Chapter 6: Principles of Data Reduction

May 12, 2023

Exercise 6.1

Yes

$$\frac{p(x|0,\sigma)}{q(|X||0,\sigma)} = \frac{p(x|0,\sigma)}{p(x|0,\sigma)+p(-x|0,\sigma)} = \frac{1}{2}$$

Does not depend on the paramters.

Exercise 6.2

The pdf for X_i is $f_{X_i}(x|\theta) = \exp(i\theta - x)\mathbb{1}_{x \geq i\theta}$. Then

$$\begin{aligned}
f_X(x_i|\theta) &= \prod_{i=1}^{n} f_{X_i}(x_i|\theta) \\
&= \exp\left(\sum_{i=1}^{n} i\theta - x_i\right) \prod_{i=1}^{n} \mathbb{1}_{x_i \geq i\theta} \\
&= \exp\left(\frac{n(n+1)\theta}{2}\right) \exp\left(\sum_i x_i\right) \prod_{i=1}^{n} \mathbb{1}_{\frac{x_i}{i} \geq \theta} \\
&= \exp\left(\frac{n(n+1)\theta}{2}\right) \mathbb{1}_{\min \frac{x_i}{i} \geq \theta} \exp\left(\sum_i x_i\right)
\end{aligned}$$

$g(T(x)|\theta) = g(\min \frac{x_i}{i}|\theta) = \exp\left(\frac{n(n+1)\theta}{2}\right) \mathbb{1}_{\min \frac{x_i}{i} \geq \theta}$ and $h(x) = \exp\left(\sum_i x_i\right)$. By factorization theorem, it is sufficient statistic.

Exercise 6.3

Given the pdf

$$f(x|\mu,\sigma) = \frac{1}{\sigma}e^{-(x-\mu)/\sigma}, \quad u < x < \infty, 0 < \sigma < \infty$$

We need to get rid of the dependency on u in the range of x. With indicator, we can rewrite it as

$$f(x|\mu,\sigma) = \frac{1}{\sigma}e^{-(x-\mu)/\sigma}\mathbb{1}_{\mu<x}, \quad -\infty < x < \infty, 0 < \sigma < \infty$$

Then

$$\begin{aligned} f(\mathbf{x}|\mu,\sigma) &= \prod_{i=0}^{n} f(x_i|\mu,\sigma) \\ &= \frac{1}{\sigma^n}\exp\left(-\frac{\sum_i x_i - n\mu}{\sigma}\right)\prod_{i=0}^{n}\mathbb{1}_{\mu<x_i} \\ &= \frac{1}{\sigma^n}\exp\left(-\frac{\sum_i x_i - n\mu}{\sigma}\right)\mathbb{1}_{\mu<x_{\min}} \end{aligned}$$

Define $T((x)) = (t_1, t_2) = (\sum_i x_i, x_{\min})$ and $h(\mathbf{x}) = 1$. By factorization theorem, it is a sufficient statistics.

Exercise 6.4

The pdf is

$$\begin{aligned} f(x|\theta) &= \left[\prod_{i}^{n} h(x_i)\right] c^n(\theta)\exp\left(\sum_{i=1}^{n}\sum_{j=1}^{k} w_j(\theta)t_j(x_i)\right) \\ &= H(x)C(\theta)\exp\left(\sum_{j=1}^{k} w_j(\theta)\sum_{i=1}^{n} t_j(x_i)\right) \\ &= H(x)C(\theta)\exp\left(\sum_{j=1}^{k} w_j(\theta)T_j(x)\right) \\ &= H(x)C(\theta)g(T(x)|\theta) \end{aligned}$$

Therefore by Factorization theorem, $T(X) = [T_j(X)] = (\sum_{i=1}^{n} t_j(x_i))$ is sufficient statistics.

Exercise 6.8

The pdf of sample X is $f(x|\theta) = \prod_{i=1}^{n} f(x_i - \theta)$. By theorem 6.2.13, we take the ratio of $f(x|\theta)$ and $f(y|\theta)$,

$$\frac{f(x|\theta)}{f(y|\theta)} = \frac{\prod_{i=1}^{n} f(x_i - \theta)}{\prod_{i=1}^{n} f(y_i - \theta)}$$

Note that the above expression is a constant of θ only when the terms cancel out which implies there exists an ordering T such that $f(T(x_i) - \theta) = f(T(y_i) - \theta)$. Order statistics is such an ordering. Therefore it is the minimal statistics.

Exercise 6.9

(a) the ratio is

$$\frac{f(x|\theta)}{f(y|\theta)} = \exp\left(-\frac{1}{2}\sum_i (x_i-\theta)^2 + \frac{1}{2}\sum_i (y_i-\theta)^2\right) = \exp\left(-\frac{1}{2}\sum_i (x_i^2 - y_i^2) + n\theta(\bar{x}-\bar{y})\right)$$

For the ration to be constant function of θ iff $\bar{x} = \bar{y}$. So sample mean is the min sufficient statistics.

(b) the joint pdf is $f(x|\theta) = \exp(-\sum_i (x_i - \theta))$ Where $x_i > \theta, \theta \in \Re$. We can rewrite this with indicator function,

$$f(x|\theta) = I_{\min x > \theta} \exp\left(-\sum_i (x_i - \theta)\right), \text{ where } x_i, \theta \in \Re$$

So the ratio becomes

$$\frac{f(x|\theta)}{f(y|\theta)} = \frac{I_{\min x>\theta}\exp(-\sum_i (x_i-\theta))}{I_{\min y>\theta}\exp(-\sum_i (y_i-\theta))} = \frac{I_{\min x>\theta}\exp(-\sum_i x_i)}{I_{\min y>\theta}\exp(-\sum_i y_i)}$$

$\frac{I_{\min x>\theta}}{I_{\min y>\theta}}$ cancels out iff $\min x = \min y$ which is the minimal sufficient statistics.

(c) The ratio is the following

$$\frac{f(x|\theta)}{f(y|\theta)} = \frac{\exp(\bar{y}-\bar{x})}{\prod_i (1+\exp(x_i-\theta))^2/(1+\exp(y_i-\theta))^2}$$

The only way for the ratio to be constant of θ is to have the bottom term cancels out which means for each i, there is j such that $1+\exp(x_i - \theta) = 1+\exp(y_j - \theta)$. This can only happen when T is order statistics of x and $T(x) = T(y)$. Therefore order statistics is the minimal sufficient statistics.

(d) Same reasoning as (c), order statistics is the minimal sufficient statistics.

(e) The ratio is

$$\frac{f(x|\theta)}{y|\theta} = \exp\left\{-\sum_i |x_i-\theta| + \sum_i |y_i - \theta|\right\}$$

The expression in the exponent is the difference between the distance from θ to x and that of y. Since θ can be any value, so the difference between x and y to θ cannot stay constant unless the data point in x is the same as data point in y which means ordered statistic is the minimal sufficient statistics.

Exercise 6.12

(a.1)The distribution of N is just $P(N = k) = p_k$ independent of θ, therefore N is ancillary statistics.

(a.2)Suppose $(X(x), N(x)) = (X(y), N(y))$, then x and y have the same successes and same length, so the ratio $f(x|\theta)/f(y|\theta)$ is constant 1.

Now Suppose $x = (x_1, \ldots, x_{n_1})$ and $y = (y_1, \ldots, y_{n_2})$ such that $f(k_1|\theta)/f(k_2|\theta)$ is constant (Here $k_1 = X(x)$ and $k_2 = X(y)$ are successes). Then

$$\frac{f(k_1, n_1|\theta)}{f(k_2, n_2|\theta)} = \frac{f(k_1|\theta, n_1)p(n_1)}{f(k_2|\theta, n_2)p(n_2)} = \frac{\binom{n_1}{k_1}}{\binom{n_2}{k_2}}\theta^{k_1-k_2}(1-\theta)^{n_1-n_2-(k_1-k_2)}\frac{p_{n_1}}{p_{n_2}}$$

For the expression to be constant, we get $k_1 = k_2$ and $n_1 - n_2 - (k_1 - k_2) = 0 \Rightarrow n_1 = n_2$ So there (X, N) is minimal statistics.

(b) Since the bias of the estimator X/N is $\mathbf{E}[X/N - \theta]$.

$$\begin{aligned}
\text{Bias}(X/N, \theta) &= \mathbf{E}[X/N - \theta] \\
&= \mathbf{E}[X/N] - \theta \\
&= \sum_{X,N} \frac{X}{N} P(X, N|\theta) - \theta \\
&= \sum_{n=1}^{\infty} \sum_{k=1}^{n} \frac{k}{n} P(X = k|N = n, \theta)P(N = n) - \theta \\
&= \sum_{n=1}^{\infty} \sum_{k=1}^{n} \frac{k}{n} \binom{n}{k} \theta^k (1-\theta)^{n-k} p_n - \theta \\
&= \sum_{n=1}^{\infty} \frac{p_n}{n} \sum_{k=1}^{n} k \binom{n}{k} \theta^k (1-\theta)^{n-k} - \theta \\
&= \sum_{n=1}^{\infty} \frac{p_n}{n} \mathbf{E}_{\text{Binomial}(n,\theta)}[X] - \theta \\
&= \sum_{n=1}^{\infty} \frac{p_n}{n} n\theta - \theta \\
&= \theta - \theta = 0
\end{aligned}$$

Exercise 6.13

Since $f(x, y|\alpha) = \alpha(xy)^{\alpha-1}e^{-x^\alpha - y^\alpha}$

Let

$$\begin{cases} s = \log x/\log y \\ t = \log x + \log y \end{cases} \Rightarrow \begin{cases} x = \exp\left(\frac{st}{s+1}\right) \\ y = \exp\left(\frac{t}{s+1}\right) \end{cases}$$

$$\Rightarrow \frac{\partial(x,y)}{\partial(s,t)} = \begin{vmatrix} \frac{t}{(s+1)^2}e^{st/(s+1)} & \frac{s}{s+1}e^{st/(s+1)} \\ -\frac{t}{(s+1)^2}e^{t/(s+1)} & \frac{1}{s+1}e^{t/(s+1)} \end{vmatrix} = \frac{t}{(s+1)^2}e^t$$

Then by change of variables,

$$f(s,t|\alpha) = f(x,y|\alpha)\left|\frac{\partial(x,y)}{\partial(s,t)}\right| = \alpha\frac{\alpha t}{(s+1)^2}e^{\alpha t}e^{-\exp\left(\frac{\alpha ts}{s+1}\right)-\exp\left(\frac{\alpha t}{s+1}\right)}$$

We want to find $f(s|\alpha) = \int_t f(s,t|\alpha)dt$. If we let $u = \alpha t$, then $\alpha dt = du$. Therefore

$$\begin{aligned} f(s|\alpha) &= \int_t f(s,t|\alpha)dt \\ &= \int_t \frac{\alpha t}{(s+1)^2}e^{\alpha t}\exp\left\{-\exp\left(\frac{\alpha ts}{s+1}\right)-\exp\left(\frac{\alpha t}{s+1}\right)\right\}\alpha dt \\ &= \int_u \frac{u}{(s+1)^2}e^{u}\exp\left\{-\exp\left(\frac{us}{s+1}\right)-\exp\left(\frac{u}{s+1}\right)\right\}du \\ &= f(s) \end{aligned}$$

α has vanished from the final expression, therefore the distribution of $s = \log x/\log y$ which does not depend on α is ancillary.