Stat511(Section001): Homework #3

Due on Feb. 2, 2021 at 5:00pm

Instructor:Xuming He

Tiejin Chen **tiejin@umich.edu**

Problem 1

For every i, X_i is a Bernoulli distribution, and we can rewrite the pmf of X_i in one line;

$$P(X_i = x) = (\frac{e^{\theta - a_i}}{1 + e^{\theta - a_i}})^x (\frac{1}{1 + e^{\theta - a_i}})^{1-x} = exp[xlog(e^{\theta - a_i}) - log(1 + e^{\theta - a_i})] = \frac{exp(x(\theta - a_i))}{1 + e^{\theta - a_i}}$$

Hence we can get:

$$f(X) = \frac{exp(\theta \sum_{i=1}^n x_i - \sum_{i=1}^n a_i x_i)}{\prod_{i=1}^n (1 + e^{\theta - a_i})} = \frac{exp(\theta \sum_{i=1}^n x_i)}{\prod_{i=1}^n (1 + e^{\theta - a_i})} exp(-\sum_{i=1}^n a_i x_i) = g_\theta(T) h(x)$$

Hence the one-dim sufficient statistics for θ is $\sum_{i=1}^n x_i$

Problem 2

(i)
we can get:

$$f(x) = \frac{\beta^{n\alpha}}{\Gamma(\alpha)^n} (\prod_{i=1}^n x_i)^{\alpha - 1} e^{-\beta(\sum_{i=1}^n x_i)}$$

Hence, we can get:

$$T(x) = (\prod_{i=1}^n x_i, \sum_{i=1}^n x_i)$$

is a two dimension sufficient statistic for (α, β).
(ii)
when we know the α. Then $(\prod_{i=1}^n x_i)^{\alpha-1}$ is no longer one part of $g_\theta(T(x))$ and become one part of $h(x)$. And thus $g_\theta(T(x)) = \beta^{n\alpha} e^{-\beta(\sum_{i=1}^n x_i)}, h(x) = \frac{1}{\Gamma(\alpha)^n} (\prod_{i=1}^n x_i)^{\alpha - 1}$. Thus we can get:

$$T(x) = \sum_{i=1}^n x_i$$

is one-dim sufficient statistic for β.

Problem 3

(a)
we can know:

$$f(x) = (\frac{1}{\theta})^n \prod_{i=1}^n I(x_i > \theta) \prod_{i=1}^n I(x_i < 2\theta)$$

Also we get$\prod_{i=1}^n I(x_i > \theta)$ is equivalence to $I(x_{(1)} > \theta)$ and $\prod_{i=1}^n I(x_i < 2\theta)$ is equivalance to $I(x_{(n)} < 2\theta)$. Thus, we can get;

$$f_\theta(x) = (\frac{1}{\theta})^n I(x_{(1)} > \theta) I(x_{(n)} < 2\theta)$$

Where we can see $f(x)$ is $g_\theta(T(x))$ and $h(x) = 1$. Thus:

$$T(x) = (x_{(1)}, x_{(n)})$$

is sufficient for θ.Now we prove it is minimal.We can find:

$$x_{(1)} = sup\{\theta : f_\theta(x) > 0\}, x_{(n)} = 2inf\{\theta : f_\theta(x) > 0\}$$

Problem 3 continued on next page. . .

Thus, if S(x) is a sufficient statistics for θ, we can get there exists functions h and g_θ so that $f_\theta(x) = g_\theta(S(x))h(x)$. And since $h(x) > 0$, we have:

$$x_{(1)} = sup\{\theta : f_\theta(S(x)) > 0\}, x_{(n)} = 2inf\{\theta : f_\theta(S(x)) > 0\}$$

Therefore, there is a function Φ such that $T(x) = \Phi(S(x))$ for every $S(x)$. Hence, we conclude that $T(x)$ is minimal sufficient statistics.

(b)

We know if sample mean and sample variance are the sufficient statistics for θ, then by factoriztion theorem(Assumed $S(x) = \bar{X}, S^2$):

$$f_\theta(x) = (\frac{1}{\theta})^n I(x_{(1)} > \theta)I(x_{(n)} < 2\theta) = g_\theta(S(x))h(x)$$

Hence $g_\theta(S(x))$ must contain $(\frac{1}{\theta})^n I(x_{(1)} > \theta)I(x_{(n)} < 2\theta)$. Therefore, $g_\theta(S(x))$ must contain $x_{(1)}$ and $x_{(n)}$, which is contradicted with the fact that $g_\theta(S(x))$ is the function of only sample mean and sample variance. Thus we prove sample mean and sample variance are not sufficient for θ.

Problem 4

(a)

We have:

$$f_\theta(x) = \frac{n!}{x_1!x_2!x_3!x_4!} p_1^{x_1} p_2^{x_2} p_3^{x_3} p_4^{x_4} = h(x)exp[x_1 log(p_1) + x_2 log(p_2) + x_3 log(p_3) + x_4 log(p_4)]$$

We place x_4 by $n - x_1 - x_2 - x_3$. And we can get:

$$\begin{aligned} f_\theta(x) &= h(x)exp[x_1 log(p_1) + x_2 log(p_2) + x_3 log(p_3) + (n - x_1 - x_2 - x_3) log(p_4)] \\ &= h(x)exp(nlog(p_4))exp[x_1 log(\frac{p_1}{p_4}) + x_2 log(\frac{p_2}{p_4}) + x_3 log(\frac{p_3}{p_4})] \\ &= h(x)a(\theta)exp[x_1 log(\frac{2+\theta}{\theta}) + x_2 log(\frac{1-\theta}{\theta}) + x_3 log(\frac{1-\theta}{\theta})] \\ &= h(x)a(\theta)exp[x_1 log(\frac{2+\theta}{\theta}) + (x_2 + x_3) log(\frac{1-\theta}{\theta})] \end{aligned}$$

And thus this is an exponential family with $b(\theta) = (log(\frac{2+\theta}{\theta}), log(\frac{1-\theta}{\theta})), T(x) = (x_1, x_2 + x_3)$. Now we prove $T(x)$ is minimal sufficient for θ. For two sample point x and y. We have:

$$\frac{f(x|\theta)}{f(y|\theta)} = \frac{y_1!y_2!y_3!y_4!}{x_1!x_2!x_3!x_4!} exp(((x_1 - y_1) log(\frac{2+\theta}{\theta}) + (x_2 + x_3 - y_2 - y_3) log(\frac{1-\theta}{\theta}))$$

if $\frac{f(x|\theta)}{f(y|\theta)}$ is constant if and only if $x_1 = y_1, x_2 + x_3 = y_2 + y_3$ since we see that $\frac{y_2!y_3!}{x_2!x_3!}$ is the constant when we think it is a function of θ. Hence, we get the minimal sufficient statistics is $(x_1, x_2 + x_3)$.

(b)

It is. We see the two term in $b(\theta)$. Assumed $log(\frac{2+\theta}{\theta}) = t$, Then we have:

$$log(\frac{2+\theta}{\theta}) = t \Rightarrow \frac{\theta + 2}{\theta} = e^t \Rightarrow \frac{2 - 2\theta}{\theta} = e^t - 3 \Rightarrow log(\frac{1-\theta}{\theta}) = log(\frac{e^t - 3}{2})$$

which is one to one relationship. One term is the function of another. Hence it will be a curved exponential family.