Data Science Homework 1

R11922066 資工碩一林庭安

Problem 1 Basic Definitions

(1)

- The one pre-condition that Σ must satisfy:
  Let X be a set and Σ a σ-algebra over X
- The 3 conditions that μ must satisfy:
  (1) Non-negativity: $\mu(E) \geq 0, \forall E \in \Sigma$
  (2) Null empty set: $\mu(\emptyset) = 0$
  (3) Countable additivity (or σ-additivity): For all countable collections $\{E_k\}_{k=1}^{\infty}$ of pairwise disjoint sets in Σ, $\mu(\sqcup_{k=1}^{\infty} E_k) = \Sigma_{k=1}^{\infty} \mu(E_k)$

(2)

- $d(x, y) \geq 0$ (non-negativity)
- $d(x, y) = 0$ if and only if $x = y$ (identity of indiscernibles)
- $d(x, y) = d(y, x)$ (symmetry)
- $d(x, z) \leq d(x, y) + d(y, z)$ (subadditivity / triangle inequality)

Problem 2 Random Variable Transformation

(1)

Let $F_Z(z)$ be the cumulative distribution function of Z. Then

$$
\begin{aligned}
F_Z(z) &= P(Z \leq z) \\
&= P(max(X, Y) \leq z) \\
&= P(X \leq z)P(Y \leq z) \\
&= F_X(z)F_Y(z)
\end{aligned}
$$

Thus, the probability density function of Z is

$$
\begin{aligned}
f_Z(z) &= \frac{d}{dz}F_Z(z) \\
&= \frac{d}{dz}(F_X(z)F_Y(z)) \\
&= f_X(z)F_Y(z) + F_X(z)f_Y(z) \\
&= e^{-z}\int_0^z e^{-y}dy + e^{-z}\int_0^z e^{-x}dx \\
&= e^{-z}(-e^{-z}+1) + e^{-z}(-e^{-z}+1) \\
&= 2(-e^{-2z}+e^{-z})
\end{aligned}
$$

Hence, $f_Z(z) = \begin{cases} 2(-e^{-2z}+e^{-z}) & \text{for z >0} \\ 0 & \text{elsewhere} \end{cases}$

(2)

Let $F_W(w)$ be the cumulative distribution function of W. Then

$$
\begin{aligned}
F_W(w) &= P(W \leq w) \\
&= P(min(X,Y) \leq w) \\
&= 1 - P(min(X,Y) > w) \\
&= 1 - P(X > w)P(Y > w) \\
&= 1 - (1 - F_X(w))(1 - F_Y(w))
\end{aligned}
$$

Thus, the probability density function of W is

$$
\begin{aligned}
f_W(w) &= \frac{d}{dw}F_W(w) \\
&= \frac{d}{dw}(1 - (1 - F_X(w))(1 - F_Y(w))) \\
&= -\frac{d}{dw}((1 - F_X(w))(1 - F_Y(w))) \\
&= -(-f_X(w)(1 - F_Y(w)) + (1 - F_X(w))(-f_Y(w))) \\
&= f_X(w)(1 - F_Y(w)) + f_Y(w)(1 - F_X(w)) \\
&= e^{-w}(1 - \int_0^w e^{-y}dy) + e^{-w}(1 - \int_0^w e^{-x}dx) \\
&= e^{-w}(1 - (-e^{-w}+1)) + e^{-w}(1 - (-e^{-w}+1)) \\
&= 2e^{-2w}
\end{aligned}
$$

Hence, $f_W(w) = \begin{cases} 2e^{-2w} & \text{for w >0} \\ 0 & \text{elsewhere} \end{cases}$

Problem 3 Random Variable Transformation

(1)

$\because X \sim \Gamma(\theta = 1, \alpha = 1)$

$\therefore$ The cumulative distribution function of X is

$$F_X(x) = \int_0^x \frac{e^{-u}}{\Gamma(1)} du$$

The cumulative distribution function of 2X is

$$\begin{aligned} F_{2X}(x) &= P(2X \leq x) \\ &= P(X \leq x/2) \\ &= F_X(x/2) \\ &= \int_0^{x/2} \frac{e^{-u}}{\Gamma(1)} du \end{aligned}$$

Now substitute z=2u. Then the integral is equal to

$$\int_0^x \frac{e^{-z/2}}{\Gamma(1)} \frac{dz}{2}$$

$\implies 2X \sim \Gamma(\theta = 2, \alpha = 1)$

(2)

The probability density function of 2X is

$$f_{2X}(x) = \frac{e^{-x/2}}{2\Gamma(1)} \quad = \frac{x^{(1-1)}e^{-x/2}}{2^1\Gamma(1)} \implies 2X \sim \Gamma(\theta = 2, \alpha = 1)$$

$$f_{2X}(x) = \frac{e^{-x/2}}{2\Gamma(1)} \quad = \frac{x^{(2/2-1)}e^{-x/2}}{2^{2/2}\Gamma(2/2)} \implies 2X \sim \chi^2(k = 2)$$

Problem 4 Statistical Distance

(1)

$d(x, y) = (x - y)^2 = \|x - y\|^2 = \langle x - y, x - y\rangle = \|x\|^2 - \|y\|^2 - \langle 2y, x - y\rangle$

The F in this Bregman divergence is $F(x) = \|x\|^2$

(2)

$\because$ Entropy has the property:

$H(X, Y) = H(X \mid Y) + H(Y) = H(Y \mid X) + H(X)$

and $H(f(X) \mid X) = 0$

$$\begin{aligned} \therefore H(x, f(x)) &= H(x) + H(f(x) \mid x) \quad = H(f(x)) + H(x \mid f(x)) \\ &= H(x) + 0 \end{aligned}$$

which means $H(x) = H(f(x)) + H(x \mid f(x))$

so $H(f(x)) \leq H(x)$

$\implies$ The entropy of the output of this neural network will always be equal or less that of the input.

Problem 5 Point Estimation

(1)

Calculate the first population moment, we get $E(X) = \dfrac{\theta}{2}$

Using the moment method, we set $\dfrac{\theta}{2} = E(X) = M1 = \dfrac{1}{n}\sum_{i=1}^{n} X_i = \overline{X}$

Therefore, the estimator using the moment method is $\hat{\theta} = 2\overline{X}$

(2)

The density function of the uniform distribution over the interval $(0, \theta)$ is

$$f(x;\theta) = \begin{cases} \frac{1}{\theta} & \text{if } 0 \le x \le \theta \\ 0 & \text{elsewhere} \end{cases}$$

The likelihood function is given by

$$L(\theta) = \begin{cases} \prod_{i=1}^{n} f(x_i;\theta) = \theta^{-n} & \text{if } 0 \le x_1 \le x_n \le \theta, n \ge 2 \\ 0 & \text{elsewhere} \end{cases}$$

The derivative of the log likelihood function wrt θ gives:

$$\frac{dlnL(\theta)}{d\theta} = -\frac{n}{\theta} < 0$$

So we can say that $L(\theta) = \theta^{-n}$ is a decreasing function for $\theta \ge x_n$

Therefore, the maximum value occurs at the left end point, which is x_n

$\implies \hat{\theta}_{MLE} = x_n$

(3)

The density function of the uniform distribution over the interval $(0, \theta)$ is

$$f(x;\theta) = \begin{cases} \frac{1}{\theta} & \text{if } 0 \le x \le \theta \\ 0 & \text{elsewhere} \end{cases}$$

The prior distribution of θ is

$$g(\theta) = \begin{cases} 1 & \text{if } 0 \le \theta \le 1 \\ 0 & \text{elsewhere} \end{cases}$$

The posterior distribution of θ is

$$f(x \mid \theta)g(\theta) = f(x \mid \theta) = \begin{cases} \prod_{i=1}^{n}(\frac{1}{\theta}) = \theta^{-n} & \text{if } 0 \le x_1 \le x_n \le \theta, n \ge 2 \\ 0 & \text{elsewhere} \end{cases}$$

The derivative of log posterior distribution wrt θ gives:

$$\frac{dlnf(x \mid \theta)g(\theta)}{d\theta} = -\frac{n}{\theta} < 0$$

So we can say that $f(x \mid \theta)g(\theta) = \theta^{-n}$ is a decreasing function for $\theta \ge x_n$

Therefore, the maximum value occurs at the left end point, which is x_n

$$\implies \hat{\theta}_{MAP}(x) = \arg\max_{\theta} f(x \mid \theta) g(\theta)$$
$$= x_n$$

(4)

The joint distribution of x and θ is

$$u(x,\theta) = f(x \mid \theta) h(\theta) = \theta^{-n} \quad , \text{for } 0 \le x_1 \le x_n \le \theta \le 1, n \ge 2$$

The marginal distribution of x is

$$\begin{aligned} g(x) &= \int_{x_n}^{1} u(x,\theta) d\theta \\ &= \int_{x_n}^{1} \theta^{-n} d\theta \\ &= \frac{\theta^{1-n}}{1-n}\bigg|_{x_n}^{1} \\ &= \frac{1}{1-n} - \frac{x_n^{1-n}}{1-n} \\ &= \frac{1-x_n^{1-n}}{1-n} \qquad ,\text{for } 0 \le x_1 \le x_n \le 1, n \ge 2 \end{aligned}$$

The conditional density of θ given x is

$$\begin{aligned} k(\theta \mid x) &= \frac{u(x,\theta)}{g(x)} \\ &= \frac{1}{\theta^n} \frac{1-n}{1-x_n^{1-n}} \qquad ,\text{for } 0 \le x_1 \le x_n \le \theta \le 1, n \ge 2 \end{aligned}$$

$$\begin{aligned} \hat{\theta} &= E[\theta \mid x] \\ &= \int_{x_n}^{1} \theta k(\theta \mid x) d\theta \\ &= \int_{x_n}^{1} \frac{1}{\theta^{n-1}} \frac{1-n}{1-x_n^{1-n}} d\theta \\ &= \frac{1-n}{1-x_n^{1-n}} \int_{x_n}^{1} \theta^{1-n} d\theta \\ &= \frac{1-n}{1-x_n^{1-n}} \left(\frac{1}{2-n} \theta^{2-n} \bigg|_{x_n}^{1} \right) \\ &= \frac{1-n}{1-x_n^{1-n}} \left(\frac{1}{2-n} - \frac{x_n^{2-n}}{2-n} \right) \\ &= \frac{1-n}{1-x_n^{1-n}} \frac{1-x_n^{2-n}}{2-n} \end{aligned}$$