PROBABILITY 2018
HANDOUT 5: MARKOV PROCESSES AND CHAINS

GIOVANNI PISTONE

Contents

1. Markov process

1. A stochastic process $Y_0, Y_1, \ldots$ is a *Markov Process* if past and future are conditionally independent given the present,

$$(Y_0, \ldots, Y_k) \perp\!\!\!\perp (Y_k, \ldots, Y_N) |\, Y_k \ , \quad k < N \ .$$

An equivalent condition is the sufficiency of the present in computing the distribution of the future given the past,

$$\mathrm{E}\left(\phi_k(Y_k) \cdots \phi_N(Y_N) | Y_0, \ldots, \phi_N(Y_k)\right) = \mathrm{E}\left(\phi_k(Y_k) \cdots \phi_N(Y_N) | Y_k\right) \ , \qquad k < N \ ,$$

for all bounded $\phi_k, \ldots, \phi_N$.

Exercise 1. The special case

$$\mathrm{E}\left(\phi(Y_{k+1}) | Y_0, \ldots, Y_k\right) = \mathrm{E}\left(\phi_k(Y_{k+1}) | Y_k\right) \ , \qquad k < N \ ,$$

for all bounded ϕ implies the Markov property. [Hint: Prove the sufficiency condition by induction conditioning first on $Y_0, \ldots, Y_{N-1}$.]

Exercise 2. Let be given a Gaussian white noise $Z_1, \ldots, Z_n$ and a further independent Gaussian random variable X_0. For a given real α define $X_k = \alpha X_{k-1} + Z_k$, $k \geqslant 1$. Show that it is a Markov process and compute the transitions. [Hint: write $\phi(X_{k+1}) = \phi(\alpha X_k + Z_{k+1}$ and use the independence.]

2 (Martingale problem)**.** Let $Y_0, Y_1, \ldots, Y_N$ be a Markov process, each random variable having values in the measurable space $(S, \mathcal{S})$. Given any bounded measurable $\phi \colon S \to \mathbb{R}$ define the new process X^ϕ by $X^\phi_0 = \phi(Y_0)$,

$$X^\phi_t = X^\phi_{t-1} + \phi(Y_t) - \mathrm{E}\left(\phi(Y_t) | Y_0, \ldots, Y_{t-1}\right) \ , \quad t \geqslant 1 \ .$$

Then, X^ϕ_t is $(Y_0, \ldots, Y_t)$-measurable and has the martingale property

$$\mathrm{E}\left(X^\phi_t \middle| Y_0, \ldots, Y_{t-1}\right) = X^\phi_{t-1} \ .$$

Because the Markov property is a property of conditional independence, we know that

$$\mathrm{E}\left(\phi(Y_t) | Y_0, \ldots, Y_{t-1}\right) = \mathrm{E}\left(\phi(Y_t) | Y_{t-1}\right) = \int \phi(y) \ \mu_{Y_t | Y_{t-1}}(dy | Y_{t-1}) \ .$$

Date: May 3, 2018.

The family of operators $\phi \mapsto A_t\phi$, $t = 1, \dots, N$, defined by $A_t\phi(x) = \phi(x) - \int \phi(y)\, \mu_{Y_t|Y_{t-1}}(dy|x)$ is called the generator of the Markov process. We have

$$X_t^\phi - X_{t-1}^\phi = \phi(Y_t) - \phi(Y_{t-1}) + A_t\phi(Y_{t-1})$$

hence,

$$X_t^\phi = \phi(Y_t) + \sum_{s=1}^{t} A_s\phi(Y_{s-1}) \ .$$

The process $\left(\sum_{s=1}^t A_s\phi(Y_{s-1})\right)_{t\geqslant 1}$ is predictable and it is called the *compensator* of $(\phi(Y_t))_{t\geqslant 1}$ because process minus compensator equals martingale. (See another example below.)

2. Markov chain with stationary transition probability

3. Let S be a finite set with $\#S = N$. A *Markov chain with stationary transition probability* (MC) is an S-valued Markov process $X_0, X_1, \dots$ such that for all couples of consecutive times $t, t+1$ the conditional distribution of X_{t+1} given X_t does not depend on t. The common transition is called the stationary transition of the MC. As the space S is finite, the transition is characterized by the numbers $[p(y|x) \colon x, y \in S]$, so that

$$\begin{aligned} \mathrm{E}\left(\phi(X_{t+1})|X_0, \dots, X_t\right) &= \mathrm{E}\left(\phi(X_{t+1})|X_t\right) \quad \text{Markov property} \\ &= \sum_{y\in S} \phi(y)p(y|X_t) \quad \text{stationary transitions .} \end{aligned}$$

Given an order or numbering on S, the transition can be given in form of a matrix $P = [P_{x,y}]_{x,y\in S}$ with $P_{x,y} = p(y|x)$. P is called *transition matrix.* If the real functions on S are identified with a column vector e.g., $[\phi(y) \colon y \in S]^*$, then $\mathrm{E}\left(\phi(X_{t+1})|X_t\right) = \hat{\phi}(X_t)$ with $\hat{\phi} = P\phi$.

Let us represent the probability functions on S as row vectors. If $\pi_t(y) = [\mathrm{P}\left(X_t = y\right) : x \in S]$, then the joint distribution of (X_t, X_{t+1}) is given by the probability function

$$(x, y) \mapsto \mathrm{P}\left(X_t = x, X_{t+1} = y\right) = \mathrm{P}\left(X_{t+1} = y|X_t = x\right) \mathrm{P}\left(X_t = x\right) = \pi_t(x)P_{x,y}$$

and the probability function of X_{t+1} is $y \mapsto \pi_{t+1}(y) = \sum_x \pi_t(x)P_{x,y}$, that is $\pi_{t+1} = \pi_t P$.

The probability function π is *invariant* if $\pi P = \pi$ that is if π is a left eigenvector with eigenvaue 1.

Given a sequence of times $s, s+1, \dots, s+k$ then the joint probability function of $X_s, X_{s+1}, \dots, X_{s+k}$ is given by

$$\mathrm{P}\left(X_s = x_s, X_{s+1} = x_{s=1}, \dots, X_{s+k} = x_{s+k}\right) = \pi_s(x_s)P_{x_s,x_{s+1}} \cdots P_{x_{s+k-1},x_{s+k}} \ .$$

The proof is by induction. If $k = 1$, then $\mathrm{P}\left(X_s = x_s, X_{s+1} = x_{s+1}\right) = \pi_s(x_s)P_{x_s,x_{s+1}}$. If it is true up to $k-1$, then the MC property implies

$$\begin{aligned} &\mathrm{P}\left(X_s = x_s, X_{s+1} = x_{s+1}, \dots, X_{s+k} = x_{s+k}\right) = \\ &\qquad \mathrm{P}\left(X_{s+k} = x_{s+k}|X_s = x_s, X_{s+1} = x_{s+1}, \dots, X_{s+k-1} = x_{s+k-1}\right) \times \\ &\qquad\qquad \mathrm{P}\left(X_s = x_s, X_{s+1} = x_{s+1}, \dots, X_{s+k-1} = x_{s+k-1}\right) = \\ &\mathrm{P}\left(X_{s+k} = x_{s+k}|X_{s+k-1} = x_{s+k-1}\right) \mathrm{P}\left(X_s = x_s, X_{s+1} = x_{s=1}, \dots, X_{s+k-1} = x_{s+k-1}\right) = \\ &\qquad\qquad \mathrm{P}\left(X_s = x_s, X_{s+1} = x_{s=1}, \dots, X_{s+k-1} = x_{s+k-1}\right) P_{x_{s+k-1},x_{s+k}} \end{aligned}$$

Given the initial distribution with probability function π_0 and the transition matrix P, then the distribution at time t is given by the probability function

$$\pi_t(y) = \sum_{x,\dots,x_{t-1}\in S} \pi_0(x_0)P_{x_0,x_1} \cdots P_{x_{t-1},y} = \pi_0(x)P^n_{x,y}$$

that is, $\pi_t = \pi_0 P^n$. If $\lim_{n\to\infty} \pi_0 P^n = \pi$ exists and is a probability function, then π is an invariant probability. In fact,

$$\pi P = \left(\lim_{n\to\infty} \pi_0 P^n\right) P = \lim_{n\to\infty} \pi_0 P^{n+1} = \pi \ .$$

We do not discuss here the existence and uniqueness of such a limit, see the Wikipedia article *Markov chain* and the references therein, but see the exercise below. *Markov Chain Monte Carlo* (MCMC) is a popular simulation method that uses $\lim_{n\to\infty} \pi_0 P^n$ to simulate π.

Exercise 3. Let be given a probability function π_0 on S and a $N \times N$ matrix P with positive elements such that $P\mathbf{1} = \mathbf{1}$. Such a matrix is called a *Markov matrix*. Take the sample space $\Omega = S^{n+1}$ and define the probability function

$$(x_0, x_1, \ldots, x_n) \mapsto \pi_0(x_0) P_{x_0,x_1} \cdots P_{x_{n-1},x_n} \ .$$

(Show by induction that it is indeed a probability function.) Let P be the probability with the given probability function and let $X_0, X_1, \ldots, X_n$ be the canonical process i.e., $X_t(x_0, \ldots, x_n) = x_t$. Check that the process is a MC with transition matrix P. In other word, the distribution of a MC is characterized by π_0 and the transition matrix. Conversely, given any probability function π and any Markov matrix P, there exist a MC with the given initial probability and the given transition.

Exercise 4 (2-state MC). Let

$$P = \begin{bmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{bmatrix} \quad \alpha, \beta \in [0,1]$$

be the generic Markov matrix on the two elements set $S = \{1, 2\}$. An invariant probability function is $\pi = \begin{bmatrix} p & 1-p \end{bmatrix}$ such that

$$\begin{cases} p = p(1-\alpha) + (1-p)\beta \\ 1-p = p\alpha + (1-p)(1-\beta) \end{cases} .$$

The two equations are dependent because the rank of $P - I$ is 1. It follows

$$p(\alpha + \beta) = \beta \quad \text{and} \quad (1-p)(\alpha+\beta) = \alpha \ .$$

If $\alpha + \beta = 0$ i.e., $P = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ then all probability functions are invariant. In the following, we assume $\alpha + \beta > 0$. In such a case, the invariant probability function is

$$\pi = \begin{bmatrix} \frac{\beta}{\alpha+\beta} & \frac{\alpha}{\alpha+\beta} \end{bmatrix} \ .$$

For example, the invariant probability of both $P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ and $P = \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{bmatrix}$ is $\pi = \begin{bmatrix} \frac12 & \frac12 \end{bmatrix}$. (Note that the first example is produces a "deterministic" process while the second produces and "independent" process.) The characteristic equation of P is

$$\det(P - \lambda I) = \det \begin{bmatrix} (1-\alpha) - \lambda & \alpha \\ \beta & (1-\beta) - \lambda \end{bmatrix} = \lambda^2 - (2 - \alpha - \beta)\lambda + (1 - \alpha - \beta) = 0 \ .$$

One solution is $\lambda_1 = 1$ (why?), while the other is $\lambda_2 = 1 - \alpha - \beta$. Let us assume $\alpha, \beta > 0$. The first eigen-vector is a vector

$$\boldsymbol{u}_1 = \begin{bmatrix} u_{12} \\ u_{22} \end{bmatrix} \quad \text{such that} \quad \begin{bmatrix} -\alpha & \alpha \\ -\beta & \beta \end{bmatrix} \begin{bmatrix} u_{11} \\ u_{21} \end{bmatrix} = 0$$

e.g., $\boldsymbol{u}_1 = \begin{bmatrix} 1 & 1 \end{bmatrix}^*$. The second eigen-vector is a vector

$$\boldsymbol{u}_2 = \begin{bmatrix} u_{12} \\ u_{22} \end{bmatrix} \quad \text{such that} \quad \begin{bmatrix} \beta & \alpha \\ \beta & \alpha \end{bmatrix} \begin{bmatrix} u_{12} \\ u_{22} \end{bmatrix} = 0$$

e.g., $\boldsymbol{u}_2 = \begin{bmatrix} -\alpha & \beta \end{bmatrix}^*$. It follows that

$$P = U \begin{bmatrix} 1 & 0 \\ 0 & 1-\alpha-\beta \end{bmatrix} U^{-1} \quad \text{with} \quad U = \begin{bmatrix} \boldsymbol{u}_1 & \boldsymbol{u}_2 \end{bmatrix} = \begin{bmatrix} 1 & -\alpha \\ 1 & \beta \end{bmatrix}$$

because $\det U = \alpha + \beta > 0$. It follows that

$$P^n = U \begin{bmatrix} 1 & 0 \\ 0 & (1-\alpha-\beta)^n \end{bmatrix} U^{-1} \ .$$

(Compute U^{-1} and P^n. Check the equation for $n = 2$.) As $-1 < 1-\alpha-\beta < 1$, we have

$$\lim_{n\to\infty} P^n =$$

$$U \lim_{n\to\infty} \begin{bmatrix} 1 & 0 \\ 0 & (1-\alpha-\beta)^n \end{bmatrix} U^{-1} = \begin{bmatrix} 1 & -\alpha \\ 1 & \beta \end{bmatrix} \begin{bmatrix} \frac{1}{\alpha+\beta} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \beta & \alpha \\ -1 & 1 \end{bmatrix} =$$

$$\begin{bmatrix} \frac{\beta}{\alpha+\beta} & \frac{\alpha}{\alpha+\beta} \\ \frac{\beta}{\alpha+\beta} & \frac{\alpha}{\alpha+\beta} \end{bmatrix} = \begin{bmatrix} \pi \\ \pi \end{bmatrix} \ .$$

In conclusion: if $\alpha + \beta = 0$ all probability functions are invariant and there is no convergence to the invariant probability; If $\alpha + \beta > 0$ there is a unique probability and there is convergence. What happens if $\alpha = 0$ while $\beta > 0$?

Exercise 5. In the previous example, consider the case $\alpha, \beta > 0$. Define the time of first visit to 1 as

$$T = \inf \{t \geqslant 0 | X_t = 1\} \ .$$

Show that T is a stopping time with probability function

$$\mathrm{P}(T = t) = \alpha(1-\beta)^n \ .$$

Explain what happens if $\beta = 1$. Otherwise, T is almost surely finite and $\mathbb{E}(T) = \alpha/\beta$.

Exercise 6. Let X_n, $n \geqslant 0$, be a Markov process with finite state space S and assume stationary transition probabilities given by the Markov matrix P. For each real function $\phi \colon S \to \mathbb{R}$, we have

$$\mathrm{E}(\phi(X_t)|X_0, \dots, X_{t-1}) - \phi(X_{t-1}) = A\phi(X_{t-1}) \quad \text{where} \quad A = P - I \ .$$

It follows that $M_t^\phi = \phi(X_t) - \sum_{s=1}^t A\phi(X_{s-1})$ is a martingale.

If T is a stopping time, then

$$M_{t\wedge T}^\phi = \phi(X_{t\wedge T}) - \sum_{1 \leqslant s \leqslant t\wedge T} A\phi(X_{s-1})$$

is martingale.

Conversely, if P is a Markov matrix, $A = P - I$, and for each ϕ the process M^ϕ is a martingale, then

$$\begin{aligned}
\mathrm{E}\left(\phi(X_t)|\mathcal{F}_{t-1}\right) &= \mathrm{E}\left(M_t^\phi + \sum_{s=1}^t A\phi(X_{s-1})\middle|\mathcal{F}_{t-1}\right) \\
&= M_{t-1}^\phi + \sum_{s=1}^t A\phi(X_{s-1}) \\
&= \phi(X_{t-1}) + A\phi(X_{s-1}) = P\phi(X_{t-1})
\end{aligned}$$

so that $(X_t)_{t\geqslant 0}$ is a MC with transition matrix P.

3. Reversible Markov chains

4. The Markov property is symmetric in the direction of time. If $X_0, \dots, X_n$ is a MC, then the time-reversed process $Y_h = X_{n-h}$ is a Markov process with transitions

$$\begin{aligned}
&\mathrm{P}\left(Y_{h+1} = x|Y_h = y\right) = \mathrm{P}\left(X_{n-h-1} = x|X_{n-h} = y\right) = \\
&\quad \frac{\mathrm{P}\left(X_{n-h-1} = x, X_{n-h} = y\right)}{\mathrm{P}\left(X_{n-h} = y\right)} = \frac{\mathrm{P}\left(X_{n-h} = y|X_{n-h-1} = x\right)\mathrm{P}\left(X_{n-h-1} = x\right)}{\mathrm{P}\left(X_{n-h} = y\right)} = \\
&\qquad \frac{P_{x,y}\pi_{n-h-i}(x)}{\pi_{n-h}(y)} .
\end{aligned}$$

If moreover the MC is stationary that is $\pi_t = \pi$, then the time-reversed process is a MC with the same invariant distribution and transitions

$$Q_{y,x} = \frac{\pi(x)P_{x,y}}{\pi(y)} .$$

Equivalently, we can say that the 2-dimensional distribution are given by

$$\mathrm{P}\left(X_s = x, X_{s+1} = y\right) = \pi(x)P_{x,y} = \pi(y)Q_{y,x} .$$

A stationary Markov chain is reversible if $Q_{j.i} = P_{j,i}$. Equivalently, if π is a probability function such that

$$\pi(x)P_{x,y} = \pi(y)P_{y,x} ,$$

we sum the previous relation over x to get

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

so that π is indeed an invariant probability and the MC constructed from π and P is reversible.

Given a Markov matrix P, if there exists a positive function $\kappa\colon S$ such that $\kappa(x)P_{x,y} = \kappa(y)P_{y,x}$ then we can normalize κ. In such a case we have an immediate way to compute the invariant probability.

Exercise 7. Let $G = (S, \mathcal{E})$ be a graph. For each vertex $x \in S$ the *degree* of x, $\deg x$, is the number of edges from x. Let E be the *adjacency matrix* of G. The degree as a row vector is $E\mathbf{1}$. Define the Markov matrix

$$P = \operatorname{diag}\left(E\mathbf{1}\right)^{-1} E .$$

i.e., the transitions

$$P_{x,y} = \begin{cases} \frac{1}{\deg x} & \text{if } y \text{ is connected with } x, \\ 0 & \text{if } y \text{ is not connected with } x. \end{cases}$$

Observe that x is connected to y if, and only if, y is connected to x, hence

$$(\deg x)P_{x,y} = (x \to y) = (y \to x) = (\deg y)P_{y,x} \ .$$

It follows that the invariant probability is

$$\pi(x) = \frac{\deg x}{\sum_{y \in S} \deg y} \ .$$

and the MC is reversible.

Exercise 8 (Hastings-Metropolis). Consider the following problem: Given a Markov matrix Q on a finite S and a probability function π on S, define the matrix

$$P_{x,y} = \begin{cases} Q_{x,y}\alpha(x,y) & \text{if } x \neq y \\ Q_{x,x} + \sum_{z \neq x} Q_{x,z}(1 - \alpha(x,z)) & \text{if } x = y \end{cases} ,$$

where $0 \leqslant \alpha(x,y) \leqslant 1$ Notice that $P_{x,y} \geqslant 0$ amd

$$\sum_{y \in S} P_{x,y} = \sum_{y \neq x} Q_{x,y}\alpha(x,y) + Q_{x,x} + \sum_{z \neq x} Q_{x,z}(1 - \alpha(x,z)) = Q_{x,x} + \sum_{z \neq x} Q_{x,z} = 1 \ .$$

The Markov matrix P is reversible with invariant probability π if

$$\pi(x)Q_{x,y}\alpha(x,y) = \pi(y)Q_{y,x}\alpha(y,x) \ , \quad x \neq y \ .$$

One possible choice is

$$\alpha(x,y) = 1 \wedge \frac{\pi(y)Q_{y,x}}{\pi(x)Q_{x,y}} \ .$$

Collegio Carlo Alberto
E-mail address: giovanni.pistone@carloalberto.org