HW5 – 209404409

(Q1)

1. Given $\theta > 0$ the PDF of a single random variable, X_i is, $f(X_i = x_i)$ depends on whether $x_i \in [0, \theta]$.
   If it is, then according to continues uniform distribution, $f(X_i = x_i) = \frac{1}{\theta - 0}$, otherwise it could not be sampled, and therefore $f(X_i = x_i) = 0$.
   $$f(X_i = x_i) = \begin{cases} \frac{1}{\theta - 0}, x_i \in [0, \theta] \\ 0 \quad\quad ,else \end{cases} = \begin{cases} \theta^{-1}, x_i \in [0, \theta] \\ 0 \quad\quad ,else \end{cases}$$
2. We are given that all the random variables are independent, and therefore,
   $L(x_1, \dots, x_{10}; \theta) \overset{likelihood\ definition}{=}$
   $\Pr(X_1 = x_1, \dots, X_{10} = x_{10}; \theta) \overset{independent}{=}$
   $\Pr(X_1 = x_1; \theta) \dots \Pr(X_{10} = x_{10}; \theta) = \prod_{i=1}^{10} \Pr(X_i = x_i; \theta)$
   - If $\forall i \in [10], x_i \in [0, \theta]$, then according to 1.1, $\Pr(X_i = x_i; \theta) = \theta^{-1}$, and therefore, $\prod_{i=1}^{10} \Pr(X_i = x_i; \theta) = \theta^{-10}$
   - Otherwise, $\exists i \in [10], x_i \notin [0, \theta]$, and according to 1.1, $\Pr(X_i = x_i; \theta) = 0$, and therefore, $\prod_{i=1}^{10} \Pr(X_i = x_i; \theta) = 0$

   That is, $L(x_1, \dots, x_{10}; \theta) = \prod_{i=1}^{10} \Pr(X_i = x_i; \theta) \overset{1.1}{=} \begin{cases} \frac{1}{\theta^{10}}, \forall i \in [10], x_i \in [0, \theta] \\ 0 \quad\quad\quad\quad ,else \end{cases}$

3. We will split by the value of θ,
   - $\forall i \in [10], x_i \in [0, \theta]$
     $\frac{dL(x_1, \dots, x_{10}; \theta)}{d\theta} \overset{1.2}{=} \frac{d}{d\theta} \theta^{-10} = -10\theta^{-11}$
   - Otherwise,
   - $\frac{dL(x_1, \dots, x_{10}; \theta)}{d\theta} \overset{1.2}{=} \frac{d}{d\theta} 0 = 0$

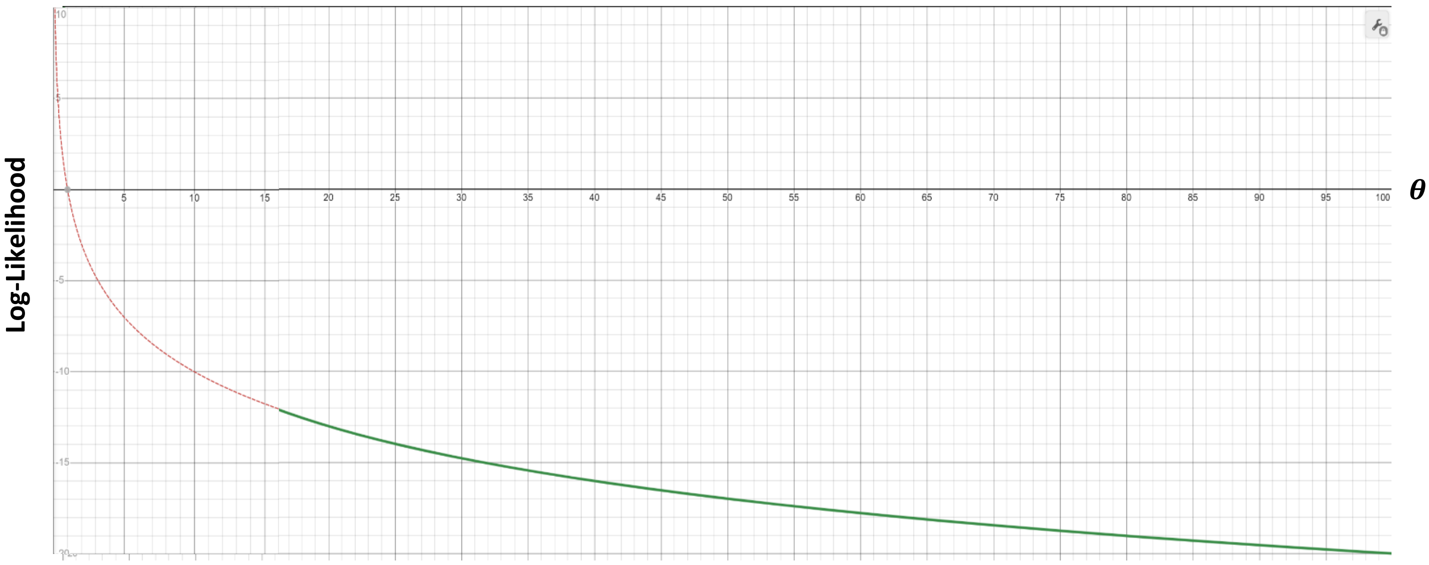
   That is, $\frac{dL(x_1, \dots, x_{10}; \theta)}{d\theta} = \begin{cases} -10\theta^{-11}, \forall i \in [10], x_i \in [0, \theta] \\ 0 \quad\quad\quad\quad ,else \end{cases}$

   As we can see, $\frac{dL(x_1, \dots, x_{10}; \theta)}{d\theta} = 0$ if $\exists i \in [10]. x_i \notin [0, \theta]$, or equally, $0 < \theta < \max\{x_i\}_{i=1}^{10}$, and for any $\theta \geq \max\{x_i\}_{i=1}^{10}$, it holds that $\forall i \in [10], x_i \in [0, \theta]$, and therefore by definition, $\frac{dL(x_1, \dots, x_{10}; \theta)}{d\theta} = -10\theta^{-11} < 0$.

   Note that It is given that $\theta > 0$ and therefore $\theta^{-11} > 0$, and therefore $-10\theta^{-11} < 0$

   Therefore, $\frac{dL(x_1, \dots, x_{10}; \theta)}{d\theta} = 0 \leftrightarrow 0 < \theta < \max\{x_i\}_{i=1}^{10}$

4. The log-likelihood will be undefined for $0 < \theta < \max\{x_i\}_{i=1}^{10}$, because the likelihood at these points is 0 and therefore the log function in undefined. Therefore,
$\log L(x_1, \dots, x_{10}; \theta) =$
$\log \prod_{i=1}^{10} \Pr(X_i = x_i; \theta) \overset{1.2}{=} \begin{cases} \log \theta^{-10} & , \forall i \in [10], x_i \in [0, \theta] \\ undefined & , else \end{cases} =$
$\begin{cases} -10 \log \theta & , \forall i \in [10], x_i \in [0, \theta] \\ undefined & , else \end{cases},$
following is a graph of the log-likelihood as a function of θ,



The red and green graphs, together, form the function $\log \frac{1}{\theta^{10}}$.
The green part shows also the values of the log-likelihood function for any $\theta \geq \max\{x_i\}_{i=1}^{10}$, and the red part is the region of the $\log \frac{1}{\theta^{10}}$ function which is **not defined** for the log-likelihood function, due to the reason described above. The meeting point of these two parts of the $\log \frac{1}{\theta^{10}}$ function, is the hypothetical value of $\max\{x_i\}_{i=1}^{10}$, that means, the meeting point may move left or right, dependent on the value of $\max\{x_i\}_{i=1}^{10}$.

5. $\hat{\theta} = \arg\max_{\theta} L(x_1, \ldots, x_{10}; \theta) = \begin{cases} \arg\max_{\theta} \frac{1}{\theta^{10}}, \forall i \in [10], x_i \in [0, \theta] \\ \arg\max_{\theta} 0 \qquad\qquad , else \end{cases}$

Lets show that $\hat{\theta} = \max\{x_i\}_{i=1}^{10}$

We will split into two cases,

- $\exists\, i \in [10], x_i \neq 0$

  Assume, by the way of contradiction that $\hat{\theta} \neq \max\{x_i\}_{i=1}^{10}$. Let's split into two cases,
  - $\hat{\theta} > \max\{x_i\}_{i=1}^{10}$,

    $\forall i \in [10], \hat{\theta} > \max\{x_i\}_{i=1}^{10} \geq x_i$ and therefore $\frac{1}{x_i} > \frac{1}{\hat{\theta}}$, and therefore also, $\frac{1}{\max\{x_i\}_{i=1}^{10}} > \frac{1}{\hat{\theta}}$, so we can chooses $\max\{x_i\}_{i=1}^{10} = \theta' \neq \hat{\theta}$, such that, $\frac{1}{\left(\max\{x_i\}_{i=1}^{10}\right)^{10}} = L(x_1, \ldots, x_{10}; \theta') > \frac{1}{\hat{\theta}^{10}} = L\left(x_1, \ldots, x_{10}; \hat{\theta}\right)$ which contradicts the optimality of $\hat{\theta}$.
  - $\hat{\theta} < \max\{x_i\}_{i=1}^{10}$

    $$L\left(x_1, \ldots, x_{10}; \hat{\theta}\right) \overset{\max\{x_i\}_{i=1}^{10} > \hat{\theta}}{=} 0 < \frac{1}{\max\{x_i\}_{i=1}^{10}} = L\left(x_1, \ldots, x_{10}; \max\{x_i\}_{i=1}^{10}\right)$$

    Which contradicts the optimality of $\hat{\theta}$.

    Note that $\max\{x_i\}_{i=1}^{10} \neq 0$ because there exists at least one $0 \leq x_i \neq 0$.

  At both cases we got a contradiction, and therefore it holds that $\hat{\theta} = \max\{x_i\}_{i=1}^{10}$

- $\forall i \in [10], x_i = 0$,

  It is given that $\theta > 0$, and indeed, as θ goes closer to 0^+ the log-likelihood gets bigger, that is,

  $$\lim_{\theta \to 0^+} Loglikelihood(\theta) =$$

  $$\lim_{\theta \to \left(\max\{x_i\}_{i=1}^{10}\right)^+} Loglikelihood(\theta) \overset{\theta > \max\{x_i\}_{i=1}^{10}}{=} \lim_{\theta \to \left(\max\{x_i\}_{i=1}^{10}\right)^+} \frac{1}{\theta^{10}} = \infty$$

  Therefore there is no maximum value for the log-likelihood, but it approaches ∞ as at gets closer to0^+.

Therefore, in both cases we get that the maximum value of the log-likelihood received for $\hat{\theta} = \max\{x_i\}_{i=1}^{10}$.

This could also be seen visually in the plotted graph, as we can see that (the plotted graph assumes there is at list one sample which is not equal to 0) the green like, which indicates the actual values of the log-likelihood function, receive its maximum value exactly at the point it meets the red line, that is the value of $\max\{x_i\}_{i=1}^{10}$.

(Q2)

1. To find the requested $c \in \mathbb{R}_{>0}$ we will first note the following,
   $\alpha\theta = \alpha\left(W^{(1)}, W^{(2)}, \dots, W^{(L-1)}, W^{(L)}\right) = \left(\alpha W^{(1)}, \alpha W^{(2)}, \dots, \alpha W^{(L-1)}, \alpha W^{(L)}\right)$, and we will denote $h^{(1)}_{new}(x) = \sigma\left(\alpha {W^{(1)}}^T \cdot x\right), h^{(l)}_{new}(x) = \sigma\left(\alpha {W^{(l)}}^T h^{(l-1)}_{new}(x)\right)$
   Therefore,
   $$F_{\alpha\theta}(x) \overset{1}{=} \alpha {W^{(L)}}^T h^{(l-1)}_{new}(x) \overset{1}{=} \alpha {W^{(L)}}^T \sigma\left(\alpha {W^{(l-1)}}^T h^{(l-2)}_{new}(x)\right) =$$
   $$\overset{1}{=} \alpha {W^{(L)}}^T \sigma\left(\alpha {W^{(l-1)}}^T \sigma\left(\alpha {W^{(l-2)}}^T h^{(l-3)}_{new}(x)\right)\right) \overset{1}{=} \dots \overset{1}{=}$$
   $$\overset{1}{=} \alpha {W^{(L)}}^T \sigma\left(\alpha {W^{(l-1)}}^T \sigma\left(\alpha {W^{(l-2)}}^T \left(\dots \sigma\left(\alpha {W^{2}}^T h^{(1)}_{new}(x)\right) \dots\right)\right)\right) =$$
   $$\overset{1}{=} \alpha {W^{(L)}}^T \sigma\left(\alpha {W^{(l-1)}}^T \sigma\left(\alpha {W^{(l-2)}}^T \left(\dots \sigma\left(\alpha {W^{2}}^T \sigma\left(\alpha {W^{(1)}}^T \cdot x\right)\right) \dots\right)\right)\right) =$$
   $$\overset{2}{=} \alpha {W^{(L)}}^T \sigma\left(\alpha {W^{(l-1)}}^T \sigma\left(\alpha {W^{(l-2)}}^T \left(\dots \sigma\left(\alpha^2 {W^{2}}^T \sigma\left({W^{(1)}}^T \cdot x\right)\right) \dots\right)\right)\right) \overset{2}{=} \dots \overset{2}{=}$$
   $$\overset{2}{=} \alpha {W^{(L)}}^T \sigma\left(\alpha {W^{(l-1)}}^T \sigma\left(\alpha^{L-2} {W^{(l-2)}}^T \left(\dots \sigma\left({W^{2}}^T \sigma\left({W^{(1)}}^T \cdot x\right)\right) \dots\right)\right)\right) =$$
   $$\overset{2}{=} \alpha {W^{(L)}}^T \sigma\left(\alpha^{L-1} {W^{(l-1)}}^T \sigma\left({W^{(l-2)}}^T \left(\dots \sigma\left({W^{2}}^T \sigma\left({W^{(1)}}^T \cdot x\right)\right) \dots\right)\right)\right) =$$
   $$\overset{2}{=} \alpha^L {W^{(L)}}^T \sigma\left({W^{(l-1)}}^T \sigma\left({W^{(l-2)}}^T \left(\dots \sigma\left({W^{2}}^T \sigma\left({W^{(1)}}^T \cdot x\right)\right) \dots\right)\right)\right) =$$
   $$\overset{1}{=} \alpha^L {W^{(L)}}^T \sigma\left({W^{(l-1)}}^T \sigma\left({W^{(l-2)}}^T \left(\dots \sigma\left({W^{2}}^T h^{(1)}(x)\right) \dots\right)\right)\right) =$$
   $$\overset{1}{=} \alpha^L {W^{(L)}}^T \sigma\left({W^{(l-1)}}^T \sigma\left({W^{(l-2)}}^T \left(\dots \left({W^{3}}^T h^{(2)}(x)\right) \dots\right)\right)\right) \overset{1}{=} \dots \overset{1}{=}$$
   $$\overset{1}{=} \alpha^L {W^{(L)}}^T \sigma\left({W^{(l-1)}}^T h^{(l-2)}(x)\right) \overset{1}{=} \alpha^L {W^{(L)}}^T h^{(l-1)}(x) \overset{1}{=} \alpha^L F_\theta(x)$$
   1 – definition
   2 – positive-homogeneous property of the ReLU function
   And therefore, the appropriate $c \in \mathbb{R}_{>0}$ is $c = \alpha^L \in \mathbb{R}_{>0}$.
2. We need to calculate $\lim_{\alpha \to 0} \frac{1}{1+\exp\{-F_{\alpha\theta}(x)\}}$.
   We will calculate the denominator separately,
   $$1 + \exp\{-F_{\alpha\theta}(x)\} \overset{2.1}{=} 1 + \exp\{-\alpha^L F_\theta(x)\}$$
   $$= 1 + (\exp\{-F_\theta(x)\})^{\alpha^L} \xrightarrow[\alpha \to 0]{} 1 + (\exp\{-F_\theta(x)\})^0 = 2$$
   When the last transition is because the function is continuous (note that $-F_\theta(x)$ is a constant in respect to α).
   Therefore, because $\frac{1}{1+\exp\{-F_{\alpha\theta}(x)\}}$ is also continues,
   $$\lim_{\alpha \to 0} \frac{1}{1 + \exp\{-F_{\alpha\theta}(x)\}} = \frac{1}{2}$$