EE3731C Tutorial - Statistical Signal 3

Department of Electrical and Computer Engineering

1.

$$\begin{aligned}
\sigma^2_{ML} &= \underset{\sigma^2}{\operatorname{argmax}}\, p(x_1 \cdots x_N | \sigma^2) \\
&= \underset{\sigma^2}{\operatorname{argmax}} \prod_{n=1}^{N} p(x_n|\sigma^2) \quad \text{By conditional independence} \\
&= \underset{\sigma^2}{\operatorname{argmax}} \log \prod_{n=1}^{N} p(x_n|\sigma^2) \\
&= \underset{\sigma^2}{\operatorname{argmax}} \sum_{n=1}^{N} \log p(x_n|\sigma^2) \\
&= \underset{\sigma^2}{\operatorname{argmax}} \sum_{n=1}^{N} \left[-\log \sigma - \frac{(x_n-\mu)^2}{2\sigma^2} \right] \quad \text{since } p(x_n|\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_n-\mu)^2}{2\sigma^2}} \\
&= \underset{\sigma^2}{\operatorname{argmax}} -N \log \sigma - \sum_{n=1}^{N} \frac{(x_n-\mu)^2}{2\sigma^2} \qquad (1)
\end{aligned}$$

Differentiating Eq. (1) with respect to σ, we get

$$\frac{\partial}{\partial \sigma} \left(-N \log \sigma - \sum_{n=1}^{N} \frac{(x_n-\mu)^2}{2\sigma^2} \right) = -\frac{N}{\sigma} + \sum_{n=1}^{N} \frac{(x_n-\mu)^2}{\sigma^3}$$

Setting the above to 0, we get

$$\sigma^2_{ML} = \frac{1}{N} \sum_{n=1}^{N} (x_n - \mu)^2$$

2. Here are three examples where the conditions of Fundamental Theorem of Markov Chain does not hold

   (a) Consider the random walk x_n defined in Q3. x_n does not satisfy the Fundamental Theorem of Markov Chain. To see this, observe that at any time n_0, $p(x_{n_0} = n_0 + 1) = 0$ (e.g., at time 5, there is zero probability of $x_5 = 6$).

   (b) Consider a Markov chain x_n with two states s_1 and s_2, where $p(x_{n+1} = s_2|x_n = s_1) = 1$ and $p(x_{n+1} = s_1|x_n = s_2) = 1$. In other words, the Markov chain will oscillate between the two states with certainty. Therefore, if $x_1 = s_1$, then $x_n = s_1$ for all odd time points and $x_n = s_2$ for all even time points. Therefore, there is no such n_0, such that for all time $n > n_0$, any two states have non-zero probability of communicating. Note that for this Markov chain, there is a unique stationary distribution $[1/2 \;\; 1/2]$, but we cannot start at any state and eventually converge at the stationary distribution. For example, if we start at state s_1, π will never converge to $[1/2 \;\; 1/2]$.

(c) Consider a Markov chain x_n with four states s_1, s_2, s_3 and s_4. Let

$$T = \begin{bmatrix} 1/2 & 1/2 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 \\ 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 1/2 & 1/2 \end{bmatrix}$$

We see that states s_1 and s_2 can transition among each other. Similarly, s_3 and s_4 can also transition among each other. However, states s_1 and s_2 cannot transition to states s_3 or s_4 (and vice versa). Therefore this Markov chain does not have a unique stationary distribution. If the Markov chain starts in state s_1 or s_2, then $\pi_\infty(s_1, s_2, s_3, s_4) = [1/2 \; 1/2 \; 0 \; 0]$. Similarly, if the Markov chain starts in state s_3 or s_4, then $\pi_\infty(s_1, s_2, s_3, s_4) = [0 \; 0 \; 1/2 \; 1/2]$.

3.

$$\begin{aligned} &p\left(\max_{1\le n\le 20} y_n = 10 | y_{20} = 0\right) \\ &= \frac{p\{\max_{1\le n\le 20} y_n = 10, y_{20} = 0\}}{p(y_{20} = 0)} \\ &= \frac{\#\text{paths where } \max_{1\le n\le 20} y_n = 10 \text{ and } y_{20} = 0}{\#\text{paths where } p(y_{20} = 0)} \quad \text{because each path is equally likely} \end{aligned}$$

There are only 2 paths where $\max_{1\le n\le 20} y_n = 10$ and $y_{20} = 0$ (see figure below). The first path (in blue) is $z_1 = \cdots = z_{10} = +1$ and $z_{11} = \cdots = z_{20} = -1$. The second path (in red) is $z_1 = \cdots = z_{10} = -1$ and $z_{11} = \cdots = z_{20} = +1$.

On the other hand, for $p(y_{20} = 0)$, 10 of the z_n must be $+1$ and 10 of the z_n must be -1. Therefore there are $\binom{20}{10}$ paths for $p(y_{20} = 0)$. Therefore

$$p\left(\max_{1\le n\le 20} y_n = 10 | y_{20} = 0\right) = \frac{2}{\binom{20}{10}}$$

4. Let $x_N = k$. If x_0 is even then $N + k$ must be even. If x_0 is odd then $N + k$ must be odd. Since $x_{11} = 2$, this means that x_0 must be odd, which means that $p(x_0 = -2 | x_{11} = 2) = p(x_0 = 0 | x_{11} = 2) = p(x_0 = 2 | x_{11} = 2) = 0$. Therefore, let's

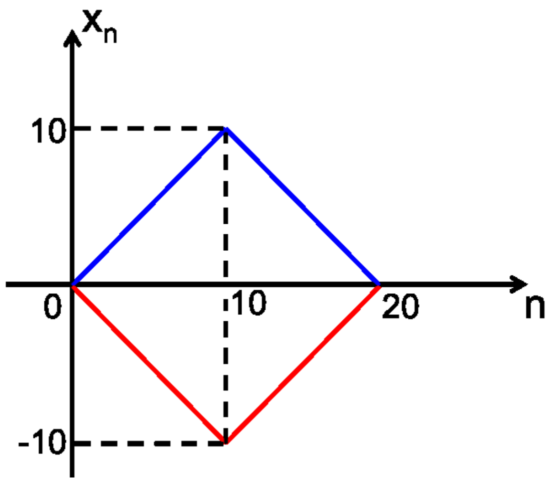

Figure 1: Only two possible paths

use Bayes rule to compute the remaining probabilities:

$$\begin{aligned} p(x_0 = -1 | x_{11} = 2) &= \frac{p(x_0 = -1) p(x_{11} = 2 | x_0 = -1)}{p(x_{11} = 2)} \\ &= \frac{\frac{1}{5} p(\text{seven of the } z_n \text{ are } +1 \text{ and four of the } z_n \text{ are -1})}{p(x_{11} = 2)} \\ &= \frac{\frac{1}{5}\binom{11}{7}\frac{1}{2^{11}}}{p(x_{11} = 2)} \end{aligned}$$

and

$$\begin{aligned} p(x_0 = 1 | x_{11} = 2) &= \frac{p(x_0 = 1) p(x_{11} = 2 | x_0 = 1)}{p(x_{11} = 2)} \\ &= \frac{\frac{1}{5} p(\text{six of the } z_n \text{ are } +1 \text{ and five of the } z_n \text{ are -1})}{p(x_{11} = 2)} \\ &= \frac{\frac{1}{5}\binom{11}{6}\frac{1}{2^{11}}}{p(x_{11} = 2)} \end{aligned}$$

Note that $p(x_{11} = 2) = p(x_0 = 1) p(x_{11} = 2 | x_0 = 1) + p(x_0 = -1) p(x_{11} = 2 | x_0 =$

$-1) = \frac{1}{5}\binom{11}{7}\frac{1}{2^{11}} + \frac{1}{5}\binom{11}{6}\frac{1}{2^{11}}$. Therefore

$$\begin{aligned}
p(x_0 = -1 | x_{11} = 2) &= \frac{\frac{1}{5}\binom{11}{7}\frac{1}{2^{11}}}{\frac{1}{5}\binom{11}{7}\frac{1}{2^{11}} + \frac{1}{5}\binom{11}{6}\frac{1}{2^{11}}} \\
&= \frac{\binom{11}{7}}{\binom{11}{7} + \binom{11}{6}} \\
&= \frac{\frac{11!}{7!4!}}{\frac{11!}{7!4!} + \frac{11!}{6!5!}} \\
&= \frac{1}{1 + \frac{7!4!}{6!5!}} \\
&= \frac{1}{1 + \frac{7}{5}} \\
&= \frac{5}{12}
\end{aligned}$$

Similarly, $p(x_0 = 1 | x_{11} = 2) = 7/12$. Therefore

$$p(x_0 | x_{11} = 2) = \begin{cases} \frac{5}{12} & x_0 = -1 \\ \frac{7}{12} & x_0 = 1 \\ 0 & \text{otherwise} \end{cases}$$

5. Let's start by considering a few facts about expectation. Suppose the continuous random variable y is a function of two continuous random variables y_1 and y_2, i.e., $y = g(y_1, y_2)$. Then

$$E_{p(y)}(y) \triangleq \int y p(y) dy = \int\int g(y_1, y_2) p(y_1, y_2) dy_1 dy_2 \triangleq E_{p(y_1, y_2)}(g(y_1, y_2)).$$

The above equation might look scary, but it is saying something intuitive: sampling y from $p(y)$ repeatedly and averaging is equivalent to sampling y_1 and y_2 from $p(y_1, y_2)$ repeatedly, computing $g(y_1, y_2)$ and averaging. Note that there is some (possibly complex) relationship between $p(y)$, $p(y_1, y_2)$ and g, so that the two integrals work out to be equal. We can replace the integral with summation if we are dealing with discrete, instead of continuous random variables. The above is true even if we condition on observing some random variable w:

$$E_{p(y|w)}(y) \triangleq \int y p(y|w) dy = \int\int g(y_1, y_2) p(y_1, y_2 | w) dy_1 dy_2 \triangleq E_{p(y_1, y_2|w)}(g(y_1, y_2))$$

The above equation again looks scary, but it is saying that sampling y from $p(y|w)$ repeatedly and averaging is equivalent to sampling y_1 and y_2 from $p(y_1, y_2|w)$ repeatedly, computing $g(y_1, y_2)$ and averaging.

Relevant to this tutorial question is when $g(y_1, y_2) = y_1 + y_2$. Applying the above equation, we get

$$\begin{aligned}
E_{p(y|w)}(y) &= \int\int (y_1 + y_2)p(y_1, y_2|w)dy_1dy_2 \\
&= \int\int y_1p(y_1, y_2|w)dy_1dy_2 + \int\int y_2p(y_1, y_2|w)dy_1dy_2 \\
&= \int y_1p(y_1|w)dy_1 + \int y_2p(y_2|w)dy_2 \\
&\triangleq E_{p(y_1|w)}(y_1) + E_{p(y_2|w)}(y_2)
\end{aligned}$$

We can extend the above equation to more variables. For example, suppose $y = y_1 + y_2 + y_3$, then

$$E_{p(y|w)}(y) = E_{p(y_1|w)}(y_1) + E_{p(y_2|w)}(y_2) + E_{p(y_3|w)}(y_3) \tag{2}$$

Importantly, we note that $\underline{y_1, y_2, y_3 \text{ and } w \text{ do NOT have to be independent.}}$

Now, let's go back to our original problem. The MMSE estimate of x_4 is

$$\begin{aligned}
x_{4,MMSE} &= E_{p(x_4|x_1=4,x_2=2,0\leq x_3\leq 4)}(x_4) \\
&= E_{p(x_4|x_1=4,x_2=2,-2\leq z_3\leq 2)}(x_4) \qquad \text{because } x_3 = x_2 + z_3
\end{aligned}$$

Note that $x_4 = x_2 + z_3 + z_4$ (position at timepoint 4 is equal to the position at timepoint 2 plus the steps taken at timepoints 3 and 4). Therefore we can apply Eq.(2) by thinking of y as x_4 and y_1, y_2, y_3 as x_2, z_3, z_4. We can also think of w as the event that $x_1 = 4, x_2 = 2, -2 \leq z_3 \leq 2$. Therefore, we get

$$\begin{aligned}
x_{4,MMSE} = E_{p(x_2|x_1=4,x_2=2,-2\leq z_3\leq 2)}(x_2) \\
+ E_{p(z_3|x_1=4,x_2=2,-2\leq z_3\leq 2)}(z_3) \\
+ E_{p(z_4|x_1=4,x_2=2,-2\leq z_3\leq 2)}(z_4)
\end{aligned}$$

Let's consider each of these three terms:

- $E_{p(x_2|x_1=4,x_2=2,-2\leq z_3\leq 2)}(x_2)$: Since we are conditioning on $x_2 = 2$, then

  $$p(x_2|x_1 = 4, x_2 = 2, -2 \leq z_3 \leq 2) = \delta(z_2 - 2),$$

  where δ is the dirac delta function. In plain English, suppose we observe x_2 to be 2, then x_2 must be equal to 2. Therefore $E_{p(x_2|x_1=4,x_2=2,-2\leq z_3\leq 2)}(x_2) = 2$.
- $E_{p(z_3|x_1=4,x_2=2,-2\leq z_3\leq 2)}(z_3)$: z_3 is a random step at timepoint 3 independent of the position at timepoint 1 (x_1) and timepoint 2 (x_2). Therefore we have

  $$\begin{aligned}
  p(z_3|x_1 = 4, x_2 = 2, -2 \leq z_3 \leq 2) &= p(z_3| - 2 \leq z_3 \leq 2) \\
  &= \begin{cases} \frac{p(z_3)}{p(-2\leq z_3\leq 2)} & \text{for } -2 \leq z_3 \leq 2 \\ 0 & \text{for } z_3 < -2 \text{ or } z_3 > 2 \end{cases}
  \end{aligned}$$

Observe that the probability distribution $p(z_3|-2 \leq z_3 \leq 2)$ is symmetric about 0, so we have

$$E_{p(z_3|x_1=4,x_2=2,-2\leq z_3\leq 2)}(z_3) = E_{p(z_3|-2\leq z_3\leq 2)}(z_3) = 0$$

- $E_{p(z_4|x_1=4,x_2=2,-2\leq z_3\leq 2)}(z_4)$: z_4 is a random step at timepoint 4 independent of the positions at timepoint 1 (x_1) and timepoint 2 (x_2) and the random step taken at timepoint 3 (z_3), therefore $p(z_4|x_1 = 4, x_2 = 2, -2 \leq z_3 \leq 2) = p(z_4)$, which is a Gaussian distribution with mean 0. Therefore $E_{p(z_4|x_1=4,x_2=2,-2\leq z_3\leq 2)}(z_4) = 0$

Putting the above three terms together, we have $x_{4,MMSE} = 2 + 0 + 0 = 2$.