \infty$ and $\mathbb{E}[|X_k|] < \infty$.
Then

$$\frac{1}{n} Y_n = \frac{1}{n} \sum_{k=1}^n X_k \to \mu \quad \text{as } n \to \infty$$

in distribution (i.e. the distr. fct. of Y_n converges to $\mathbb{1}_{[\mu,\infty)}(x)$ for $x \neq \mu$) .

Theorem 1.3 (Central limit theorem (CLT))

Let $X_1, X_2, \ldots \in \mathbb{R}$ be a sequence of iidrv's with $\mu := \mathbb{E}[X_k] < \infty$ and
$\sigma^2 := \mathrm{var}[X_k] < \infty$. Then

$$\frac{Y_n - n\mu}{\sqrt{n}} = \frac{1}{\sqrt{n}} \sum_{k=1}^n (X_k - \mu) \to \xi \quad \text{as } n \to \infty$$

in distr., where $\xi \sim N(0, \sigma^2)$ is a **Gaussian** with PDF $\quad f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-x^2/(2\sigma^2)}$.

Expansion. as $n \to \infty$, $\quad Y_n = n\mu + \sqrt{n}\sigma\xi + o(\sqrt{n})\,, \quad \xi \sim N(0,1)$

1. Discrete-time Markov processes

Definition 1.4

A **discrete-time stochastic process** with **state space** S is a sequence
$Y_0, Y_1, \ldots = (Y_n : n \in \mathbb{N}_0)$ of random variables taking values in S.
The process is called **Markov**, if for all $A \subseteq S$, $n \in \mathbb{N}_0$ and $s_0, \ldots, s_n \in S$

$$\mathbb{P}(Y_{n+1} \in A | Y_n = s_n, \ldots, Y_0 = s_0) = \mathbb{P}(Y_{n+1} \in A | Y_n = s_n)\,.$$

A Markov process (MP) is called **homogeneous** if for all $A \subseteq S$, $n \in \mathbb{N}_0$ and $s \in S$

$$\mathbb{P}(Y_{n+1} \in A | Y_n = s) = \mathbb{P}(Y_1 \in A | Y_0 = s)\,.$$

If S is discrete, the MP is called a **Markov chain (MC)**.

The generic probability space Ω is the **path space**

$$\Omega = D(\mathbb{N}_0, S) := S^{\mathbb{N}_0} = S \times S \times \ldots$$

which is uncountable even when S is finite. For a given $\omega \in \Omega$ the function
$n \mapsto Y_n(\omega)$ is called a **sample path**.
Up to finite time N and with finite S, $\Omega_N = S^N$ is finite.

1. Discrete-time Markov processes

Examples.

- For the simple random walk we have state space $S = \mathbb{Z}$ and $Y_0 = 0$. Up to time N, $\mathbb{P}$ is a distribution on the finite path space Ω_N with

$$\mathbb{P}(\omega) = \prod_{n=1}^{N} \big(p\,\delta_{1,Y_n(\omega)-Y_{n-1}(\omega)} + q\,\delta_{-1,Y_n(\omega)-Y_{n-1}(\omega)}\big)$$
$$= p^{\text{\# of up-steps}}\, q^{\text{\# of down-steps}} \qquad (= (1/2)^N \text{ for } p = q = 1/2)\ .$$

  There are only 2^N paths in Ω_N with non-zero probability.
- For the generalized random walk with $Y_0 = 0$ and increments $Y_{n+1} - Y_n \in \mathbb{R}$, we have $S = \mathbb{R}$ and $\Omega_N = \mathbb{R}^N$ with an uncountable number of possible paths.
- A sequence $Y_0, Y_1, \ldots \in S$ of iidrv's is also a Markov process with state space S.
- Let $S = \{1, \ldots, 52\}$ be a deck of cards, and $Y_1, \ldots, Y_{52}$ be the cards drawn at random without replacement. Is this a Markov process?

1. Discrete-time Markov chains

Proposition 1.4

Let $(X_n : n \in \mathbb{N}_0)$ by a homogeneous DTMC with state space S. Then the **transition function**

$$p_n(x,y) := \mathbb{P}[X_n = y | X_0 = x] = \mathbb{P}[X_{k+n} = y | X_k = x] \quad \text{for all } k \geq 0$$

is well defined and fulfills the **Chapman Kolmogorov equations**

$$p_{n+k}(x,y) = \sum_{z\in S} p_n(x,z)\, p_k(z,y) \quad \text{for all } k,n \geq 0,\ x,y \in S\ .$$

Proof. We use the law of total probability, the Markov property and homogeneity

$$\mathbb{P}[X_{n+k} = y|X_0 = x] = \sum_{z\in S} \mathbb{P}[X_{n+k} = y|X_k = z,\ X_0 = x]\,\mathbb{P}[X_k = z|X_0 = x]$$
$$= \sum_{z\in S} \mathbb{P}[X_{n+k} = y|X_k = z]\,\mathbb{P}[X_k = z|X_0 = x]$$
$$= \sum_{z\in S} \mathbb{P}[X_n = y|X_0 = z]\,\mathbb{P}[X_k = z|X_0 = x]$$

1. Markov chains

- In matrix form with $P_n = \big(p_n(x,y) : x,y \in S\big)$ the Chapman Kolmogorov equations read

$$P_{n+k} = P_n\, P_k \quad \text{and in particular} \quad P_{n+1} = P_n\, P_1\ .$$

  With $P_0 = \mathbb{I}$, the obvious solution to this recursion is

$$P_n = P^n \quad \text{where we write} \quad P_1 = P = \big(p(x,y) : x,y \in S\big)\ .$$

- The **transition matrix** P and the initial condition $X_0 \in S$ completely determine a homogeneous DTMC, since for all $k \geq 1$ and all events $A_1, \ldots, A_k \subseteq S$

$$\mathbb{P}[X_1 \in A_1, \ldots, X_k \in A_k] = \sum_{s_1\in A_1} \cdots \sum_{s_k \in A_k} p(X_0, s_1)p(s_1,s_2)\cdots p(s_{k-1},s_k)\ .$$

- Fixed X_0 can be replaced by an **initial distribution** $\quad \pi_0(x) := \mathbb{P}[X_0 = x]$.
  The distribution at time n is then

$$\pi_n(x) = \sum_{y\in S}\sum_{s_1\in S}\cdots\sum_{s_{n-1}\in S} \pi_0(y)p(y,s_1)\cdots p(s_{n-1},x) \quad \text{or} \quad \pi_n = \pi_0 P^n\ .$$

1. Transition matrices

The transition matrix P is **stochastic**, i.e.

$$p(x,y) \in [0,1] \quad \text{and} \quad \sum_y p(x,y) = 1 \ .$$

Example 1 (Random walk with boundaries)

Let $(X_n : n \in \mathbb{N}_0)$ be a SRW on $S = \{1, \ldots, L\}$ with $p(x,y) = p\delta_{y,x+1} + q\delta_{y,x-1}$.
The boundary conditions are

- **periodic** if $\quad p(L,1) = p \ , \quad p(1,L) = q \ ,$
- **absorbing** if $\quad p(L,L) = 1 \ , \quad p(1,1) = 1 \ ,$
- **closed** if $\quad p(1,1) = q \ , \quad p(L,L) = p \ ,$
- **reflecting** if $\quad p(1,2) = 1 \ , \quad p(L, L-1) = 1 \ .$

1. Stationary distributions

Definition 1.5

Let $(X_n : n \in \mathbb{N}_0)$ be a homogeneous DTMC with state space S. The distribution $\pi(x), \ x \in S$ is called **stationary** if for all $y \in S$

$$\sum_{x \in S} \pi(x) p(x,y) = \pi(y) \quad \text{or} \quad \boldsymbol{\pi} P = \boldsymbol{\pi} \ .$$

π is called **reversible** if it fulfills the **detailed balance** conditions

$$\pi(x) p(x,y) = \pi(y) p(y,x) \quad \text{for all } x, y \in S \ .$$

- reversibility implies stationarity, since

$$\sum_{x \in S} \pi(x) p(x,y) = \sum_{x \in S} \pi(y) p(y,x) = \pi(y) \ .$$

- Stationary distributions are left **eigenvectors** with **eigenvalue** 1 .

1. Distribution at time n

Consider a DTMC on a finite state space with $|S| = L$, and let $\lambda_1, \ldots, \lambda_L \in \mathbb{C}$ be the **eigenvalues** of the transition matrix P with corresponding

left (row) eigenvectors $\langle u_i|$ and **right (column) eigenvectors** $|v_i\rangle$

in bra-ket notation. Assuming that **all eigenvalues are distinct** we have

$$A = \sum_{i=1}^{L} \lambda_i |v_i\rangle\langle u_i| \quad \text{and} \quad A^n = \sum_{i=1}^{L} \lambda_i{}^n |v_i\rangle\langle u_i|$$

since eigenvectors can be chosen **orthonormal** $\quad \langle u_i|v_j\rangle = \delta_{i,j}$.
Since $\boldsymbol{\pi}_n = \boldsymbol{\pi}_0 P^n$ we get

$$\langle \boldsymbol{\pi}_n| = \langle \boldsymbol{\pi}_0|v_1\rangle \lambda_1{}^n \langle u_1| + \ldots + \langle \boldsymbol{\pi}_0|v_L\rangle \lambda_L{}^n \langle u_L| \ .$$

- The **Gershgorin theorem** implies that $|\lambda_i| \leq 1$ and the dependence on the initial condition $\boldsymbol{\pi}_0$ decays exponentially in directions where $|\lambda_i| < 1$.
- $\lambda_1 = 1$ corresponds to the stationary distribution and $|v_1\rangle = (1, \ldots, 1)^T$.
- Other eigenvalues with $|\lambda_i| = 1$ and $\lambda_i \neq 1$ correspond to persistent oscillations.

1. Lazy Markov chains

Definition 1.6

Let $(X_n : n \in \mathbb{N}_0)$ be a DTMC with transition matrix $p(x,y)$. The DTMC with transition matrix

$$p^\epsilon(x,y) = \epsilon\delta_{x,y} + (1-\epsilon)\,p(x,y)\ , \quad \epsilon \in (0,1)$$

is called a **lazy version** of the original chain.

- Since all diagonal elements are bounded below by $\epsilon > 0$, the Gershgorin theorem now implies for the eigenvalues of P^ϵ

  $$|\lambda_i| = 1 \quad \Rightarrow \quad \lambda_i = 1\ .$$

  Such a matrix P^ϵ is called **aperiodic**, and there are no persistent oscillations.
- The stationary distribution is unique if and only if the eigenvalue $\lambda = 1$ has multiplicity 1, which is independent of lazyness and is discussed later.

1. Absorbing states

Definition 1.7

A state $s \in S$ is called **absorbing** for a DTMC with transition matrix $p(x,y)$, if

$$p(s,y) = \delta_{s,y} \quad \text{for all } y \in S\ .$$

RW with absorbing BC.
Let h_k be the **absorption probability** for $X_0 = k \in S = \{1,\ldots,L\}$,

$$h_k = \mathbb{P}[\text{absorption}|X_0 = k] = \mathbb{P}[X_n \in \{1,L\} \text{ for some } n \geq 0|X_0 = k]\ .$$

Conditioning on the first jump and using Markov, we have the recursion

$$h_k = p h_{k+1} + q h_{k-1} \quad \text{for } k = 2,\ldots,L-1\ ; \quad h_1 = h_L = 1\ .$$

Ansatz for solution $\quad h_k = \lambda^k\ , \quad \lambda \in \mathbb{C}$:

$$\lambda = p\lambda^2 + q \quad \Rightarrow \quad \lambda_1 = 1\ , \quad \lambda_2 = q/p$$

General solution of 2nd order linear recursion

$$h_k = a\lambda_1^k + b\lambda_2^k = a + b(q/p)^k\ , \quad a,b \in \mathbb{R}\ .$$

Determine coefficients from boundary condition $\quad \Rightarrow \quad h_k \equiv 1$