r11921091 電機所 楊冠彥

Problem 1

1-1 P.6 Pre-condition : Let X be a set and Σ a algebra over X.

μ must satisfy

- a. Non-negativity : $\mu(E) \geq 0, \forall E \in \Sigma$
- b. Null empty set : $\mu(\phi) = 0$
- c. Countable additivity : $\forall$ countable collections $\{E_k\}_{k=1}^{\infty}$ or pairwise disjoint sets in Σ,

$$\mu\left(\sqcup_{k=1}^{\infty} E_k\right) = \sum_{k=1}^{\infty} \mu(E_k)$$

1-2 P5 From the description in Problem 1, we know that $f(): X \times X \to R^+$

If $f()$ is considered as a metric of X, then $\forall\ x, y, z$ in X, $f()$ must satisfy the following condition:

a. $f(x, y) \geq 0$ (non-negativity)

b. $f(x, y) = 0$ (identity of indiscernibles)

c. $f(x, y) = f(y, x)$ (symmetry)

d. $f(x, y) \leq f(x, y) + f(y, z)$ (triangle inequality or subactivity)

Problem 2.

(1) Note that $Z \leq z \Leftrightarrow$ both X and Y are $\leq z$

$\because$ independence

$\therefore F_Z(z) = P(Z \leq z) = P(X \leq z) P(Y \leq z)$

$$= \left(\int_0^z e^{-x} dx\right)\left(\int_0^z e^{-y} dy\right)$$

$$= (1-e^{-z})(1-e^{-z})$$

$$= (1-e^{-z})^2$$

This is valid $\forall z > 0$. (If $z \leq 0$, then $F_Z(z) = 0$)

$$\Rightarrow f_Z(z) = \frac{1-2e^{-z}+e^{-2z}}{dz} = 2e^{-z} - 2e^{-2z}$$

$$= 2e^{-z}(1-e^{-z}), \text{ for } z > 0$$

$$\Rightarrow f_Z(z) = \begin{cases} 2e^{-z} - 2e^{-2z}, & \text{if } z > 0 \\ 0, & \text{otherwise.} \end{cases}$$

(2) Note that $W \leq w \Leftrightarrow$ both X and Y are $\leq w$

$$\therefore F_W(w) = P(W \leq w) = P(\min(X,Y) \leq w)$$

$$= 1 - P(W > w)$$

$$= 1 - \left(P(X > w)\, P(Y > w)\right)$$

$$= 1 - [1 - P(X \leq w)][1 - P(Y \leq w)]$$

$$= 1 - \left[1 - \int_0^w e^{-x}\,dx\right]\left[1 - \int_0^w e^{-y}\,dy\right]$$

$$= 1 - [1 - (1 - e^{-w})][1 - (1 - e^{-w})]$$

$$= 1 - e^{-2w}$$

This is valid $\forall\, z > 0$. (If $z \leq 0$, then $F_W(w) = 0$)

$$\Rightarrow f_W(w) = \frac{F_W(w)}{dw} = \frac{1 - e^{-2w}}{dw}$$

$$= 2e^{-2w}, \quad \text{if } w > 0$$

$$\Rightarrow f_W(w) = \begin{cases} 2e^{-2w}, & \text{if } w > 0 \\ 0, & \text{otherwise} \end{cases}$$

Problem 3

- Drive the distribution of $2X$

$$f(x) = \begin{cases} \frac{1}{\Gamma(\alpha)\,\theta^{\alpha}} x^{\alpha-1} e^{-\frac{x}{\theta}} & \text{if } 0 < x < \infty \\ 0 & \text{otherwise} \end{cases}$$

$\theta = 1$, $\alpha = 1$

$$f(x) = \frac{1}{\Gamma(1)} e^{-x} \quad , \text{if } 0 < x < \infty$$

$$= e^{-x} \quad , \text{if } 0 < x < \infty$$

$$\Rightarrow f(x) = \begin{cases} e^{-x} & , \text{if } 0 < x < \infty \\ 0 & , \text{otherwise} \end{cases}$$

When $X' = 2X$

Let $C(x')$ be the cdf of X'

$$C(x') = P(X' \le x') = P(2X \le x')$$

$$= P\left(0 \le X \le \frac{x'}{2}\right)$$

$$= \int_0^{\frac{x'}{2}} e^{-x}\, dx$$

$$= 1 - e^{-\frac{x'}{2}}$$

$$c(x') = \frac{d\,C(x')}{dx'} = \frac{d}{dx'}\left(1 - e^{-\frac{x'}{2}}\right)$$

$$= \left(-e^{-\frac{x'}{2}}\right)\left(-\frac{1}{2}\right)$$

$$= \frac{1}{2} e^{-\frac{x'}{2}}$$

$$\therefore f(x') = \begin{cases} \frac{1}{2} e^{-\frac{x'}{2}} & , \text{ if } 0<X<\infty \\ 0 & , \text{ otherwise} \end{cases}$$

◦ Expressive the distribution of 2X in terms of

(1) Gamma distribution

(2) Chi ^2 distribution

(1) From above answer, we know that

$X'=2X$, $$f(x') = \begin{cases} \frac{1}{2} e^{-\frac{x'}{2}} & , \text{ if } 0<X<\infty \\ 0 & , \text{ otherwise} \end{cases}$$

$$= \begin{cases} \frac{1}{\Gamma(\alpha)\theta^{\alpha}} x^{\alpha-1} e^{-\frac{x}{\theta}} & , \text{ if } 0<X<\infty \\ 0 & , \text{ otherwise} \end{cases}$$

$\Rightarrow \alpha = 1 \quad (\because x^{\alpha-1} = x^{0})$

$\Rightarrow \theta = 2$

(2) $X \sim \chi^2(r) = \text{Gamma}(2, \frac{r}{2}) = \text{Gamma}(2,1)$

$\Rightarrow r=2 \Rightarrow X \sim \chi^2(2)$

$$f(x) = \begin{cases} \frac{1}{\Gamma(\frac{r}{2}) 2^{\frac{r}{2}}} x^{\frac{r}{2}-1} e^{-\frac{x}{2}} dx \\ 0 \end{cases}$$

$$= \begin{cases} \frac{1}{2} e^{-\frac{x}{2}} & , \text{ if } 0<X<\infty \\ 0 & , \text{ otherwise} \end{cases}$$

Problem 4.

a. Bregman Divergence

$$d(x,y) = (x-y)^2 = x^2 - 2xy + y^2$$

$$= x^2 - (y^2 + (2y)(x-y))$$

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

Let $F(x) = x^2 \wedge \in$ stricly convex function

$$d(x,y) = F(x) - (F(y) + \langle \nabla F(y), x-y \rangle)$$

$$= D_F(x,y)$$

$\Rightarrow$ the squared Eudidean distance is a Bregman divergence

QED.

b. Entropy: $H(x) = -\sum_{i=1}^{n} P(x_i) \log P(x_i)$

$$= \sum_{i=1}^{n} P(x_i) \log\left(\frac{1}{P(x_i)}\right)$$

$$H(y) = \sum_{i=1}^{n} P(y_i) \log\left(\frac{1}{P(y_i)}\right)$$

$$= \sum_{i=1}^{n} P(F(x_i)) \log\left(\frac{1}{P(F(x_i))}\right)$$

$$H(X,Y) = H(X|Y) + H(Y) = H(Y|X) + H(X)$$

$$\because H(F(X)|X) = 0$$

$$\Rightarrow H(X, F(X)) = H(X) + H(F(X)|X) = H(F(X)) + H(X|F(X))$$

which means $H(X) = H(F(X)) + H(X|F(X))$

$\therefore H(F(X)) \leq H(X)$ QED

Problem 5.

$X \sim Unif(0, \theta)$ Note that uniform distribution

$\theta \sim Unif(0, 1)$ $f(x) = \frac{1}{b-a}$, $x \in (a,b)$

a. moment method

$$\mu = E(X) = \int_a^b x f(x) dx$$

$$= \int_a^b \frac{x}{b-a} dx$$

$$= \frac{x^2}{2(b-a)}\Big|_a^b = \frac{b^2-a^2}{2(b-a)} = \frac{b+a}{2}$$

Set $\frac{\theta}{2} = \bar{X} \Rightarrow \theta = 2\bar{X}$

$\therefore$ The moment estimator is $\hat{\theta} = 2\bar{X}$

b. MLE method

$$f(x_i|\theta) = \frac{1}{\theta} \quad \text{if } 0 < x_i < \theta$$

$$L(\theta) = f(x_1, \cdots, x_n|\theta)$$

$$= \prod_{i=1}^n f(x_i|\theta) = \frac{1}{\theta^n} \quad \text{if } 0 < x_i < \theta$$

Claim 1. Given $x_1, \cdots, x_n \in \mathbb{R}$, $x_1, \cdots, x_n < k$

$$\Leftrightarrow X_{(n)} := \max_{1 \leq i \leq n} X_i < k$$

Claim 2. Given $x_1, \cdots, x_n \in \mathbb{R}$, $x_1, \cdots, x_n > k$

$$\Leftrightarrow X_{(1)} := \min_{1 \leq i \leq n} X_i > k$$

Let I denote the indicator function,

where $I(x) = \begin{cases} 1 & , \text{if } x \in A \\ 0 & , \text{if } x \notin A \end{cases}$

$$L(\theta) = \frac{1}{\theta^n} \quad , \text{ if } 0 < x_i < \theta$$

$$= \frac{1}{\theta^n} \prod_{i=1}^{n} I(0 < x_i < \theta)$$

$$= \frac{1}{\theta^n} \prod_{i=1}^{n} \left[I(x_i > 0)\, I(x_i < \theta) \right]$$

$$= \frac{1}{\theta^n} \prod_{i=1}^{n} \left[I(x_i > 0) \right] \prod_{j=1}^{n} \left[I(x_j < \theta) \right]$$

$$\prod_{i=1}^{n} \left[I(x_i > 0) \right] = I(x_1 > 0 \cap x_2 > 0 \cap \cdots \cap x_n > 0) = I(x_{(1)} > 0)$$

$$\prod_{j=1}^{n} \left[I(x_i < \theta) \right] = I(x_1 < \theta \cap x_2 < \theta \cap \cdots \cap x_n < \theta) = I(x_{(n)} < \theta)$$

$$\therefore L(\theta) = \frac{1}{\theta^n} I(x_{(1)} > \theta)\, I(x_{(n)} < \theta)$$

if $x_{(1)} \leq 0$, then

$$L(\theta) = 0$$

if $x_{(1)} > 0$,

(a) if $\theta > x_{(n)}$, then

$$L(\theta) = \frac{1}{\theta^n} I(\theta > x_{(n)})$$

(b) if $\theta \leq x_{(n)}$, then

$$L(\theta) = 0 \qquad (\because \text{ indicator function})$$

$\Rightarrow \theta = x_{(n)}$ can obtain maximum

$\therefore$ θ 的 MLE 為 $\hat{\theta} = x_{(n)} = \max\{x_1, \cdots, x_n\}$

c. MAP method

$$\hat{\theta}_{MAP(X)} = \underset{\theta}{\operatorname{argmax}}\ f(\theta|x)$$

$$= \underset{\theta}{\operatorname{argmax}}\ f(x|\theta)\, g(\theta)$$

$\because X \sim Unif(0,\theta) \wedge \theta \sim Unif(0,1)$ with prior distribution

$$f(x|\theta) = \frac{1}{\theta} \qquad 0 < X < \theta$$

$$g(\theta) = 1$$

i.e. we want to find θ to maximize

$$\hat{\theta}_{MAP(X)} = \underset{\theta}{\operatorname{argmax}}\ \frac{1}{\theta^n}$$

$$= \max\{X_1, \cdots, X_n\}$$

d. Bayesian method

Population $\quad f(x|\theta) = \frac{1}{\theta} \quad$ for $0<X<\theta$

Prior $\quad h(\theta) = 1 \quad$ for $0<\theta<1$

Joint dist. $\quad u(X,\theta) = h(\theta)\, f(x|\theta)$

$$= \frac{1}{\theta}, \quad \text{for } 0<X<\theta<1$$

Marginal dist. of x

$$g(x) = \int_x^1 u(x,\theta)\,d\theta = \int_x^1 \frac{1}{\theta}\,d\theta = -\ln x, \text{ for } 0<X<1$$

$$k(\theta|x) = \frac{u(x,\theta)}{g(x)} = -\frac{1}{\theta \ln x}, \quad \text{for } 0<X<\theta<1$$

$$\hat{\theta} = E(\theta|x) = \int_x^1 \theta\, k(\theta|x)\,d\theta = \int_x^1 \theta \frac{-1}{\theta \ln x}\,d\theta$$

$$= -\frac{1}{\ln x}\int_x^1 d\theta$$

$$= \frac{x-1}{\ln x}$$

$$\Rightarrow \hat{\theta} = \frac{x-1}{\ln x}$$