MVE550 2019 Lecture 11

Petter Mostad

Chalmers University

December 13, 2019

Distributions with varying parametrizations

- In the **normal distribution** $\text{Normal}(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x - \mu^2)\right)$ the second parameter is the variance.
- However, in R, the second parameter is the standard deviation: `dnorm(x, mu, sigma)`.

- In the **negative binomial distribution** $\text{NegativeBinomial}(x; r, p) = \binom{x+r-1}{x}(1-p)^x p^r$ the second parameter is the chance of success.
- This corresponds to R: `dnorm(x, r, p)`.
- In the original version of the lecture notes, the second parameter was the chance of failure. This has now been changed to the above.
- With the notation above, if $x \sim \text{Poisson}(\theta)$ and $\theta \sim \text{Gamma}(\alpha, \beta)$, then $x \sim \text{NegativeBinomial}(\alpha, \beta/(1+\beta))$.

More distributions with varying parametrizations

- In the **Gamma distribution** $\text{Gamma}(x; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp(-\beta x)$ the second parameter is the *rate*.
- Sometimes, one instead uses the *scale* $1/\beta$ as the second parameter.
- In R, we can use either `dgamma(x, alpha, beta)` or `dgamma(x, alpha, scale=1/beta)`.

- In the **Geometric distribution** $\text{Geometric}(x; p) = (1-p)^x p$ one assumes the possible values for x are $\{0, 1, 2, 3, \ldots\}$.
- Sometimes, the possible values for x are assumed to be $\{1, 2, 3, \ldots\}$ and the probabilities are shifted one step.
- R uses the first definition, e.g., `dgeom(0, 0.3)` is 0.3.

A review of the Exponential distribution

- The **exponential distribution** has density $\text{Exponential}(x; \lambda) = \lambda e^{-\lambda x}$.
- The expectation is $1/\lambda$.
- Sometimes, the expectation is used as the parameter.
- R (and our course) uses the density above: `dexp(x, lambda)` has expectation `1/lambda`.
- The cumulative density is $F(x) = 1 - e^{-\lambda x}$. Thus $\Pr(X > t) = e^{-\lambda t}$.
- The variance is $1/\lambda^2$.
- The exponential distribution is the only distribution on $\mathbb{R}^+$ that is *memoryless*:

$$\Pr(X > s + t \mid X > s) = \Pr(X > t).$$

Overview

- Definition of continuous-time Markov processes.
- Some basic properties.
- The matrix exponential.
- Properties for the long run.

Continuous time Markov chains

- A continuous time stochastic process $\{X_t\}_{t \geq 0}$ with discrete state space S is a *continuous time Markov chain* if

$$P(X_{t+s} = j \mid X_s = i, X_u, 0 \leq u < s) = P(X_{t+s} = j \mid X_s = i)$$

  where $s, t \geq 0$ and $i, j, x_u \in S$.
- The process is *time-homogeneous* if for $s, t \geq 0$ and all $i, j \in S$

$$P(X_{t+s} = j \mid X_s = i) = P(X_t = j \mid X_0 = i)$$

  .
- We then define the *transition function* as the matrix function $P(t)$ with

$$P(t)_{ij} = P(X_t = j \mid X_0 = i)$$

The Chapman-Kolmogorov Equations

For the transition function $P(t)$ we have

- $$P(s+t) = P(s)P(t)$$
- $$P(0) = I$$
- Note similarity to the properties of the exponential function!

Holding times are exponentially distributed

- Define T_i as the time the continuous-time Markov chain started in i stays in i before moving to a different state, so that for any $s > 0$

$$P(T_i > s) = P(X_u = i, 0 \leq u \leq s)$$

- The distribution of T_i is *memoryless* and thus exponential.
- We define q_i so that

$$T_i \sim \text{Exponential}(q_i)$$

- Remember that this means that the average time the process stays in i is $1/q_i$. The *rate* of transition out of the state is q_i.
- Note that we can have $q_i = 0$ meaning that the state i is *absorbing*: $P(T_i > s) = 1$.

The embedded chain

- Define a new stochastic process by listing the states the chain visits. This will be a discrete step Markov chain.
- It is called the *embedded chain*; transition matrix is denoted $\tilde{P}$.
- Note that $\tilde{P}$ has zeros along its diagonal!
- Note that the continuous time Markov chain is completely determined by the expected holding times $(1/q_1, \ldots, 1/q_k)$ and the transition matrix $\tilde{P}$ of the embedded chain. We will soon find explicit formulas relating $(q_1, \ldots, q_k)$ and $\tilde{P}$ with $P(t)$.

Transition rates

A way to describe a continuous-time Markov chain is to describe $k \times (k-1)$ independent "alarm clocks":

- For states i and j so that $i \neq j$, let q_{ij} be the parameter of an Exponentially distributed random variable representing the time until an "alarm clock" rings.
- When in state i, wait until the first alarm clock rings, then move to the state given by the index j of that alarm clock. This defines a continuous-time Markov chain.
- Note that the time until the first alarmclock rings is Exponentially distributed with parameter equal to the sum of the q_{ij}'s. Thus, for the holding time parameter q_i we get

$$q_i = q_{i1} + q_{i2} + \cdots + q_{i,i-1} + q_{i,i+1} + \cdots + q_{ik}$$

- Thus the rate of leaving state i is the sum of the rates of moving to each of the other states.
- The probability that the j'th clock rings first is given by $q_{ij} / \sum_k q_{ik}$. Thus

$$\tilde{P}_{ij} = \frac{q_{ij}}{q_{i1} + \cdots + q_{i,i-1} + q_{i,i+1} + \cdots + q_{ik}} = \frac{q_{ij}}{q_i}$$

The derivative at zero

- To relate $P(t)$ to the q_{ij}'s, we first relate them to $P'(0)$.
- Assuming $P(t)$ is differentiable we get

$$P'(0) = \begin{bmatrix} -q_1 & q_{12} & q_{13} & \cdots & q_{1k} \\ q_{21} & -q_2 & q_{23} & \cdots & q_{2k} \\ q_{31} & q_{31} & -q_3 & \cdots & q_{3k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ q_{k1} & q_{k2} & q_{k3} & \cdots & -q_k \end{bmatrix} = Q$$

  where the q_i and the q_{ij} are those defined earlier.
- In fact we don't need to require a finite state space; discrete is enough.
- Q is called the *(infinitesimal) generator* of the chain.

Kolmogorov Forward Backward

- Prove: We get that for all $t \geq 0$,

$$P'(t) = P(t)Q = QP(t)$$

- Note what this means in terms of the components of $P(t)$:

$$\begin{aligned} P'(t)_{ij} &= -P_{ij}(t)q_j + \sum_{k \neq j} P_{ik}(t)q_{kj} \\ P'(t)_{ij} &= -q_i P_{ij}(t) + \sum_{k \neq i} q_{ik} P_{kj}(t) \end{aligned}$$

- The equations above define a set of differential equations which the components of the matrix function $P(t)$ needs to fulfill.

The matrix exponential

- For any square matrix A define the *matrix exponential* as

$$e^A = \sum_{n=0}^{\infty} \frac{1}{n!} A^n = I + A + \frac{1}{2}A^2 + \frac{1}{6}A^3 + \frac{1}{24}A^4 + \dots$$

- Some important properties:
  - $e^0 = I$.
  - $e^A e^{-A} = I$.
  - $e^{(s+t)A} = e^{sA} e^{tA}$.
  - If $AB = BA$ then $e^{A+B} = e^A e^B = e^B e^A$.
  - $\frac{\partial}{\partial t} e^{tA} = A e^{tA} = e^{tA} A$.
- $P(t) = e^{tQ}$ is the unique solution to the differential equations $P'(t) = QP(t)$ for all $t \geq 0$ and $P(0) = I$.

Computing the matrix exponential

- Assume there exists an invertible matrix S and a matrix D such that $Q = SDS^{-1}$. Then (show!)

$$e^{tQ} = Se^{tD}S^{-1}$$

- If $D = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_k \end{bmatrix}$ is a diagonal matrix, then (show!)

$$e^{tD} = \begin{bmatrix} e^{t\lambda_1} & 0 & \dots & 0 \\ 0 & e^{t\lambda_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & e^{t\lambda_k} \end{bmatrix}.$$

- Recall that if Q is *diagonalizable* it can be written as $Q = SDS^{-1}$ where D is diagonal with the eigenvalues along the diagonal, and S has the corresponding eigenvectors as columns.
- In `R` you may use `expm` from R package `expm`.

Limiting and stationary distributions

- A probability vector v represents a *limiting distribution* if, for all states i and j,
$$\lim_{t \to \infty} P_{ij}(t) = v_j.$$
- A probability vector v represents a *stationary distribution*, if, for all $t \geq 0$,
$$v = vP(t)$$
- A limiting distribution is a stationary distribution but not necessarily vice versa.
- A continuous-time Markov chain is *irreducible* if for all i and j there exists a $t > 0$ such that $P_{ij}(t) > 0$.
- However, periodic continuous-time Markov chains do not exist: If $P_{ij}(t) > 0$ for some $t > 0$ then $P_{ij}(t) > 0$ for all $t > 0$.

The fundamental limit theorem

- An absorbing communication class is one where there is zero probability (i.e., zero rate) of leaving it to other commuication classes.
- For a finite-state continuous-time Markov chain with finite holding time parameters, there are two possibilities:
  - The process is irreducible, and $P_{ij}(t) > 0$ for all $t > 0$ and all i, j.
  - The process contains one or more absorbing communication classes.
- *Fundamental Limit Theorem*: Let $\{X_t\}_{t \geq 0}$ be a finite, irreducible, continuous-time Markov chain with transition funciton $P(t)$. Then there exists a unique stationary distribution vector v which is also the limiting distribution.
- A probability vector v is a stationary distribution of a continuous-time Markov chain with infinitesimal generator Q if and only if $vQ = 0$.