EE 641 Final Exam

December 12, Fall 2022

Name: **Key** ______________________________

Q1: Instructions (4pt)

Rules: I understand that this is an open book exam that shall be done within the allotted time of 180 minutes. I can use my notes, previous posted exams and exam solutions, and web resources. However, I will not communicate with any other person other than the official exam proctors during the exam; I will not seek or accept help from any other persons other than the official proctors; and I will not use GPT-3 or any other variant of an AI response engine.

Signature: ______________________________

Q2: Generating Random Variables (10pt)

Let X be a random variable with the CDF given by

$$F(\lambda) = P\{X \leq \lambda\} \ ,$$

where F is continuous and strictly monotone increasing.

Q2.1:

Give a method for generating a new random variable, X', with the same distribution as X.

Q2.2:

Prove that if $F'(\lambda)$ is the CDF of X', then $F'(\lambda) = F(\lambda)$.

Solution:

Q2.1:

Since F is a CDF, we know that $\lim_{\lambda \to \infty} F(\lambda) = 1$ and $\lim_{\lambda \to -\infty} F(\lambda) = 0$. In addition, since $F : \Re \to [0,1]$ is a continuous and strictly monotone increasing function, we know that there exits an inverse function $F^{-1} : (0,1) \to \Re$ such that $\forall u \in (0,1)$

$$F(F^{-1}(u)) = u \ .$$

So then we generate a random variable U that is uniformly distributed on the interval $(0,1)$, and we can then generate the desired random variable with

$$X' = F^{-1}(U) \ .$$

Q2.2:

We know that

$$\begin{aligned} F'(\lambda) &= P\{X' \leq \lambda\} \\ &= P\{F^{-1}(U) \leq \lambda\} \\ &= P\{F(F^{-1}(U)) \leq F(\lambda)\} \\ &= P\{U \leq F(\lambda)\} \\ &= F(\lambda) \ . \end{aligned}$$

Q3: Properties of Discrete Distribution (25pt)

Consider $X = (X_0, \cdots, X_{N-1})$ where X_n are i.i.d. random variables such that

$$P\{X_n = i\} = \theta_i \ .$$

Also let $p_\theta(x)$ denote the associated family of distributions such that $\theta \in S$ where S denotes the M dimensional simplex given by

$$S = \left\{ \theta \in S : \forall i \in \{0, \cdots, M-1\} \ , \ \theta_i \geq 0 \text{ and } \sum_{i=0}^{M-1} \theta_i = 1 \right\} \ .$$

Q3.1:

Show that

$$K_i = \sum_{n=0}^{N-1} \delta(X_n - i) \ ,$$

is a sufficient statistic for the family of distributions $p_\theta(x)$.

Q3.2:

Show that $p_\theta(x)$ is an exponential distribution with natural sufficient statistics of $\{K_i\}_{i=0}^{M-1}$.

Q3.3:

Derive the maximum likelihood estimate of θ given the observations X.

Solution:

Q3.1:

In order to show that K_i is a sufficient statistic, we need to show that the density can be put in the form of equation (12.21) given by

$$p_\theta(x) = g(K, x) \ .$$

Manipulating the expressions results in the form

$$\begin{aligned} p_\theta(x) &= \prod_{m=0}^{M-1} \theta_i^{K_i} \\ &= \exp\left\{ \sum_{m=0}^{M-1} K_i \log(\theta_i) \right\} \ . \end{aligned}$$

So therefore, K is a sufficient statistic for the family of distributions.

Q3.2:

In order to show that $p_\theta(x)$ is an exponential family of distributions with natural sufficient statistic K, we must show that it has the form of equation (12.25).

$$p_\theta(x) = \exp\left\{ \langle \eta(\theta), K \rangle + d(\theta) + s(x) \right\} \ .$$

To show this, we take

$$\begin{aligned} [\eta(\theta)]_i &= \log(\theta_i) \\ d(\theta) &= 0 \\ s(x) &= 0 \ . \end{aligned}$$

Q3.3:

In order to compute the maximum likelihood estimate, we need to minimize the negative log likelihood subject to the constraint that $1 = \sum_{i=0}^{M-1} \theta_i$. We can do this using the Lagrange multiplier, λ, with

$$\begin{aligned} \nabla_{\theta_m} \left\{ \sum_{i=0}^{M-1} K_i \log(\theta_i) + \lambda \sum_{i=0}^{M-1} \theta_i \right\} \Bigg|_{\theta = \hat{\theta}} &= 0 \\ \frac{K_i}{\hat{\theta}_m} + \lambda &= 0 \\ \hat{\theta}_m &= \frac{K_i}{-\lambda} \ . \end{aligned}$$

We can then solve for λ by

$$1 = \sum_{i=0}^{M-1} \theta_i = \sum_{i=0}^{M-1} \frac{K_i}{-\lambda} = \frac{1}{-\lambda} \sum_{i=0}^{M-1} K_i = \frac{1}{-\lambda} N \ .$$

So we have that $\lambda = -1/N$ and

$$\hat{\theta}_m = \frac{K_i}{-\lambda} = \frac{K_i}{-\lambda} = \frac{K_i}{N} \ .$$

Q4: ADMM Optimization (20pt)

Let $X = (X_0, \cdots, X_{N-1})$ be i.i.d. random variables with distribution

$$P\{X_n = m\} = \pi_m \ ,$$

where $\pi_i \geq 0$ and $\sum_{m=0}^{M-1} \pi_m = 1$.

Also, let $Y = (Y_0, \cdots, Y_{N-1})$ be conditionally independent discrete random variables given X with each Y_n having the conditional distribution given by

$$P\{Y_n = j | X_n = i\} = P_{i,j}$$

where $P_{i,j} \geq 0$ and $\sum_{j=0}^{M-1} P_{i,j} = 1$.

Furthermore, let $\theta = (\pi_0, P_{0,0}, \ldots, P_{0,M-1}, \cdots, \pi_{M-1}, P_{M-1,0}, \ldots, P_{M-1,M-1})$ parameterize the model.

Q4.1:

Using the statistic,

$$N_i = \sum_{n=0}^{N-1} \delta(X_n - i) \ ,$$

write out an expression for the density of X.

Q4.2:

Using the statistic,

$$K_{i,j} = \sum_{n=0}^{N-1} \delta(X_n - i)\delta(Y_n - j) \ ,$$

write out an expression for the conditional density of Y given X.

Q4.3:

Write out the negative log likelihood, $-\log p_\theta(x, y)$, in terms of the sufficient statistics N_i and $K_{i,j}$.

Q4.4:

Write out the maximum likelihood estimate of θ given the complete data, (X, Y).

Q4.5:

Write out the explicit expression for the E-step of the EM algorithm for estimating θ given Y.

Q4.6:

Write out the explicit expression for the M-step of the EM algorithm for estimating θ given Y.

Solution:

Q4.1:

$$p_\theta(x) = \exp\left\{ \sum_{i=0}^{M-1} N_i \log(\pi_i) \right\}$$

Q4.2:

$$p_\theta(y|x) = \exp\left\{ \sum_{i=0}^{M-1} \sum_{j=0}^{M-1} K_{i,j} \log(P_{i,j}) \right\}$$

Q4.3:

$$\begin{aligned} -\log p_\theta(x,y) &= -\log p_\theta(y|x) - \log p_\theta(x) \\ &= -\sum_{i=0}^{M-1} \sum_{j=0}^{M-1} K_{i,j} \log(P_{i,j}) - \sum_{i=0}^{M-1} N_i \log(\pi_i) \end{aligned}$$

Q4.4:

$$\begin{aligned} \hat{\pi}_i &= \frac{N_i}{N} \\ \hat{P}_{i,j} &= \frac{K_{i,j}}{N_i} \end{aligned}$$

Q4.5:

We first need to compute

$$f_n(i|\theta') = P\{X_n = i | Y_n = y_n\} = \frac{\pi'_i P'_{i,y_n}}{\sum_{m=0}^{M-1} \pi'_m P'_{m,y_n}}$$

Then we compute the expected sufficient statistics

$$\begin{aligned} \bar{N}_i &\leftarrow E_{\theta'}\left[\sum_{n=0}^{N-1} \delta(X_n - i) | y_n \right] = \sum_{n=0}^{N-1} E_{\theta'}\left[\delta(X_n - i)|y_n\right] = \sum_{n=0}^{N-1} f_n(i|\theta') \\ \bar{K}_{i,j} &\leftarrow E_{\theta'}\left[\sum_{n=0}^{N-1} \delta(X_n - i)\delta(Y_n - j) | y_n \right] = \sum_{n=0}^{N-1} E_{\theta'}\left[\delta(X_n - i)|y_n\right] \delta(y_n - j) = \sum_{n=0}^{N-1} \delta(y_n - j) f_n(i|\theta') \end{aligned}$$

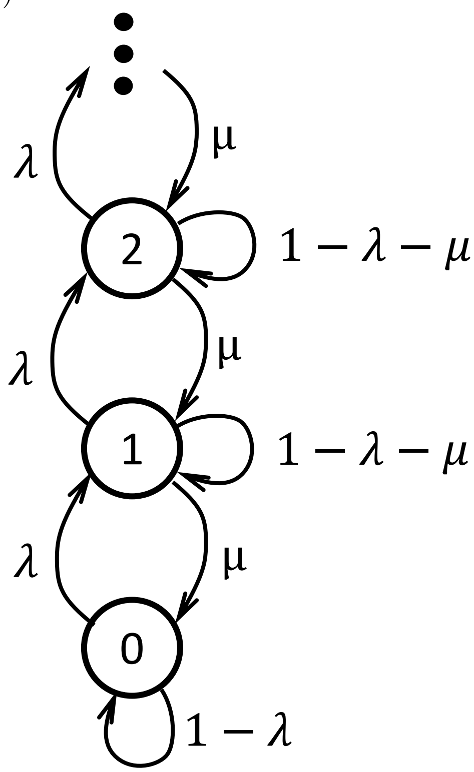
Q4.6:

The M-step is then given by

$$\hat{\pi}_i \leftarrow \frac{\bar{N}_i}{N}$$

$$\hat{P}_{i,j} \leftarrow \frac{\bar{K}_{i,j}}{N_i}$$

Q5: Markov Chain (25pt)



Let $\{X_n\}_{n=0}^{\infty}$ be a homogeneous Markov chain with states $\{0, 1, 2, \cdots\}$ and state-transition diagram as shown above. Furthermore, assume that $\rho = \lambda/\mu < 1$.

Q5.1:

Write out an explicit form for the state transition probabilities, $P_{i,j}$.

Q5.2:

Is there a solution to the detailed balance equations for this Markov chain? If so, give the solution.

Q5.3:

Is there a solution to the full balance equations for this Markov chain? If so, give the solution.

Q5.4:

Is the Markov chain reversible? Justify your answer.

Q5.5:

Determine the stationary distribution for the Markov chain.

Solution:

Q5.1:

$$P_{i,j} = \begin{cases} \lambda & \text{if } j = i+1 \\ \mu & \text{if } j = i-1 \text{ and } i > 0 \\ 1-\lambda-\mu & \text{if } j = i \text{ and } i > 0 \\ 1-\lambda & \text{if } j = i = 0 \\ 0 & \text{otherwise} \end{cases}$$

Q5.2:

Since this is a birth-death process, it must be reversible, and therefore, there must be a solution to the DBE. Detailed balance equations are given by

$$\pi_i P_{i,j} = \pi_j P_{j,i} \ .$$

If we take $j = i + 1$, then we have that

$$\begin{aligned} \pi_i P_{i,i+1} &= \pi_{i+1} P_{i+1,i} \\ \pi_i \lambda &= \pi_{i+1} \mu \\ \pi_{i+1} &= \frac{\lambda}{\mu} \pi_i \end{aligned}$$

So therefore we have that

$$\pi_i = \left(\frac{\lambda}{\mu} \right)^i \pi_0 \ .$$

If we define $\rho = \frac{\lambda}{\mu}$, and using the fact that $1 = \sum_{i=0}^{\infty} \pi_i$, we have that

$$1 = \sum_{i=0}^{\infty} \rho^i \pi_0 = \pi_0 \sum_{i=0}^{\infty} \rho^i = \pi_0 \frac{1}{1-\rho} \ .$$

So we have that $\pi_0 = 1 - \rho$, which results in

$$\pi_i = \rho^i (1 - \rho) \ .$$

Q5.3:

Any solution to the DBE must also be a solution to the FBE because

$$\sum_{i=0}^{M-1} \pi_i P_{i,j} = \sum_{i=0}^{M-1} \pi_j P_{j,i} = \pi_j \sum_{i=0}^{M-1} P_{j,i} = \pi_j \ .$$

So therefore, the solution to the FBE is also

$$\pi_i = \rho^i (1 - \rho) \ .$$

Q5.4:

Yes, the Markov chain is reversible because it is a birth-death process and also because there is a solution to the DBE.

Q5.5:

The stationary distribution is given by

$$\pi_i = \rho^i(1-\rho) \ .$$

Q6: Plug-and-Play Methods (25pt)

Define

$$f(x) = \frac{1}{2\sigma_y^2}\|y - Ax\|^2$$

and its associated proximal map as

$$\hat{x} = F(z) = \arg\min_x \left\{ f(x) + \frac{1}{2\sigma^2}\|x - z\|^2 \right\} .$$

Let $H(z)$ be a firmly non-expansive function so that

$$X \approx H(X + W) ,$$

where X is a typical image and $W \sim N(0, \sigma^2 I)$. Then define

$$T = (2H - I)(2F - I) .$$

Furthermore, assume that the fixed point problem $Tw^* = w^*$ has a unique solution denoted by w^*.

Q6.1:

Use the results of Appendix B, Properties B.5, B.3, and B.1 to prove that T is non-expansive.

Q6.2:

Give an algorithm for computing the solution to the fixed point problem $Tw^* = w^*$. Why do you know that this algorithm converges to w^*.

Q6.3:

Prove that there is a solution to the equilibrium equation

$$F(x^* - u^*) = x^*$$

$$H(x^* + u^*) = x^* .$$

(Hint: Reverse the argument of Section 10.3.3 page 161.)

Q6.4:

In the equilibrium equations,

$$F(x^* - u^*) = x^*$$

$$H(x^* + u^*) = x^* ,$$

give an interpretation for the quantities x^* and u^*.

Q6.5:

Explain how one might obtain an agent, $H(z)$?

Q6.6:

What is the advantage of this approach over more conventional MAP estimation?

Solution:

Q6.1:

By B.5 we know that since F is a proximal map, F must be firmly non-expansive.

By B.3, we know that since F is firmly non-expansive, then $(2F - I)$must be non-expansive.

By B.3 and the fact that H is assumed to be firmly non-expansive, we know that $(2H - I)$ must be firmly non-expansive.

By B.2 and the fact that $(2F - I)$ and $(2H - I)$ are firmly non-expansive, then both $(2F - I)$ and $(2H - I)$ must be non-expansive.

By B.1 and the fact that $(2F - I)$ and $(2H - I)$ are non-expansive, then we know that $T = (2H - I)(2F - I)$ must be non-expansive.

Q6.2:

For $\rho \in (0, 1)$, the Mann algorithm given by

initialize w
Repeat {
  $w \leftarrow (1 - \rho)w + \rho T w$
} .

It must converge to a fixed point w^* because T is non-expansive and a fixed point exists.

Q6.3:

Let w_1^* be a fixed point so that $Tw_1^* = w_1^*$. Then we know that

$$(2H - I)(2F - I)w_1^* = w_1^*$$

So if we define $w_2^* = (2F - I)w_1^*$, then we have that

$$\begin{aligned} (2F - I)w_1^* &= w_2^* \\ (2H - I)w_2^* &= w_1^* \ . \end{aligned}$$

From this we have that

$$\begin{aligned} Fw_1^* &= \frac{w_2^* + w_1^*}{2} \\ Hw_2^* &= \frac{w_2^* + w_1^*}{2} \ . \end{aligned}$$

If we define the following transformed variables as

$$\begin{aligned} x^* &= \frac{w_1^* + w_2^*}{2} \\ u^* &= \frac{w_1^* - w_2^*}{2} \ , \end{aligned}$$

then we have that

$$\begin{aligned} w_1^* &= x^* + u^* \\ w_2^* &= x^* - u^* \ , \end{aligned}$$

so therefore we have that

$$\begin{aligned} F(x^* + u^*) &= x^* \\ H(x^* + u^*) &= x^* \ . \end{aligned}$$

Q6.4:
The quantity x^* is the solution to the inverse problem $y = Ax + W$. The quantity u^* has the interpretation of noise that is removed by the operator H.

Q6.5:
One can obtain a agent H by generating training pairs $(X^{(k)}, Z^{(k)})$ where

$$Z^{(k)} = X^{(k)} + W^{(k)} \ ,$$

where $X^{(k)}$ is a typical image that is expected in the application and $W^{(k)}$ is with noise with variance σ^2.
Then the denoising agent $H_\theta(z)$ can be trained to minimize the loss function given by

$$L(\theta) = \sum_k \|Z^{(k)} - H_\theta(X^{(k)})\|^2 \ .$$

Q6.6:
The method is more general than conventional MAP estimation because H does not need to be a proximal map. This also allows for the use for more advanced agents such as deep neural networks.