資料科學 HW1 Student ID: P10942A08 Name: 鞠[illegible]浩

Problem 1: Random Number Transformation

(1) What is the probability density function of $Z = \max(X, Y)$?

Ans: $P(Z \le a) = P(X \le a \text{ and } Y \le a)$

$= P(X \le a)\, P(Y \le a)$

$= F_X(a)\, F_Y(a)$

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

$= (1-e^{-a})^2 = 1 - 2e^{-a} + e^{-2a}$

$\therefore$ The pdf of Z is

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

(2) What is the probability density function of $W = \min(X, Y)$?

Ans: $P(W > a) = P(X > a \text{ and } Y > a) = P(X > a)\, P(Y > a)$

$= 1 - P(W \le a)$

$F_W(a) = 1 - P(W > a) = 1 - P(X > a) P(Y > a)$

$= 1 - (1 - P(X \le a))(1 - P(Y \le a))$

$= 1 - (1 - F_X(a)) \cdot (1 - F_Y(a))$

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

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

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

$\therefore$ The pdf of W is

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

Problem 2: Statistical Distances

(1) What are the four conditions that f must satisfy for f() to be considered a metric?

Ans: $f : X \times X \to R^+$ (R^+: the set of non-negative real numbers) such that for all x, y, z in X, f satisfies the following conditions:

1. $f(x,y) \ge 0$ (non-negativity)
2. $f(x,y) = 0$ if and only if $x = y$ (identity of indiscernibles)
3. $f(x,y) = f(y,x)$ (symmetry)
4. $f(x,z) \le f(x,y) + f(y,z)$ (subadditivity / triangle inequality)

(2) Prove that the squared Euclidean distance is a Bregman divergence.

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

$= \|x\|^2 - 2xy + \|y\|^2$

$= \|x\|^2 - \|y\|^2 + 2\|y\|^2 - 2xy$

$= \|x\|^2 - \|y\|^2 - 2y(x-y)$

$= \|x\|^2 - (\underbrace{\|y\|^2 + 2y(x-y)}_{\text{Tangent of } f \text{ at } y})$

$\Rightarrow F(x) - (F(y) + \langle \nabla F(y), x-y \rangle) \rightarrow$ The Bregman distance definition

$\therefore$ squared Euclidean distance is a Bregman divergence.

(3) Prove that the entropy of the output of this neural network will always be equal or less that of the input.

Ans: $H(X,Y) = H(X|Y) + H(Y) = H(Y|X) + H(X)$ ... ①

We know that $H(Y|X) = H(F(x)|X) = 0$ ... ②

use ①, $H(X, F(x)) = H(X) + H(F(x)|X) = H(F(x)) + H(X|F(x))$

$= H(X) + 0 = H(F(x)) + H(X|F(x))$

$\Rightarrow H(X) = H(F(x)) + H(X|F(x))$ ... ③

and we know $H(X|F(x))$ always ≥ 0, $\because H(x) = -\sum_{i=1}^{n} P(x_i)\log(P_{xi})$

P_{xi} always ≤ 1, then $\log(P_{xi}) < 0$, $\therefore H(x)$ always ≥ 0 ... ④

$\therefore$ with ③ and ④, we know $H(X|F(x)) \ge 0$, $\therefore H(x) - H(F(x)) \ge 0$

$\therefore$ the entropy of the output of this NN will always be equal or less that of the input.

Problem 4: Goodness of Estimation

(1) What is the unbiased estimator with the lowest variance that you can construct from a linear combination of θ1 and θ2, and what's its variance?

Ans: $Var(\theta 1) = 1$, $Var(\theta 2) = 2$, $Cov(\theta 1, \theta 2) = \frac{1}{4}$,

set linear combination of θ1 and θ2 is $C_1\hat{\theta}_1 + C_2\hat{\theta}_2$

We set $C_1 + C_2 = 1$, then $E(C_1\hat{\theta}_1 + C_2\hat{\theta}_2) = C_1\theta + C_2\theta = \theta$, Using $C_2 = 1 - C_1$,

we have: $Var(C_1\hat{\theta}_1 + C_2\hat{\theta}_2) = C_1^2 Var(\hat{\theta}_1) + C_2^2 Var(\hat{\theta}_2) + 2C_1C_2 Cov(\hat{\theta}_1, \hat{\theta}_2)$

$= C_1^2 + 2C_2^2 + \frac{1}{2}C_1C_2$

$= C_1^2 + 2(1-C_1)^2 + \frac{1}{2}C_1(1-C_1)$

$= \frac{5}{2}C_1^2 - \frac{7}{2}C_1 + 2$

$= \frac{5}{2}(C_1 - \frac{7}{10})^2 + \frac{31}{40}$

minimizing the expression, we have $C_1 = \frac{7}{10}$, $C_2 = \frac{3}{10}$

$Var(C_1\hat{\theta}_1 + C_2\hat{\theta}_2) = \frac{5}{2} \times \frac{49}{100} - \frac{7}{2} \times \frac{7}{10} + 2 = \frac{31}{40}$

The minimum variance estimator of this unbiased class is

$\frac{7}{10}\hat{\theta}_1 + \frac{3}{10}\hat{\theta}_2$

(2) Also answer the same question for θ_3 and θ_4.

Ans: $Var(\theta_3)=1, Var(\theta_4)=2, Cov(\theta_3,\theta_4)=\frac{3}{4}$

Set linear combination of θ_3 and θ_4 is $C_3\hat{\theta}_3+C_4\hat{\theta}_4$

We set $C_3+C_4=1$, then $E(C_3\hat{\theta}_3+C_4\hat{\theta}_4)=C_3\theta+C_4\theta=\theta$

Using $C_4=1-C_3$, we have

$$\begin{aligned}Var(C_3\hat{\theta}_3+C_4\hat{\theta}_4)&=C_3^2Var(\hat{\theta}_3)+C_4^2Var(\hat{\theta}_4)+2C_3C_4Cov(\hat{\theta}_3,\hat{\theta}_4)\\&=C_3^2+2C_4^2+\frac{3}{2}C_3C_4\\&=C_3^2+2(1-C_3)^2+\frac{3}{2}C_3(1-C_3)\\&=\frac{3}{2}C_3^2-\frac{5}{2}C_3+2\\&=\frac{3}{2}(C_3-\frac{5}{6})^2+\frac{23}{24}\end{aligned}$$

minimizing the expression, we have $C_3=\frac{5}{6}, C_4=\frac{1}{6}$.

$Var(C_3\hat{\theta}_3+C_4\hat{\theta}_4)=\frac{3}{2}\times\frac{25}{36}-\frac{5}{2}\times\frac{5}{6}+2=\frac{23}{24}$

The minimum variance estimator of this unbiased class is

$\frac{5}{6}\hat{\theta}_3+\frac{1}{6}\hat{\theta}_4$

(3) Ans: When $Cov(\theta_1,\theta_2)=\frac{1}{4}$ can produce an estimator with lower variance, so when the lower Covariance can produce an estimator with lower variance.

Problem 3: Point Estimation

(1) The moment method

for $f(x)=\begin{cases}\frac{1}{\theta}, & \text{for } 0\le x\le\theta\\ 0, & \text{elsewhere}\end{cases} \Rightarrow F(x)=\begin{cases}0, & \text{for } x<0\\ \frac{x}{\theta}, & \text{for } 0\le x\le\theta\\ 1, & \text{for } \theta\le x\end{cases}$

in this case, we have $\mu=E(X_1)=\frac{\theta}{2}$.

$\because \bar{X}=\frac{\theta}{2}$, $\therefore \hat{\theta}=2\bar{X}$

(2) The MAP method

$\hat{\theta}_{MAP}(X)=\arg\max_{\theta} f(x|\theta)g(\theta)$.

$\because g(\theta)$ is the prior distribution and is a uniform prior,

$\therefore \hat{\theta}_{MAP}(X)=\arg\max_{\theta} f(X|\theta)=\hat{\theta}_{MLE}(X)$.

For a uniform distribution, the likelihood function can be written as:

$$L(\theta)=\prod_{i=1}^{n}f(x_i;\theta)=\prod_{i=1}^{n}\frac{1}{\theta}=\frac{1}{\theta^n}$$

$$\ln L(\theta)=\ln\prod_{i=1}^{n}f(x_i;\theta)=\ln\prod_{i=1}^{n}\frac{1}{\theta}=\ln(\theta^{-n})=-n\ln\theta$$

$$\frac{\partial \ln L(\theta)}{\partial\theta}=\frac{-n}{\theta}$$

notice that the derivative with respect to θ is monotonically decreasing. Thus, the $\hat{\theta}_{MLE}$ would be the smallest θ possible, which is:

$$\hat{\theta}_{MLE}=\max(X_1,X_2,\dots,X_n)$$

$$\therefore \hat{\theta}_{MAP}=\max(X_1,X_2,\dots,X_n)$$

(3) The bayesian method using squared error loss function.

Population $f(x;\theta)=\frac{1}{\theta}$, for $0\le x<\theta$

Prior distribution $h(\theta)=1$, for $0\le\theta\le1$

Joint distribution $u(x,\theta)=h(\theta)f(x;\theta)=1(\frac{1}{\theta})=\frac{1}{\theta}$, for $0<x<\theta<1$

Marginal distribution of x

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

The conditional density of θ given x:

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

$$\hat{\theta}=E[\theta/x]=\int_x^1\theta k(\theta/x)d\theta$$

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

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

Problem 4: Interval Estimation.

(1) Show that $Q=X(1)-\theta$ is a pivotal quantity.

$F(x)=\int_\theta^x e^{-(x-\theta)}dx=1-e^{-(x-\theta)}$, for $\theta<x<\infty$

for $f(x;\theta)$, the probability density of the r-th order statistic X_r is given by

$$f_{X_r}(x)=\frac{n!}{(r-1)!(n-r)!}[F(x)]^{r-1}[1-F(x)]^{n-r}f(x)$$

when $r=1$, $f_{X_1}(x)=\frac{n!}{(1-1)!(n-1)!}[F(x)]^{1-1}[1-F(x)]^{n-1}f(x)$

$$=n(e^{-(x-\theta)})^{n-1}\cdot e^{-(x-\theta)}$$

$$=ne^{-n(x-\theta)}, \text{ for } x>\theta$$

for $Q=X(1)-\theta$, $f_Q(q)=ne^{-nq}$, for $q>0$

$F_Q(q)=1-e^{-nq}$, for $q>0$

the distribution of Q is not depend on θ. $\therefore Q=X(1)-\theta$ is a pivotal quantity.

(2) Use this pivotal quantity find a $100(1-\alpha)\%$ confidence interval for θ.

we know $X(1)=\min\{X_1,X_2,\dots,X_n\}$ and $Q=X(1)-\theta$, $F_Q(q)=1-e^{-nq}$, then $p(Q>q)=e^{-nq}$

for $p(\mu_1\le Q\le\mu_2)=1-\alpha$ $\therefore \frac{\alpha}{2}=p(Q\le\mu_1)=1-e^{-n\mu_1}$

$$\Rightarrow e^{-n\mu_1}=1-\frac{\alpha}{2}$$

$$\Rightarrow -n\mu_1=\ln(1-\frac{\alpha}{2})$$

$$\Rightarrow \mu_1=-\frac{\ln(1-\frac{\alpha}{2})}{n}$$

$$\frac{\alpha}{2}=P(Q>\mu_2)=e^{-n\mu_2}$$

$$\Rightarrow e^{-n\mu_2}=\frac{\alpha}{2}$$

$$\Rightarrow -n\mu_2=\ln(\frac{\alpha}{2})$$

$$\Rightarrow \mu_2=-\frac{\ln(\frac{\alpha}{2})}{n}$$

$$\therefore -\frac{\ln(1-\frac{\alpha}{2})}{n}\le Q\le-\frac{\ln(\frac{\alpha}{2})}{n}$$

$$\Rightarrow -\frac{\ln(1-\frac{\alpha}{2})}{n}\le X(1)-\theta\le-\frac{\ln(\frac{\alpha}{2})}{n}$$

$$\Rightarrow \frac{\ln(\frac{\alpha}{2})}{n}+X(1)\le\theta\le\frac{\ln(1-\frac{\alpha}{2})}{n}+X(1).$$

Thus, this pivotal quantity find a $100(1-\alpha)\%$ confidence interval for θ is

$$\left[\frac{\ln(\frac{\alpha}{2})}{n}+X(1),\ \frac{\ln(1-\frac{\alpha}{2})}{n}+X(1)\right]$$