Lecture 7. Q-Process

Tiejun Li[1,2]

[1]School of Mathematical Sciences (SMS),
&
[2]Center for Machine Learning Research (CMLR),
Peking University,
Beijing 100871,
P.R. China
tieli@pku.edu.cn

Office: No. 1 Science Building, Room 1376E

Table of Contents

Poisson Process

Definition (Poisson Process)

Let $X(t)$ be the number of calls received up to time t. $X(t)$ is called a Poisson process if:

Poisson Process

Definition (Poisson Process)

Let $X(t)$ be the number of calls received up to time t. $X(t)$ is called a Poisson process if:

1. $X(0) = 0$;

Poisson Process

Definition (Poisson Process)

Let $X(t)$ be the number of calls received up to time t. $X(t)$ is called a Poisson process if:

1. $X(0) = 0$;
2. $X(t)$ has independent increments, i.e. for any $0 \leq t_1 < t_2 < \cdots < t_n$,

$$X(t_2) - X(t_1), X(t_3) - X(t_2), \ldots, X(t_n) - X(t_{n-1})$$

   are independent;

Poisson Process

Definition (Poisson Process)

Let $X(t)$ be the number of calls received up to time t. $X(t)$ is called a Poisson process if:

1. $X(0) = 0$;
2. $X(t)$ has independent increments, i.e. for any $0 \leq t_1 < t_2 < \cdots < t_n$,
$$X(t_2) - X(t_1), X(t_3) - X(t_2), \ldots, X(t_n) - X(t_{n-1})$$
are independent;
3. for any $t \geq 0, s \geq 0$, we have the distribution of the increment $X(t+s) - X(t)$ is independent of t (time-homogeneous);

Poisson Process

Definition (Poisson Process)

Let $X(t)$ be the number of calls received up to time t. $X(t)$ is called a Poisson process if:

1. $X(0) = 0$;
2. $X(t)$ has independent increments, i.e. for any $0 \leq t_1 < t_2 < \cdots < t_n$,
$$X(t_2) - X(t_1), X(t_3) - X(t_2), \ldots, X(t_n) - X(t_{n-1})$$
are independent;
3. for any $t \geq 0, s \geq 0$, we have the distribution of the increment $X(t+s) - X(t)$ is independent of t (time-homogeneous);
4. for any $t \geq 0, h > 0$, we have
$$\mathbb{P}\{X(t+h) = X(t) + 1 |\ X(t)\} = \lambda h + o(h),$$
$$\mathbb{P}\{X(t+h) = X(t) | X(t)\} = 1 - \lambda h + o(h),$$
$$\mathbb{P}\{X(t+h) \geq X(t) + 2\} = o(h),$$
where λ is called the rate.

Poisson Process: Distribution

Let $p_m(t) = \mathbb{P}\{X(t) = m\}$. Then

$$p_0(t+h) = p_0(t)p_0(h) = p_0(t)(1-\lambda h) + o(h).$$

This gives

$$\frac{p_0(t+h) - p_0(t)}{h} = -\lambda p_0(t) + o(1).$$

As $h \to 0$, we obtain

$$\frac{dp_0(t)}{dt} = -\lambda p_0(t), \quad p_0(0) = 1.$$

The solution is given by

$$p_0(t) = e^{-\lambda t}.$$

Poisson Process: Distribution

Let $p_m(t) = \mathbb{P}\{X(t) = m\}$. Then

$$p_0(t+h) = p_0(t)p_0(h) = p_0(t)(1-\lambda h) + o(h).$$

This gives

$$\frac{p_0(t+h) - p_0(t)}{h} = -\lambda p_0(t) + o(1).$$

As $h \to 0$, we obtain

$$\frac{dp_0(t)}{dt} = -\lambda p_0(t), \quad p_0(0) = 1.$$

The solution is given by

$$p_0(t) = e^{-\lambda t}.$$

For $m > 0$, we have

$$p_m(t+h) = p_m(t)p_0(h) + p_{m-1}(t)p_1(h) + \sum_{i=2}^{m} p_{m-i}(t)p_i(h).$$

Poisson Process: Distribution

From the definition of Poisson process, we get

$$p_m(t+h) = p_m(t)(1-\lambda h) + p_{m-1}(t)\lambda h + o(h).$$

Taking the limit as $h \to 0$, we get

$$\frac{dp_m(t)}{dt} = -\lambda p_m(t) + \lambda p_{m-1}(t)$$

Using the fact $p_m(0) = 0 (m > 0)$, we get

$$p_m(t) = \frac{(\lambda t)^m}{m!} e^{-\lambda t}$$

by induction method. This means that for any fixed t, the distribution of $X(t)$ is Poisson with parameter λt.

Poisson Process: Waiting Time Distribution

- The waiting times can be obtained in the following way. Define

$$\mu_t = \mathbb{P}\{\text{Waiting time} \geq t\}.$$

Poisson Process: Waiting Time Distribution

- The waiting times can be obtained in the following way. Define
$$\mu_t = \mathbb{P}\{\text{Waiting time} \geq t\}.$$
- Then $\mu_0 = 1$, and it obeys $\mu_t - \mu_{t+h} = \mu_t \lambda h + o(h)$, thus $\mu_t' = -\lambda \mu_t$, we get
$$\mu_t = e^{-\lambda t}.$$
i.e. The waiting times are i.i.d. exponentially distributed with rate λ.

Table of Contents

Q-Process: Definition

- Now let us turn to general continuous time Markov chains. We will restrict only on finite state space case in this text. We define
$$p_{ij}(t) = \mathrm{Prob}\{X(t+s) = j | X(s) = i\}.$$

Q-Process: Definition

- Now let us turn to general continuous time Markov chains. We will restrict only on finite state space case in this text. We define
$$p_{ij}(t) = \mathrm{Prob}\{X(t+s) = j | X(s) = i\}.$$
- Here we also assumed the stationarity of the Markov chain, i.e. the right hand side is independent of s. By definition we have
$$p_{ij}(t) \geq 0, \qquad \sum_{j=1}^{N} p_{ij}(t) = 1.$$

Q-Process: Definition

In addition we require that

- $$p_{ii}(h) = 1 - \lambda_i h + o(h), \quad \lambda_i > 0,$$

  It is a statement about the regularity in time of the Markov chain; together with the obvious constraint that $p_{jj}(0) = 1$.

Q-Process: Definition

In addition we require that

- $$p_{ii}(h) = 1 - \lambda_i h + o(h), \quad \lambda_i > 0,$$

  It is a statement about the regularity in time of the Markov chain; together with the obvious constraint that $p_{jj}(0) = 1$.

- $$p_{ij}(h) = \lambda_{ij} h + o(h), \quad j \neq i.$$

  It states that if the process is in state j at time t and a change occurs between t and $t + h$, the process must have jumped to some state $i \neq j$; λ_{ij} is the rate of switching from state i to state j.

Q-Process: Basic Properties

- From the non-negativity and normalization condition of the probability, we have

$$\lambda_{ij} \geq 0, \qquad \sum_{j=1, j\neq i}^{N} \lambda_{ij} = \lambda_i.$$

Q-Process: Basic Properties

- From the non-negativity and normalization condition of the probability, we have

$$\lambda_{ij} \geq 0, \qquad \sum_{j=1, j\neq i}^{N} \lambda_{ij} = \lambda_i.$$

- The Markov property of the process requires the Chapman-Kolmogorov equation

$$p_{ij}(t+s) = \sum_{k=1}^{N} p_{ik}(t) p_{kj}(s).$$

Q-Process: Basic Properties

- From the non-negativity and normalization condition of the probability, we have

$$\lambda_{ij} \geq 0, \qquad \sum_{j=1, j\neq i}^{N} \lambda_{ij} = \lambda_i.$$

- The Markov property of the process requires the Chapman-Kolmogorov equation

$$p_{ij}(t+s) = \sum_{k=1}^{N} p_{ik}(t) p_{kj}(s).$$

- Using matrix notation $P(t) = (p_{ij}(t))$, we can express the Chapman-Kolmogorov relation as

$$P(t+s) = P(t)P(s) = P(s)P(t).$$

Q-Process: Q-matrix

Similarly, if we define

$$Q = \lim_{h \to 0+} h^{-1}(P(h) - I),$$

and denote $Q = (q_{ij})$, it can be stated as

$$q_{ii} = -\lambda_i, \qquad q_{ij} = \lambda_{ij} \quad (i \neq j), \qquad \sum_{j=1}^{N} q_{ij} = 0.$$

Q is called the generator of the Markov chain.

Q-Process: TPM

- Since
$$\frac{P(t+h)-P(t)}{h} = \frac{P(h)-I}{h}P(t)$$
as $s \to 0+$, we get
$$\frac{dP(t)}{dt} = QP(t) = P(t)Q$$

Q-Process: TPM

- Since
$$\frac{P(t+h)-P(t)}{h} = \frac{P(h)-I}{h}P(t)$$
as $s \to 0+$, we get
$$\frac{dP(t)}{dt} = QP(t) = P(t)Q$$
- The solution of this equation is given by
$$P(t) = e^{Qt}P(0) = e^{Qt},$$
since $P(0) = I$.

Q-Process: Forward Kolmogorov Equation

Next we discuss how the distribution of the Markov chain evolves in time.

- Let $\nu(t)$ be the distribution of $X(t)$. Then

$$\begin{aligned}\nu_j(t+dt) &= \sum_{i \neq j} \nu_i(t) p_{ij}(dt) + \nu_j(t) p_{jj}(dt) \\ &= \sum_{i \neq j} \nu_i(t) q_{ij} dt + \nu_j(t)(1 + q_{jj} dt) + o(dt)\end{aligned}$$

  for infinitesimal dt.

Q-Process: Forward Kolmogorov Equation

Next we discuss how the distribution of the Markov chain evolves in time.

- Let $\nu(t)$ be the distribution of $X(t)$. Then

$$\begin{aligned}\nu_j(t+dt) &= \sum_{i\neq j} \nu_i(t)p_{ij}(dt) + \nu_j(t)p_{jj}(dt) \\ &= \sum_{i\neq j} \nu_i(t)q_{ij}dt + \nu_j(t)(1+q_{jj}dt) + o(dt)\end{aligned}$$

  for infinitesimal dt.

- This gives

$$\frac{d\nu(t)}{dt} = \nu(t)Q,$$

  which is called the forward Kolmogorov equation for the distribution.

Q-Process: Distribution

- Its solution can be given as

$$\nu_j(t) = \sum_{i=1}^{N} \nu_i(0) p_{ij}(t),$$

or, in matrix notation,

$$\nu(t) = \nu(0) e^{Qt}.$$

Q-Process: Waiting Time Distribution

- Similar as the Poisson process, we can consider the waiting time distribution for each state j,

$$\mu_j(t) = \mathrm{Prob}\{\tau \geq t | X(0) = j\}.$$

Q-Process: Waiting Time Distribution

- Similar as the Poisson process, we can consider the waiting time distribution for each state j,

$$\mu_j(t) = \mathrm{Prob}\{\tau \geq t | X(0) = j\}.$$

- The same procedure as previous section leads to

$$\frac{d\mu_j(t)}{dt} = q_{jj}\mu_j(t), \qquad \mu_j(0) = 1.$$

Q-Process: Waiting Time Distribution

- Similar as the Poisson process, we can consider the waiting time distribution for each state j,

$$\mu_j(t) = \mathrm{Prob}\{\tau \geq t | X(0) = j\}.$$

- The same procedure as previous section leads to

$$\frac{d\mu_j(t)}{dt} = q_{jj}\mu_j(t), \qquad \mu_j(0) = 1.$$

- Thus the waiting time at state j is exponentially distributed with rate $-q_{jj} = \sum_{k \neq j} q_{jk}$. From the memoryless property of exponential distribution, the waiting time can be counted from any starting point.

Q-Process: Conditional Transition Probability

- It is interesting to investigate the probability

$$p(\theta, j|0, i)d\theta := \text{Prob}\{\text{The jump time } \tau \text{ is in } [\theta, \theta + d\theta) \\ \text{and } X(\tau) = j \text{ given } X(0) = i\}.$$

Q-Process: Conditional Transition Probability

- It is interesting to investigate the probability

$$p(\theta, j|0, i)d\theta := \text{Prob}\{\text{The jump time } \tau \text{ is in } [\theta, \theta + d\theta) \\ \text{and } X(\tau) = j \text{ given } X(0) = i\}.$$

- We have

$$\begin{aligned} p(\theta, j|0, i)d\theta =& \text{Prob}\{\text{No jump occurs in } [0, \theta) \text{ given } X(0) = i\} \\ & \times \text{Prob}\{\text{One jump occurs from } i \text{ to } j \text{ in } [\theta, \theta + d\theta)\} \\ =& \mu_i(\theta) q_{ij} d\theta = \exp(q_{ii}\theta) q_{ij} d\theta. \end{aligned}$$

Q-Process: Conditional Transition Probability

- It is interesting to investigate the probability

$$p(\theta, j|0, i)d\theta := \text{Prob}\{\text{The jump time } \tau \text{ is in } [\theta, \theta + d\theta) \\ \text{and } X(\tau) = j \text{ given } X(0) = i\}.$$

- We have

$$\begin{aligned} p(\theta, j|0, i)d\theta =& \text{Prob}\{\text{No jump occurs in } [0, \theta) \text{ given } X(0) = i\} \\ & \times \text{Prob}\{\text{One jump occurs from } i \text{ to } j \text{ in } [\theta, \theta + d\theta)\} \\ =& \mu_i(\theta) q_{ij} d\theta = \exp(q_{ii}\theta) q_{ij} d\theta. \end{aligned}$$

- Thus we obtain the marginal probability

$$\text{Prob}(X(\tau) = j|X(0) = i) = p(j|0, i) = -\frac{q_{ij}}{q_{ii}} = \frac{q_{ij}}{\sum_{j \neq i} q_{ij}}$$

where τ is the waiting time. These results are particularly useful for the numerical simulation of the trajectories of the Q-process.

Q-Process: Jump Time

Define the jump times of $(X_t)_{t\geq 0}$

$$J_0 = 0, \quad J_{n+1} = \inf\{t : t \geq J_n, X_t \neq X_{J_n}\}, \quad n \in \mathbb{N}$$

where we take the convention $\inf \emptyset = \infty$, and *holding times*

$$H_n = \begin{cases} J_n - J_{n-1}, & \text{if } J_{n-1} < \infty, \\ \infty, & \text{otherwise.} \end{cases}$$

for $n = 1, 2, \ldots$. We define $X_\infty = X_{J_n}$ if $J_{n+1} = \infty$. Define the jump/embedded chain induced by X_t

$$Y_n = X_{J_n}, \quad n \in \mathbb{N}.$$

Q-Process: Embedded Chain

- From strong Markov property[1] and the derivation of $p(\theta, j|0, i)$, we know that the holding times $H_1, H_2, \ldots$ are independent exponential random variables with parameters $q_{Y_0}, q_{Y_1}, \ldots$, respectively, and the jump chain Y_n is a Markov chain with $\tilde{\boldsymbol{Q}}$ as the transition probability matrix, where $\tilde{\boldsymbol{Q}} = (\tilde{q}_{ij})$ defined as

$$\tilde{q}_{ij} = \begin{cases} q_{ij}/q_i, & \text{if } i \neq j \text{ and } q_i > 0, \\ 0, & \text{if } i \neq j \text{ and } q_i = 0, \end{cases}$$

$$\tilde{q}_{ii} = \begin{cases} 0, & \text{if } q_i > 0, \\ 1, & \text{if } q_i = 0. \end{cases}$$

[1] J.R. Norris, Markov Chains, CUP,1997

Q-Process: Embedded Chain

- From strong Markov property[1] and the derivation of $p(\theta, j|0, i)$, we know that the holding times $H_1, H_2, \ldots$ are independent exponential random variables with parameters $q_{Y_0}, q_{Y_1}, \ldots$, respectively, and the jump chain Y_n is a Markov chain with $\tilde{\boldsymbol{Q}}$ as the transition probability matrix, where $\tilde{\boldsymbol{Q}} = (\tilde{q}_{ij})$ defined as

$$\tilde{q}_{ij} = \begin{cases} q_{ij}/q_i, & \text{if } i \neq j \text{ and } q_i > 0, \\ 0, & \text{if } i \neq j \text{ and } q_i = 0, \end{cases}$$

$$\tilde{q}_{ii} = \begin{cases} 0, & \text{if } q_i > 0, \\ 1, & \text{if } q_i = 0. \end{cases}$$

- It is called the jump matrix, and the corresponding Markov chain is called the embedded chain or jump chain of the original Q-process.

[1] J.R. Norris, Markov Chains, CUP,1997

Q-Process: Ergodicity

It is natural to consider the invariant distribution for the Q-processes as in the discrete time Markov chains. From the forward Kolmogorov equation, the invariant distribution must satisfy

$$\pi Q = 0, \quad \pi \cdot \mathbf{1}^T = 1.$$

But to ensure the convergence $\nu(t) \to \pi$, we need the following theorem on the finite state space.

Q-Process: Ergodic Theorem

Theorem (Convergence to equilibrium)

Suppose the matrix Q is irreducible with invariant distribution π, then for all states i, j we have

$$p_{ij}(t) \to \pi_j \text{ as } t \to \infty.$$

Q-Process: Ergodic Theorem

Theorem (Convergence to equilibrium)

Suppose the matrix Q is irreducible with invariant distribution π, then for all states i, j we have

$$p_{ij}(t) \to \pi_j \text{ as } t \to \infty.$$

Note that we do NOT need the primitive condition since in the continuous time case if $q_{ij} > 0$ we have for any $t > 0$

$$\begin{aligned} p_{ij}(t) \geq \mathbb{P}_i(J_1 \leq t, Y_1 = j, H_2 > t) &= \int_0^t e^{-q_i u} q_{ij} du \cdot e^{-q_j t} \\ &= \frac{q_{ij}}{q_i}(1 - e^{-q_i t})e^{-q_j t} > 0 \quad \text{for any } i \end{aligned}$$

For the general case, we have a transition chain such that $i \to k_1 \to \cdots \to k_m \to j$, and we take $\delta t = t/m$ and apply the above argument for each transition state pair. One obtains $p_{i,k_1}(\delta t) > 0, \ldots, p_{k_m,j}(\delta t) > 0$. This yields $p_{ij}(t) > 0$.

Q-Process: Ergodic Theorem

Similarly we also have the ergodic theorem

Theorem (Ergodic theorem)

Suppose the matrix Q is irreducible with invariant distribution π, then for any bounded function f we have

$$\frac{1}{t}\int_0^t f(X(s))ds \to \langle f \rangle_\pi, \qquad a.s.$$

Q-Process: Ergodic Theorem

Similarly we also have the ergodic theorem

Theorem (Ergodic theorem)

Suppose the matrix Q is irreducible with invariant distribution π, then for any bounded function f we have

$$\frac{1}{t}\int_0^t f(X(s))ds \to \langle f \rangle_\pi, \qquad a.s.$$

We should remark that the irreducibility condition is not enough to establish the above ergodic theorems in the countable state space case. We need the so-called positive recurrent condition in both theorems.