UC Berkeley
Department of Electrical Engineering and Computer Sciences

ELECTRICAL ENGINEERING 126: PROBABILITY AND RANDOM PROCESSES

Final Review
Fall 2017

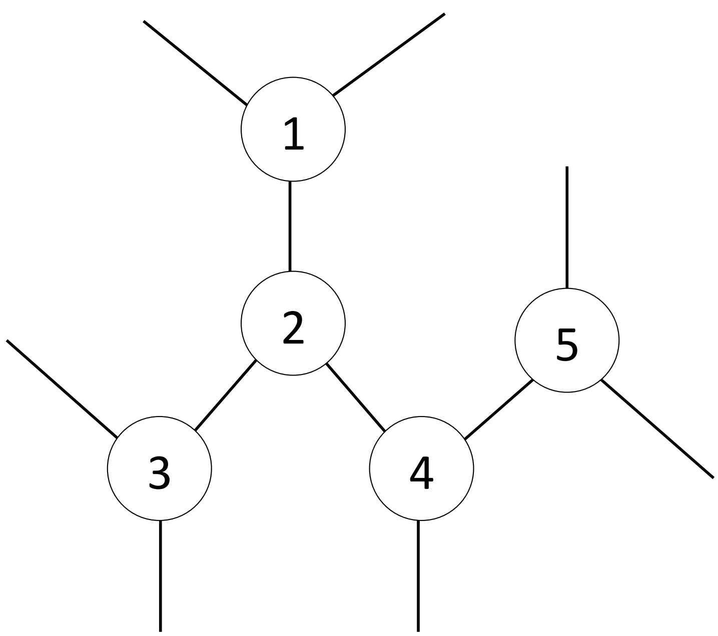
---

1. **Infection Source Detecion**

   Consider a graph where each node represent each person and edges represent connectivity between them. At time 1, the source of the rumor u^* appears. At time 2, the source chooses one of its neighbors, and infects the chosen neighbor. Similarly, in the following time slots, one of the uninfected nodes that are neighboring the nodes that are already infected in the previous time slots is chosen uniformly at random, and get infected. Right after time n, for each $n \in \mathbb{Z}_+$, you observe the infected network with n infected nodes, and you want to detect the source of the infection.

   (a) Consider an infinitely long linear-network: node i is connected with node $(i-1)$ and node $(i+1)$ for all $i \in \mathbb{Z}$.
   At time 11, 11 nodes, $\{-5, -4, \ldots, 4, 5\}$, are infected. Find the MLE of the source of the infection.

   (b) Consider the following infection graph: at time 5, the following 5 nodes are infected. Find the MLE of the source of the infection.



   (c) Consider the same graph. Given that node 4 has twice higher probability of being the source than the others, find the MAP estimate of the source of the infection.

(d) Consider an infinitely large 2D grid: node (i, j) is connected with node $(i+u, j+v)$ for all $(u, v) \in \{(\pm 1, 0), (0, \pm 1)\}$ and all $(i, j) \in \mathbb{Z}^2$.

At time 4, 4 nodes $\{(0,0), (1,0), (0,1), (-1,0)\}$ are infected. Find the MLE of the source of the infection. Which node is the second most likely source?

Solution:

(a) The MLE of the source is as follows.

$$\hat{u} = \arg\max_{u} \mathbb{P}(G \mid u)$$

Note that the number of people that can be chosen is always 2. Thus,

$$\begin{aligned}
\mathbb{P}(G \mid \pm 5) &= \frac{10!}{10!0!} \times \frac{1}{2^{10}} \\
\mathbb{P}(G \mid \pm 4) &= \frac{10!}{9!1!} \times \frac{1}{2^{10}} \\
\mathbb{P}(G \mid \pm 3) &= \frac{10!}{8!2!} \times \frac{1}{2^{10}} \\
\mathbb{P}(G \mid \pm 2) &= \frac{10!}{7!3!} \times \frac{1}{2^{10}} \\
\mathbb{P}(G \mid \pm 1) &= \frac{10!}{6!4!} \times \frac{1}{2^{10}} \\
\mathbb{P}(G \mid 0) &= \frac{10!}{5!5!} \times \frac{1}{2^{10}}
\end{aligned}$$

Thus, the MLE of the source is node 0.

(b) Note that the number of people that can be chosen is always $3, 4, 5, 6$ as the number of infected nodes increase. Thus:

$$\begin{aligned}
\mathbb{P}(G \mid 1) &= \frac{3!}{2!1!} \times \frac{1}{360} = \frac{3}{360} \\
\mathbb{P}(G \mid 2) &= \frac{4!}{2!1!1!} \times \frac{1}{360} = \frac{12}{360} \\
\mathbb{P}(G \mid 3) &= \frac{3!}{2!1!} \times \frac{1}{360} = \frac{3}{360} \\
\mathbb{P}(G \mid 4) &= \frac{4!}{3!1!} \frac{2!}{1!1!} \times \frac{1}{360} = \frac{8}{360} \\
\mathbb{P}(G \mid 5) &= \frac{2!}{1!1!} \times \frac{1}{360} = \frac{2}{360}
\end{aligned}$$

Thus, the MLE of the source is node 2.

To see why these combinatorial expressions hold, consider starting from node 2. In order to infect all of the nodes, the infection must travel up (U) once (to node 1), left (L) once (to node 3), and right (R) two times. Thus, we must take the actions ULRR in any action, and the number of arrangements of this string is $4!/(2!1!1!)$.

(c) Note that

$$\pi(4) = \frac{1}{3},$$
$$\pi(i) = \frac{1}{6}, \qquad i = 1, 2, 3, 5.$$

The MLE of the source is as follows.

$$\hat{u} = \arg\max_u \pi(u)\mathbb{P}(G \mid u)$$

Thus:

$$\begin{aligned}
\pi(1)\mathbb{P}(G \mid 1) &= \frac{1}{6}\frac{3!}{2!1!} \times \frac{1}{360} = \frac{3}{2160} \\
\pi(2)\mathbb{P}(G \mid 2) &= \frac{1}{6}\frac{4!}{2!1!1!} \times \frac{1}{360} = \frac{12}{2160} \\
\pi(3)\mathbb{P}(G \mid 3) &= \frac{1}{6}\frac{3!}{2!1!} \times \frac{1}{360} = \frac{3}{2160} \\
\pi(4)\mathbb{P}(G \mid 4) &= \frac{2}{6}\frac{4!}{3!1!}\frac{2!}{1!1!} \times \frac{1}{360} = \frac{16}{2160} \\
\pi(5)\mathbb{P}(G \mid 5) &= \frac{1}{6}\frac{2!}{1!1!} \times \frac{1}{360} = \frac{2}{2160}
\end{aligned}$$

Thus, the MAP estimate of the source is node 4.

(d) As the size of the set of neighbors depends on path, one cannot just count the number of paths.

$$\begin{aligned}
\mathbb{P}\big(G \bigm| (0,0)\big) &= 4 \cdot \frac{1}{4} \cdot \frac{1}{6} \cdot \frac{1}{7} + 2 \cdot \frac{1}{4} \cdot \frac{1}{6} \cdot \frac{1}{8} = \frac{1}{42} + \frac{1}{96} \\
\mathbb{P}\big(G \bigm| (1,0)\big) = \mathbb{P}\big(G \bigm| (-1,0)\big) &= \frac{1}{4} \cdot \frac{1}{6} \cdot \frac{1}{7} + \frac{1}{4} \cdot \frac{1}{6} \cdot \frac{1}{8} = \frac{1}{168} + \frac{1}{192} \\
\mathbb{P}\big(G \bigm| (0,1)\big) &= 2 \cdot \frac{1}{4} \cdot \frac{1}{6} \cdot \frac{1}{7} = \frac{2}{168}
\end{aligned}$$

Thus, the MLE of the source is node $(0,0)$. Also, node $(0,1)$ is more likely to be the source than node $(1,0)$ or node $(-1,0)$.

2. Voltage MAP

You are trying to detect whether voltage V_1 or voltage V_2 was sent over a channel with independent Gaussian noise $Z \sim N(V_3, \sigma^2)$. Assume that both voltages are equally likely to be sent.

(a) Derive the MAP detector for this channel.

(b) Using the Gaussian Q-function, determine the average error probability for the MAP detector.

(c) Suppose that the average transmit energy is $(V_1^2 + V_2^2)/2$ and that the average transmit energy is constrained such that it cannot be more than $E > 0$. What voltage levels V_1, V_2 should you choose to meet this energy constraint but still minimize the average error probability?

Solution:

(a) Note that since both outputs are equiprobable, the MAP rule is equivalent to the ML rule. Note that the likelihood ratio is:

$$L(y) = \frac{f(y \mid x = V_1)}{f(y \mid x = V_2)}$$

$$\hat{x} = \begin{cases} V_1, & \text{if } L(y) > 1 \\ V_2, & \text{if } L(y) \le 1 \end{cases}$$

Thus, we have:

$$f(y \mid x = V_1) = \frac{1}{\sqrt{2\pi\sigma^2}} \mathrm{e}^{-(y - V_1 - V_3)^2/2\sigma^2}$$

$$f(y \mid x = V_2) = \frac{1}{\sqrt{2\pi\sigma^2}} \mathrm{e}^{-(y - V_2 - V_3)^2/2\sigma^2}$$

so:

$$L(y) = \mathrm{e}^{(V_1 - V_2)(y - (V_1 + V_2)/2 - V_3)/\sigma^2}$$

WLOG, we may assume $V_1 > V_2$, so the ML rule is:

$$\hat{x} = \begin{cases} V_1, & \text{if } y > \dfrac{(V_1 + V_2)}{2} + V_3 \\ V_2, & \text{if } y \le \dfrac{(V_1 + V_2)}{2} + V_3 \end{cases}$$

(b) We have:

$$\begin{aligned} \mathbb{P}(\text{error}) &= \mathbb{P}(\hat{x} = V_2 \mid x = V_1)\mathbb{P}(x = V_1) + \mathbb{P}(\hat{x} = V_1 \mid x = V_2)\mathbb{P}(x = V_2) \\ &= \frac{1}{2}\mathbb{P}\Big(y < \frac{V_1 + V_2}{2} + V_3 \;\Big|\; x = V_1\Big) \\ &\quad + \frac{1}{2}\mathbb{P}\Big(y > \frac{V_1 + V_2}{2} + A_3 \;\Big|\; x = V_2\Big) \\ &= \frac{1}{2}\Big[1 - Q\Big(\frac{V_2 - V_1}{2\sigma}\Big)\Big] + \frac{1}{2}Q\Big(\frac{V_1 - V_2}{2\sigma}\Big) \\ &= Q\Big(\frac{V_1 - V_2}{2\sigma}\Big). \end{aligned}$$

(c) We would like to

$$\text{minimize } Q\Big(\frac{V_1 - V_2}{2\sigma}\Big) \text{ subject to } \frac{V_1^2 + V_2^2}{2} \le E.$$

Note that this is equivalent to maximizing $V_1 - V_2$ which is equivalent to maximizing $(V_1 - V_2)^2$, subject to the same constraint. Now, we have:

$$(V_1 - V_2)^2 \le (|V_1| + |V_2|)^2 \le 4E$$

where we have equality iff $V_1 = -V_2$, so the optimal choice is $V_1 = \sqrt{E}$, $V_2 = -\sqrt{E}$.

3. **Conditional Expectation Identity**

Prove that $\mathbb{E}[X\,\mathbb{E}[Y \mid Z]] = \mathbb{E}[Y\,\mathbb{E}[X \mid Z]]$.

Solution:

Observe that

$$\mathbb{E}\big[\mathbb{E}[X \mid Z]\,\mathbb{E}[Y \mid Z]\big] = \mathbb{E}[\mathbb{E}\{X\,\mathbb{E}[Y \mid Z] \mid Z\}] = \mathbb{E}\big[X\,\mathbb{E}[Y \mid Z]\big].$$

Similarly, $\mathbb{E}[\mathbb{E}[X \mid Z]\,\mathbb{E}[Y \mid Z]] = \mathbb{E}[Y\,\mathbb{E}[X \mid Z]]$. The identity follows.

4. **Geometric MMSE**

Let N be a geometric random variable with parameter $1-p$, and $(X_i)_{i\in\mathbb{N}}$ be i.i.d. exponential random variables with parameter λ. Let $T = X_1 + \cdots + X_N$. Compute the LLSE and MMSE of N given T.

Solution:

First, we calculate $\mathbb{P}(N = n \mid T = t)$, for $t > 0$ and $n \in \mathbb{Z}_+$.

$$\begin{aligned}
\mathbb{P}(N = n \mid T = t) &= \frac{\mathbb{P}(N = n) f_{T|N}(t \mid n)}{\sum_{k=1}^{\infty} \mathbb{P}(N = k) f_{T|N}(t \mid k)} \\
&= \frac{(1-p)p^{n-1}\lambda^n t^{n-1} \mathrm{e}^{-\lambda t}/(n-1)!}{\sum_{k=1}^{\infty}(1-p)p^{k-1}\lambda^k t^{k-1}\mathrm{e}^{-\lambda t}/(k-1)!} \\
&= \frac{\lambda(\lambda p t)^{n-1}/(n-1)!}{\lambda\sum_{k=1}^{\infty}(\lambda p t)^{k-1}/(k-1)!} = \frac{(\lambda p t)^{n-1}}{\mathrm{e}^{\lambda p t}(n-1)!}, \qquad n \in \mathbb{Z}_+.
\end{aligned}$$

Next, we calculate $\mathbb{E}[N \mid T = t]$.

$$\begin{aligned}
\mathbb{E}[N \mid T = t] &= \sum_{n=1}^{\infty} n \frac{(\lambda p t)^{n-1}}{\mathrm{e}^{\lambda p t}(n-1)!} \\
&= \sum_{n=1}^{\infty} \frac{(\lambda p t)^{n-1}}{\mathrm{e}^{\lambda p t}(n-1)!} + \sum_{n=1}^{\infty}(n-1)\frac{(\lambda p t)^{n-1}}{\mathrm{e}^{\lambda p t}(n-1)!} \\
&= 1 + \frac{\lambda p t}{\mathrm{e}^{\lambda p t}} \sum_{n=2}^{\infty} \frac{(\lambda p t)^{n-2}}{(n-2)!} = 1 + \frac{\lambda p t}{\mathrm{e}^{\lambda p t}} \mathrm{e}^{\lambda p t} = 1 + \lambda p t.
\end{aligned}$$

Hence, the MMSE is $\mathbb{E}[N \mid T] = 1 + \lambda p T$. The MMSE is linear, so it is also the LLSE.

In terms of a Poisson process, T represents the first arrival of a marked Poisson process with rate λ, where arrivals are marked independently with probability $1-p$. The marked Poisson process has rate $\lambda(1-p)$. The unmarked points form a Poisson process of rate λp. In time T, the expected number of unmarked points is $\lambda p T$, so the conditional expectation of the number of points at time T, N, is $1 + \lambda p T$.

5. **Machine**

A machine, once in production mode, operates continuously until an alarm signal is generated. The time up to the alarm signal is an exponential random variable with parameter 1. Subsequent to the alarm signal, the machine is tested for an exponentially distributed amount of time with parameter 5. The test results are positive, with probability 1/2, in which case the machine returns to production mode, or negative, with probability 1/2, in which case the machine is taken for repair. The duration of the repair is exponentially distributed with parameter 3.

(a) Let states $1, 2, 3$ correspond to production mode, testing, and repair, respectively. Let $(X(t))_{t\geq 0}$ denote the state of the system at time t. Is $(X(t))_{t\geq 0}$ a CTMC?

(b) Find the rate and transition matrices.

(c) Find the steady state probabilities.

Solution:

(a) We note that given that the current state is $i \in \{1, 2, 3\}$ in this process, the future is independent of the past.

(b) The transition rates are $\nu_1 = 1, \nu_2 = 5, \nu_3 = 3$ and the transition probabilities are given, so we have:

$$P = \begin{bmatrix} 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \\ 1 & 0 & 0 \end{bmatrix}, \qquad Q = \begin{bmatrix} -1 & 1 & 0 \\ 5/2 & -5 & 5/2 \\ 3 & 0 & -3 \end{bmatrix}.$$

(c) We set up the balance equations:

$$\begin{aligned} \pi_1 &= \frac{5}{2}\pi_2 + 3\pi_3 \\ 5\pi_2 &= \pi_1 \\ 3\pi_3 &= \frac{5}{2}\pi_2 \\ \pi_1 + \pi_2 + \pi_3 &= 1 \end{aligned}$$

We thus have:

$$\pi_1 = \frac{30}{41}, \qquad \pi_2 = \frac{6}{41}, \qquad \pi_3 = \frac{5}{41}$$