UC Berkeley
Department of Electrical Engineering and Computer Sciences

EE126: Probability and Random Process

Final Solution
Fall 2014

Problem 1. (a) First we find the conditional pdf:

$$f(y_1 \dots, y_n|x) = \frac{1}{x^n} e^{-\frac{1}{x}\sum_i y_i}.$$

Thus, $\hat{X}_{MAP} = 1$ if

$$pe^{-\sum_i y_i} > (1-p)/2^n . e^{-\frac{1}{2}\sum_i y_i},$$

and $\hat{X}_{MAP} = 2$ otherwise.

(b) The statement is wrong. For example, $X = Y = U[0,1]$!

(c) Let $X \sim N(0,1)$ and $Y = X$ with probability 1/2 and $Y = -X$ with probability 1/2. Clearly X and Y are not jointly Gaussian, but one can easiliy check that Y is normal distributed.

(d) Given $T_{10} = t$, the previous arrivals are uniformly distributed between 0 and t. Thus, the second arrival has expected value of $2t/10$.

(e) Let q_n be the probability of having a path to level n. Similar to HW 2, we have

$$q_n = 3p.q_{n-1} - 3p^2 q_{n-1}^2 + p^3 q_{n-1}^3.$$

Note that in the derivation we used the property

$$\Pr(A \cup B \cup c) = \Pr(A) + \Pr(B) + \Pr(C) - \Pr(A \cap B) - \Pr(A \cap C) - \Pr(B \cap C) + \Pr(A \cap B \cap C).$$

Problem 2. (a) By memoryless property, $E[Y_2|Y_1] = Y_1 + 1/\lambda_A$. The joint pdf is

$$f(y_2, y_1) = f(y_1)f(y_2|y_1) = \lambda_A e^{-\lambda_A y_1} \lambda_A e^{-\lambda_A (y_2 - y_1)} 1\{0 \le y_1 \le y_2\} = \lambda_A^2 e^{-\lambda_A y_2} 1\{0 \le y_1 \le y_2\}.$$

(b) Let N_t be the number of emails sent by time t. We have

$$\arg\max_\lambda \Pr(N_1 = 5|\lambda) = \arg\max_\lambda e^{-\lambda}\lambda^5/5!.$$

Thus, taking log of the expression and setting derivative of it to 0 we have $\lambda_{ML} = 5$ which is also intuitive.

(c) Let's say we observe $N = n$. Then, we estimate λ to be n. Thus, we need to find c such that $\Pr(\lambda \in (n-c, n+c)) = 0.95$. Equivalently, we can find c such that

$$\Pr(|n - \lambda| > c) = 0.05 \Rightarrow \Pr(\frac{n-\lambda}{\sqrt{\lambda}} > c/\sqrt{\lambda}) \simeq \Pr(\frac{n-\lambda}{\sqrt{\lambda}} > c/\sqrt{n}) = 0.025.$$

Thus, from the table we find that $c = 2\sqrt{n}$.

(d) The sum of two independent Poisson random variables is again Poisson. The number of emails Alice sends in $[0, 1]$ is Poisson distributed with rate λ_A and the number of emails both send in $[1, 3]$ is Poisson distributed with rate $2\lambda_A + 2\lambda_B$. So $\Pr(N_{sum} = n) = \frac{(3\lambda_A+2\lambda_B)^n e^{-3\lambda_A-2\lambda_B}}{n!}$.

(e) By memoryless property of Poisson process, the expected time until the email is finished is $1/\lambda_A$. The expected time from the starting time, S, can be found as follows. If Alice has not sent any emails this time is 1. If Alice has sent an email this time is again exponential with rate λ_A; call it T. So $S = \min(1, T)$. We can find $E(S)$ as follows.

$$E(S) = 1 \times e^{-\lambda_A} + \int_{s=0}^{1} s\lambda_A e^{-\lambda s} ds = \frac{1}{\lambda_A}(1 - e^{-\lambda_A}).$$

Therefore, the expected total typing time is $\frac{1}{\lambda_A}(2 - e^{-\lambda_A})$.

(f) Let A be the event that Alice sends 4 email in $[0, 2]$ and B be the event that a total of 10 emails are sent in $[0, 2]$. Then,

$$\begin{aligned}\Pr(A|B) &= \Pr(A \cap B)/\Pr(B)\\ &= \frac{(2\lambda_A)^4 e^{-2\lambda_A}/4! \times (\lambda_B)^6 e^{-\lambda_B}/6!}{(2\lambda_A + \lambda_B)^1 0 e^{-2\lambda_A-\lambda_B}/10!}\\ &= \binom{10}{4}(\frac{2\lambda_A}{2\lambda_A + \lambda_B})^4(\frac{\lambda_B}{2\lambda_A + \lambda_B})^6.\end{aligned}$$

Problem 3. (a) The transition diagram is shown in Figure 1.

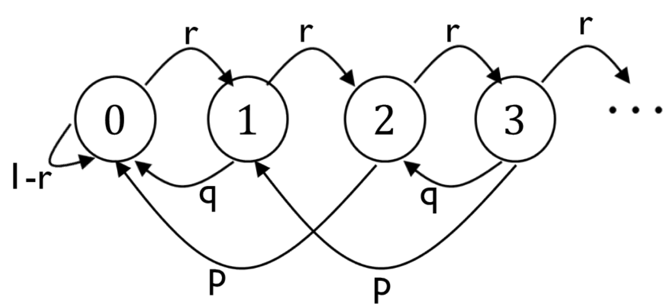

Figure 1: Markov chain.

(b) $\pi(2) = r\pi(1) + q\pi(3) + p\pi(4)$.

(c) We have that $X_n - X_{n-1} \geq V_n$ where V_n is an iid sequence with the following PMF: $\Pr(V = -2) = p$, $\Pr(V_n = -1) = q$ and $\Pr(V_n = 1) = r$. Note that the greater than or equal is because of the boundary condition, $X_n = 0$ or $X_n = 1$. Now summing both sides of the inequality over n we have

$$X_n \geq X_0 + \sum_{i=1}^{n} V_i \Rightarrow \frac{X_n}{n} \geq \frac{X_0}{n} + \frac{\sum_{i=1}^{n} V_i}{n}.$$

Thus, by law of large numbers as n tends to infinity, we have $\frac{X_n}{n} \geq E[V] = r - 2p - q > 0$. Thus, the Markov chain is transient since X_n grows linearly with n as n gets large.

Problem 4. (a) We know that $L[X|Y] = E(X) + \frac{cov(X,Y)}{var(Y)}(Y - E(Y))$. We calculate each term: $E(X) = E(Y^2) = 1/3$, $E(Y) = 0$, $var(Y) = 1/3$,

$$cov(X,Y) = E(XY) - E(X)E(Y) = E(Y^2 + Y^3 + 2ZY) = 1/3.$$

So $L[X|Y] = 1/3 + Y$.

(b) First, note that the pdf of Y and Z is symmetric around 0. Now by orthogonality principle we have

$$\begin{aligned}
&E[X - aY^2 - bY - c] = 0 \Rightarrow 1/3 - a/3 - c = 0 \\
&E[XY - aY^3 - bY^2 - cY] = 0 \Rightarrow 1/3 - b/3 = 0 \\
&E[XY^2 - aY^4 - bY^3 - cY^2] = 0 \Rightarrow (1-a) \times 2/5 - c/3 = 0.
\end{aligned}$$

For the last equation we used $E[Y^4] = 2\int_0^1 y^4 dy = 2/5$ and

$$E[XY^2] = E[Y^3 + 2ZY^2 + Y^4] = E[Y^4] = 2/5.$$

Problem 5. (Viterbi algorithm)

(a) Table 1 summarizes the state transition diagram.

$x[n-1]$	$x[n]$	$y_0[n]$	$y_1[n]$
0	0	0	0
0	1	1	0
1	0	1	1
1	1	0	1

Table 1: Truth table

From this, it is clear that $y_0[n] = x[n-1] + x[n]$ and $y_1[n] = x[n-1]$. Thus, the following circuit in Figure 2 implements the given encoder.

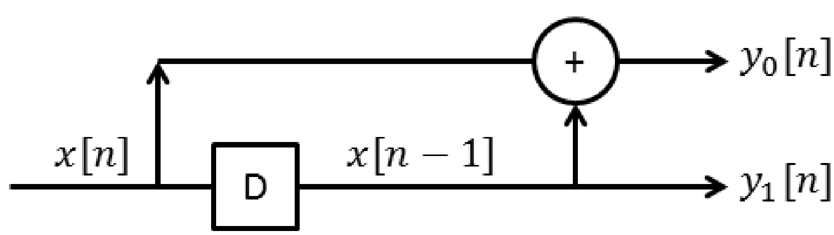

Figure 2: An example of circuit implementing the given encoder

(b) Figure 3 shows the one stage of the trellis-diagram.

(c) Table 2 shows the output sequence.

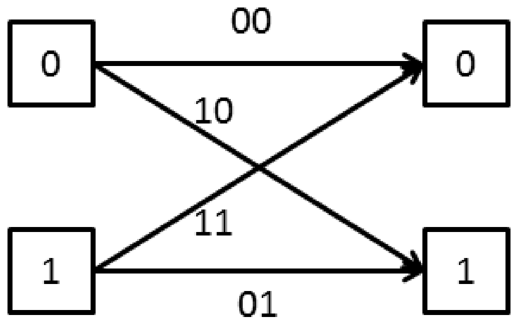

Figure 3: One stage of the trellis-diagram

n	0	1	2	3	4
$x[n]$	0	1	0	1	1
$y_0[n]$		1	1	1	0
$y_1[n]$		0	1	0	1

Table 2: Output sequence

(d) Figure 4 depicts the trellis-diagram. The MAP estimate of $\{x[n]\}$ is $(1, 0, 0, 1)$.

Problem 6. (EM algorithm)

(a) It is clear that the following estimates are the ML estimates.

$$\hat{\theta_A} = \frac{\text{Number of heads from type-A coins}}{10 \times \text{Number of type-A coins}} = \frac{5}{6}$$
$$\hat{\theta_B} = \frac{\text{Number of heads from type-B coins}}{10 \times \text{Number of type-B coins}} = \frac{1}{3}$$

(b)

$$\hat{\theta_A} = \frac{\sum_{z_i=A} h_i}{10 \sum_{z_i=A} 1} = \frac{5}{6}$$
$$\hat{\theta_B} = \frac{\sum_{z_i=B} h_i}{10 \sum_{z_i=B} 1} = \frac{1}{3}$$

(c) We want to maximize the likelihood of θ given y, i.e.,

$$\mathcal{L} = f(y|\theta) = \sum_z f(y, z|\theta) = \sum_z f(y|z, \theta) f(z|\theta).$$

Since z is independent of θ, one can instead maximize the following quantity.

$$\begin{aligned}
\mathcal{L}_1 &= \sum_z f(y|z, \theta) = \sum_z \prod_{i=1}^{3} f(y_i|z_i, \theta) \\
&= \sum_z \prod_{i=1}^{3} \left[\mathbf{1}\{z_i = A\} \left(\theta_A^{h_i} (1 - \theta_A)^{3-h_i} \right) + \mathbf{1}\{z_i = B\} \left(\theta_B^{h_i} (1 - \theta_B)^{3-h_i} \right) \right]
\end{aligned}$$

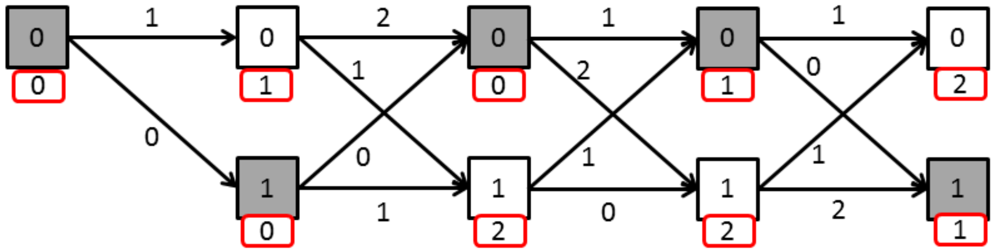

Figure 4: The trellis-diagram corresponding to the output sequence

Now, the summation is over all possible labels: $z \in \{A, B\}^3$. Even though the computation can be heavy, one can always find the MLE by optimizing $\mathcal{L}_1$ over θ_A and θ_B. The scaled objective $\mathcal{L}_1$ is plotted in Figure 5. The MLE estimates of (θ_A, θ_B) are $(\frac{2}{3}, \frac{2}{3})$. This can be indeed seen by observing the output sequences are symmetric, or the $\mathcal{L}_1$ is symmetric. Thus, $\theta_A^{ML} = \theta_B^{ML} = \frac{6}{9} = \frac{2}{3}$.

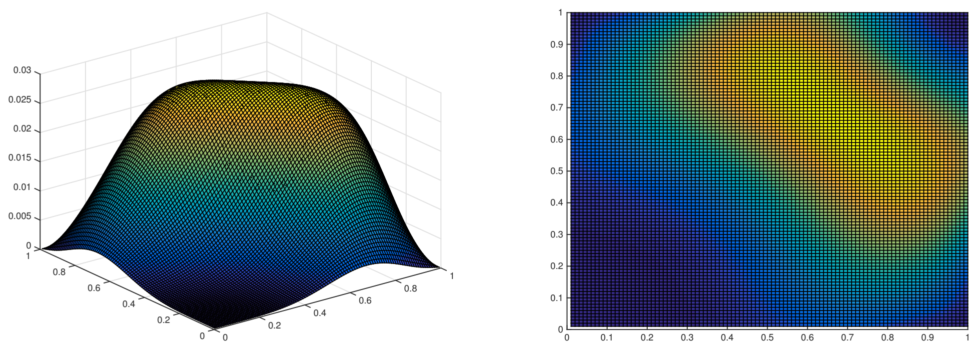

Figure 5: $\mathcal{L}_1$

(d) **(HARD EM)** First, we assign labels to each coin based on the current estimates of θ. We first find a threshold that determines labels by solving the following equation.

$$\left(\frac{2}{3}\right)^x \left(\frac{1}{3}\right)^{3-x} = \left(\frac{1}{3}\right)^x \left(\frac{2}{3}\right)^{3-x} \Rightarrow x = 1.5$$

That is, we label the i-th dice as A if $h_i > 1.5$, and we label it as B otherwise. Thus, in the first E-step of the algorithm, we get the following labels.

$$z_1 = B, z_2 = z_3 = A$$

In the following M-step, we update θ_A, θ_B as follows.

$$\hat{\theta_A} = \frac{\sum_{z_i=A} h_i}{10 \sum_{z_i=A} 1} = \frac{5}{6}$$

$$\hat{\theta_B} = \frac{\sum_{z_i=B} h_i}{10 \sum_{z_i=B} 1} = \frac{1}{3}$$

(SOFT EM) First, we find 'soft' label of each coin using the Bayes' rule. We define p_i as the probability of dice i being type-A. Then,

$$p_1 = \frac{\theta_A^1(1-\theta_A)^2}{\theta_A^1(1-\theta_A)^2 + \theta_B^1(1-\theta_B)^2} = \frac{2}{2+4} = \frac{1}{3}$$
$$p_2 = \frac{\theta_A^3(1-\theta_A)^0}{\theta_A^3(1-\theta_A)^0 + \theta_B^3(1-\theta_B)^0} = \frac{8}{8+1} = \frac{8}{9}$$
$$p_3 = \frac{\theta_A^2(1-\theta_A)^1}{\theta_A^2(1-\theta_A)^1 + \theta_B^2(1-\theta_B)^1} = \frac{4}{4+2} = \frac{2}{3}$$

Then, the following M-step of the soft EM algorithm maximizes the following objective.

$$\sum_z \log(f(y|z,\theta))P(z|y,\theta) \propto \sum_z \log(f(y|z,\theta))P(z|y)$$
$$= \sum_i p_i \log\left(\theta_A^{h_i}(1-\theta_A)^{3-h_i}\right) + (1-p_i)\log\left(\theta_B^{h_i}(1-\theta_B)^{3-h_i}\right)$$

Taking the partial derivatives of the above objective function with respect to θ_A and θ_B give the following update equations.

$$\hat{\theta_A} = \frac{\sum_i p_i \frac{h_i}{3}}{\sum_i p_i} = \frac{\frac{1+8+4}{9}}{\frac{3+8+6}{9}} = \frac{13}{17} \simeq 0.76$$
$$\hat{\theta_B} = \frac{\sum_i (1-p_i)\frac{h_i}{3}}{\sum_i (1-p_i)} = \frac{\frac{2+1+2}{9}}{\frac{6+1+3}{9}} = \frac{1}{2} = 0.5$$

These update equations have an intuitive interpretation: the new estimates are weighted average of $\frac{h_i}{3}$.

(e) Both algorithms do not guarantee convergence to the MLE. (Optional) For this problem, however, the hard EM does not converge to the MLE but the soft EM does. Figure 6 shows how the soft EM algorithm converges from two different initial points: $(\theta_A, \theta_B) = (\frac{2}{3}, \frac{1}{3})$ and $(\theta_A, \theta_B) = (0.99, 0.01)$. For both cases, the algorithm converges to the MLE $(\theta_A, \theta_B) = (\frac{2}{3}, \frac{2}{3})$.

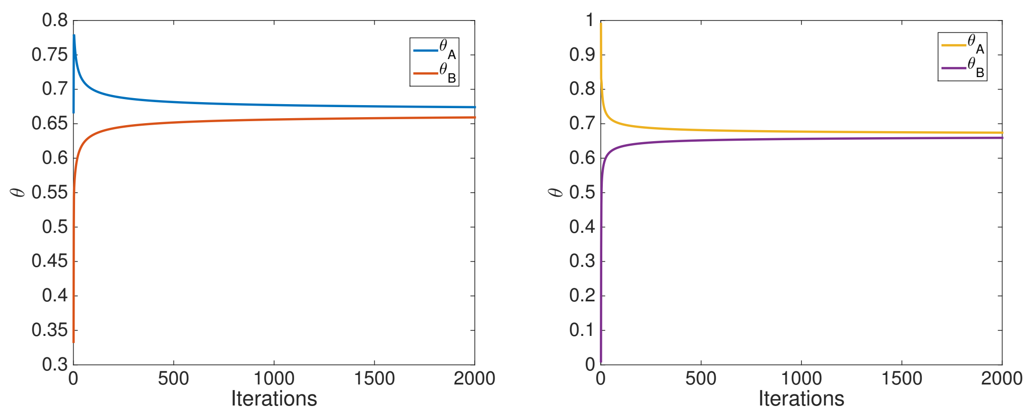

Figure 6: Convergence of the soft EM algorithm