On Markov processes

Yukio NAGAHATA
Niigata univ.
nagahata@eng.niigata-u.ac.jp

2016 January 4-8

stationary measure, stationary distribution

Definition (stationary measure, stationary distribution)

Suppose that we set $\mu = (\mu_{x_1}, \mu_{x_2}, \dots, \mu_{x_n})$ if the state space is a finite set and $\#S = n$, and $\mu = (\mu_{x_1}, \mu_{x_2}, \dots,)$ if the state space S is a countable set. We also suppose that $\forall x$, $\mu_x \geq 0$ and $\mu \neq 0$. If it satisfies $\mu P = \mu$, then we call μ a stationary measure or invariant measure, furthermore if it also satisfies $\sum_x \mu_x = 1$, then we call μ a stationary distribution or invariant distribution.

Note that if μ is a stationary measure, then $C\mu$ for some positive constant C is also a stationary measure. If S is a finite set, then we set $1/C = \sum_x \mu_x$, then $C\mu$ becomes a stationary distribution.

Markov processes

If the size of the transition matrix P is large, it is always hard to find a stationary measure μ. But following results is applicable to find a stationary measure.

Problem

Show the following;
Suppose that $\pi > 0$ and P is symmetric with respect to P, then π is a stationary measure for P.

We only say that for any $\pi > 0$

$$Q = \frac{1}{12}\left(\begin{array}{ccc} 2 & 8 & 2 \\ 5 & 2 & 5 \\ 5 & 2 & 5 \end{array}\right)$$

is not symmetric with respect to π.
If we prove this by means of matrix theory, then we consider the system of equations

$$\pi_x Q_{x,y} = \pi_y Q_{y,x}, \quad x, y = a, b, c,$$

and find some nontrivial solution. Canonical solution is compute the rank of corresponding matrix.

Markov processes

But if we prove this by means of Markov process theory, then we use the last Problem.
If there is π such that Q is symmetric with respect to π, then π is a stationary measure for Q. By the definition of the stationary measure, π is the left eigenvector of Q whose eigenvector is 1. In this example, we have already computed the eigenvalues $(1, 0, -1/4)$. It is easy to compute that the corresponding eigenvector is $\pi = (1, 1, 1)$. Finally we only check that this Q is symmetric with respect to this π or not.

Markov processes

For a given measure $\mu > 0$, we define a (weighted) inner product $\langle \cdot, \cdot \rangle_\mu$ by

$$\langle x, y \rangle_\mu = \sum_i x_i y_i \mu_i.$$

Usual inner product is given by $\langle x, y \rangle_1$ for $1 = (1, 1, \dots, 1)$.
For given a matrix A and a measure μ. If it satisfies $\langle x, Ay \rangle_\mu = \langle y, A^* x \rangle_\mu$ for all x, y then we call A^* an adjoint matrix with respect to μ.

Problem

Show that $(A^*_\mu)_{i,j} = \mu_i^{-1} A_{j,i} \mu_j$.
Show also that if A is symmetric with respect to μ then $(A^*_\mu) = A$.

First part of this Problem implies following Proposition;

Proposition (relation between P^* and stationary measure)

Let P be a transition matrix. The adjoint matrix P^*_μ is also a transition matrix, iff μ is a stationary distribution of P.

Proof

By the first part of the last Problem, we have $(P^*_\mu)_{i,j} = \mu_i^{-1} P_{j,i} \mu_j$.
Hence we have $(P^*_\mu)_{i,j} \geq 0$.
We have only to show that $\forall i$, $\sum_j (P^*_\mu)_{i,j} = 1$ iff μ is a stationary distribution. Note that

$$\sum_j (P^*_\mu)_{i,j} = \sum_j \mu_i^{-1} P_{j,i} \mu_j = \mu_i^{-1} \sum_j \mu_j P_{j,i}.$$

If μ is a stationary distribution, then the right hand side of this expression is 1 and if not, then there is at least one i such that $\sum_j \mu_j P_{j,i} \neq \mu_i$. $\square$

Theorem (Perron-Frobenius)

Let A be a nonnegative irreducible square matrix. Then we have following results;
(1) Let $\rho(A)$ be a spectral radius of A. Namely, $\rho(A) = \max\{|\lambda_i|\}$ where λ_i are eigenvalues of A. Then A has an eigenvalue $\rho(A)$.
(2) The eigenvalue related to $\rho(A)$ is positive.
(3) $\rho(A)$ is increasing function of each elements of A.
(4) $\rho(A)$ is simple.

Note that $\rho(A)$ is called the Perron-Frobenius eigenvalue, etc.
Proof
In order to prove this theorem, we prepare some notations and lemmas.

Markov processes

For nonnegative vector x, we define

$$r_x := \min\{\frac{(Ax)_i}{x_i}; x_i > 0\}.$$

Furthermore, we define

$$r := \sup\{r_x; x \geq 0, x \neq 0\}.$$

Note that it is easy to see that

$$r := \sup\{r_x; x \geq 0, \|x\| = 1\}.$$

A nonnegative vector z is called extremal or an extremal vector iff

$$Az \geq rz$$

Markov processes

Since A is irreducible, $\exists n$ s.t. $(I + A)^n > 0$. For nonnegative vector y with $\|y\| = 1$, we define

$$T(y) := y' = \tilde{y}'/\|\tilde{y}'\|, \quad \tilde{y}' = (I + A)^n y$$

Then we have $y' > 0$ and $\|y'\| = 1$. Since $Ay \geq r_y y$, we have

$$\frac{(Ay')_i}{y'_i} = \frac{((I + A)^n Ay)_i}{(I + A)^n y)_i} \geq \frac{(r_y (I + A)^n y)_i}{(I + A)^n y)_i} \geq r_y$$

Hence if we set $B := \{y;$ nonnegative vectors with $\|y\| = 1\}$, then we have $TB \subset \{y;$ positive vectors with $\|y\| = 1\}(= B') \subset B$ and

$$r := \sup\{r_x ; xinTB\}.$$

The definition of r_x is continuous in B' and there is a closed set B'' such that $TB \subset B'' \subset B'$. Hence a set of extremal vectors is not empty.

Lemma (extremal vector)

Let A be a nonnegative irreducible square matrix. Then we have
(1) $r > 0$
(2)The extremal vector z is positive and an eigenvector of A corresponding to r. Namely, $Az = rz$

Proof

(1) For any positive vector x, Ax is also positive vector. By the definition of r, we have $r > 0$.
(2) Since A is irreducible, $(I + A)^n$ is positive matrix for large enough n. By the definition of the extremal vector, we have $y = Az - rz \geq 0$. Suppose that $y \neq 0$ and we set $w = (I + A)^n z > 0$. Then we have

$$Aw - rw = A(I + A)^n z - r(I + A)^n z = (I + A)^n y > 0.$$

Then we have

$$r_w = \min_i \{\frac{Aw_i}{w_i}\} = \min_i \{\frac{rw_i + ((I + A)^n y)_i}{w_i}\} > r$$

This inequality contradict to the definition of r. Namely $y = 0$ and $Az = rz$.
Under this condition we have $0 < w = (I + A)^n z = (1 + r)^n z$, namely z is a positive vector. $\square$

Lemma (comparison of matrices)

Let A be a nonnegative irreducible square matrix and B be a complex matrix such that $|B| \leq A$.
If β is an eigenvalue of B, then

$$|\beta| \leq r.$$

Furthermore $|\beta| = r$ iff $|B| = A$ and if $\beta = re^{i\phi}$ then there is a diagonal matrix D whose diagonal elements d_i satisfy $|d_i| = 1$ such that

$$B = e^{i\phi} DAD^{-1}$$

Proof

Suppose that $y \neq 0$ and $By = \beta y$. By comparing each elements,we have

$$|\beta||y| \leq |B||y| \leq A|y|$$

namely $|\beta| \leq r_{|y|} \leq r$.

Suppose that $|\beta| = r$. Then we have $|\beta||y| \leq A|y|$. Hence $|y|$ is an extremal vector. By using last Lemma, $|y|$ is an eigenvector of A corresponding to r, i.e., $r|y| = A|y|$. This identity and above inequality imply

$$r|y| = |B||y| = A|y|$$

Since $|y|$ is positive vector and $|B| \leq A$ we conclude that $|B| = A$.

Markov processes

For $|y| > 0$, we set

$$D = \begin{pmatrix} y_1/|y_1| & & \\ & \ddots & \\ & & y_n/|y_n| \end{pmatrix}$$

Then this matrix satisfies the condition of D in this Lemma and we can rewrite $y = D|y|$. Suppose that $\beta = re^{i\phi}$. Then we can rewrite $By = \beta y$ as

$$BD|y| = By = \beta y = e^{i\phi} rD|y|, \text{ namely } e^{-i\phi} D^{-1} BD|y| = r|y|$$

Hence we conclude that

$$|e^{-i\phi} D^{-1} BD| = |B| = A, \quad e^{-i\phi} D^{-1} BD|y| = r|y| = A|y|$$

Finally, following Problem implies $B = e^{i\phi} DAD^{-1}$. □

Problem

Prove following;
Suppose that C be a complex matrix. If $\exists y > 0$ s.t.

$$|C|y = Cy$$

then $|C| = C$.

Markov processes

Taking $B = A$ in this lemma, we have

Corollary (comparison of matrices)

Suppose that A is a nonnegative and irreducible square matrix. Then $r = \rho(A)$.

A matrix B obtained from a matrix A by removing j-th row and j-th column (We allow removing more than one row and column) is called a principal submatrix.

Lemma (comparison of spectral radii)

Suppose that A is a nonnegative and irreducible square matrix. Then for any principal submatrix B, we have $\rho(A) > \rho(B)$.

Proof

Let P be a matrix such that it only changes the order of columns such that

$$A' = PAP^{-1} = \begin{pmatrix} B & A_{1,2} \\ A_{2,1} & A_{2,2} \end{pmatrix}$$

Since the eigenvalues of A and that of A' coincides. Hence we compare the value of $\rho(A')$ and $\rho(B)$. We set

$$C = \begin{pmatrix} B & 0 \\ 0 & 0 \end{pmatrix}$$

Then it is obvious that $\rho(B) = \rho(C)$ and $0 \leq C \leq A'$, $C \neq A$. Hence we can apply last lemma and obtain $\rho(B) < \rho(A)$. $\square$

Markov processes

Problem

Prove following;
Suppose that each element of a square matrix $A(t)$ is differentiable. We set $A(t) = (a_1(t), a_2(t), \ldots, a_n(t))$. Then we have

$$\frac{d}{dt} detA(t) = \sum_i \det(a_1(t), \ldots, \frac{d}{dt} a_i(t), \ldots, a_n(t))$$

where $\left(\frac{d}{dt} a\right)_i = \frac{d}{dt}(a(t))_i$. Applying this identity by substituting $A(t) = tI - B$ for constant matrix B we have

$$\frac{d}{dt} \det(tI - B) = \sum_i \det(tI - B_i)$$

where B_i is a principal submatrix by removing i-th row and column from B.

Markov processes

We restart the proof of the Theorem.
(1) and (2) are given by Lemma (extremal vedtor) and Corollary (comparison of matrices).
(3) is given by Lemma (comparison of matrices).
(4) is given as follows. Let A_i be a principal submatrix by removing i-th row and column from A. By Lemma (comparison of spectral radii), we have $\rho(A_i) < \rho(A)$. Hence we have $\det(\rho(A)I - A_i) > 0$ $(\forall i)$. This and the last Problem imply that

$$\frac{d}{dt}\det(tI - A)|_{t=\rho(A)} = \sum_i \det(tI - A_i)|_{t=\rho(A)} > 0$$

This identity implies that $\rho(A)$ is simple. □

Proposition

Suppose that A is a nonnegative irreducible square matrix. If nonnegative vector u satisfies $Au = \lambda u$, then $\lambda = \rho(A)$.

In probability theory, by using this Proposition it is easy to see that for any transition matrix P, we have $\rho(P) = 1$.

Proof

We can apply Perron-Frobenius theorem to ${}^tA$. Hence there is a positive vector v such that ${}^tvA = \rho(A){}^tv$. Since $Au = \lambda u$, u is positive and

$$\rho(A){}^tvu = ({}^tvA)u = {}^tv(Au) = \lambda{}^tvu$$

we have $\rho(A) = \lambda$. $\square$

Proposition

Suppose that A is a positive square matrix. If λ, an eigenvalue of A, satisfies $\lambda \neq \rho(A)$, then we have $|\lambda| < \rho(A)$.

Proof

We can apply Perron-Frobenius theorem to ${}^tA$. Hence there is a positive vector v such that ${}^tvA = \rho(A){}^tv$.

Suppose that $Ax = \lambda x$ and $\lambda \neq x$. Then we have

$$\lambda x_i = \sum_j a_{i,j} x_j, \qquad |\lambda||x_i| = |\lambda x_i| = |\sum_j a_{i,j} x_j| \leq \sum_j a_{i,j} |x_j|$$

Since ${}^t vA = \rho(A){}^t v$, multiplying $v_i > 0$ to both side of the last inequality and summing over i we have

$$|\lambda| \sum_i v_i |x_i| \leq \sum_{i,j} v_i a_{i,j} |x_j| = \rho(A) \sum_j v_j |x_j|$$

If it satisfies $|\lambda| = \rho(A)$, then both this and previous inequalities become identities, namely it satisfies

$$|\sum_j a_{i,j} x_j| = \sum_j a_{i,j} |x_j|$$

Since all elements of A are positive we can regard x as a positive vector (up to multiple complex constant). By using last Proposition, we conclude that if $|\lambda| = \rho(A)$ then $\lambda = \rho(A)$. $\square$

Corollary

Suppose that A is a nonnegative square matrix and $\exists k$ s.t. A^k is a positive square matrix. Then if λ, an eigenvalue of A, satisfies $\lambda \neq \rho(A)$, then $|\lambda| < \rho(A)$.

Note that if an irreducible transition matrix P has period d, then $\rho(P) = 1$ and $\omega, \omega^2, \dots, \omega^{d-1}$ where $\omega = \exp(i2\pi/d)$ are eigenvalues of P.

Problem

Prove this Corollary.
Hint: Give a formula for eigenvalues of A^k by means of eigenvalues of A.

Proposition (limit of transition matrix)

Suppose that P is an irreducible transition matrix with period 1. Then there exists $\lim_{n\to\infty} P^n$. Furthermore we have

$$\lim_{n\to\infty} P^n = \begin{pmatrix} \mu \\ \vdots \\ \mu \end{pmatrix}$$

where $\mu > 0$ is an unique stationary distribution.

Proof

Since P is a transition matrix, if we set $u = {}^t(1, \ldots, 1)$, then

$$Pu = u$$

By applying Perron-Frobenius theorem, we have $\rho(P) = 1$.

Markov processes

Since we assumed that period of P is 1, there is k, large enough such that P^k is positive matrix. Hence by last Corollary, for any λ, eigenvalues of P such that $\lambda \neq 1$ satisfies $|\lambda| < \rho(P) = 1$.
Since 1 is an eigenvalue of P, there is v such that

$$vP = v$$

By applying Perron-Frobenius theorem, we have the eigenvalue 1 is simple and $v > 0$. We normalize v as

$$\mu = \frac{1}{\sum_i v_i} v$$

then we get the unique stationary distribution.

Let $Q^{-1}PQ = \Lambda$ be a diagonalization (or Jordan normal form) of P with $\lambda_1 = 1$. Suppose that we can write $Q = (q_1, q_2, \ldots, q_n)$, $Q^{-1} = {}^t(q'_1, q'_2, \ldots, q'_n)$. Then q_1 and q'_1 are eigenvectors related to the eigenvalue 1 and it satisfies ${}^t q'_1 q_1 = 1$. Hence we can set

$$q_1 = {}^t(1, 1, \ldots, 1), \qquad {}^t q'_1 = (\mu_1, \mu_2, \ldots, \mu_n) = \mu$$

Since for any λ, eigenvalues of P such that $\lambda \neq 1$ satisfies $|\lambda| < \rho(P) = 1$, we have

$$\lim_{n\to\infty} P^n = \lim_{n\to\infty} Q\Lambda^n Q^{-1}$$

$$= Q \begin{pmatrix} 1 & & & \\ & 0 & & \\ & & \ddots & \\ & & & 0 \end{pmatrix} Q^{-1} = \begin{pmatrix} 1 & 0 \cdots 0 \\ 1 & 0 \cdots 0 \\ \vdots & \vdots \\ 1 & 0 \cdots 0 \end{pmatrix} Q^{-1} = \begin{pmatrix} \mu \\ \vdots \\ \mu \end{pmatrix}$$

□