Mathematical Statistics I
Assignment 3 - Solutions

1. (a) Consider two sample points x and y, we have

$$\begin{aligned}\frac{f(x,n;\theta)}{f(y,n';\theta)} &= \frac{f(x \mid N=n)P(N=n)}{f(y \mid N=n')\,P(N=n')} \\ &= \frac{\binom{n}{x}\theta^x(1-\theta)^{n-x}p_n}{\binom{n'}{y}\theta^y(1-\theta)^{n'-y}p_{n'}} = \theta^{x-y}(1-\theta)^{n-n'-x+y}\frac{\binom{n}{x}p_n}{\binom{n'}{y}p_{n'}}.\end{aligned}$$

The last ratio does not depend on θ. The other terms are constant as a function of θ if and only if $n = n'$ and $x = y$. So (X, N) is minimal sufficient for θ. Because $P(N = n) = p_n$ does not depend on θ, N is ancillary for θ. The point is that although N is independent of θ, the minimal sufficient statistic contains N in this case. A minimal sufficient statistic may contain an ancillary statistic.

(b) Using the fact that $X \mid N = n \sim \text{binomial}(n, \theta)$, we have

$$\begin{aligned}\mathrm{E}\left(\frac{X}{N}\right) &= \mathrm{E}\left(\mathrm{E}\left(\frac{X}{N}\middle| N\right)\right) = \mathrm{E}\left(\frac{1}{N}\mathrm{E}(X \mid N)\right) = \mathrm{E}\left(\frac{1}{N}N\theta\right) = \mathrm{E}(\theta) = \theta \\ \mathrm{Var}\left(\frac{X}{N}\right) &= \mathrm{Var}\left(\mathrm{E}\left(\frac{X}{N}\middle| N\right)\right) + \mathrm{E}\left(\mathrm{Var}\left(\frac{X}{N}\middle| N\right)\right) = \mathrm{Var}(\theta) + \mathrm{E}\left(\frac{1}{N^2}\mathrm{Var}(X \mid N)\right) \\ &= 0 + \mathrm{E}\left(\frac{N\theta(1-\theta)}{N^2}\right) = \theta(1-\theta)\mathrm{E}\left(\frac{1}{N}\right).\end{aligned}$$

2. To prove $T(\mathbf{X}) = \left(X_{(1)}, X_{(n)}\right)$ is not complete, we want to find $g[T(\mathbf{X})]$ such that $\mathrm{E}g[T(\mathbf{X})] = 0$ for all θ, but $g[T(\mathbf{X})] \not\equiv 0$. A natural candidate is $R = X_{(n)} - X_{(1)}$, the range of $\mathbf{X}$, because its distribution does not depend on θ due to the fact that parameter of the distribution being a location parameter. Also notice that, $R \sim \text{Beta}(n-1, 2)$. Thus $\mathrm{E}(R) = (n-1)/(n+1)$ does not depend on θ, and $\mathrm{E}(R-\mathrm{E}(R)) = 0$ for all θ. Thus $g\left[X_{(n)}, X_{(1)}\right] = X_{(n)} - X_{(1)} - (n-1)/(n+1) = R-\mathrm{E}(R)$ is a nonzero function whose expected value is always 0. So, $\left(X_{(1)}, X_{(n)}\right)$ is not complete. This problem can be generalized to show that if a function of a sufficient statistic is ancillary, then the sufficient statistic is not complete, because the expectation of that function does not depend on θ.

3. (a) Note that,

$$f(x_1, x_2, \ldots x_n; \theta) = \prod_{i=1}^{n} 2x_i\theta^{-2}I(x_i < \theta) = \left(2\theta^{-2}\right)^n \left(\prod_{i=1}^{n} x_i\right) I(x_{(n)} < \theta),$$

hence $Y = X_{(n)}$ is sufficient by factorization theorem and we have,

$$f_Y(y) = \frac{2n}{\theta^{2n}}y^{2n-1}, \quad 0 < y < \theta.$$

For a function $g(y)$,

$$\mathrm{E}[g(Y)] = \int_0^\theta g(y)\frac{2n}{\theta^{2n}}y^{2n-1}dy = 0 \text{ for all } \theta \text{ implies } g(\theta)\frac{2n\theta^{2n-1}}{\theta^{2n}} = 0 \text{ for all } \theta$$

by taking derivatives. This can only be zero if $g(\theta) = 0$ for all θ, so $Y = X_{(n)}$ is complete.
(b) Consider sample points $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, x_y)$ we then have,

$$\frac{f(x;\theta)}{f(y;\theta)} = \frac{e^{-\sum_{i=1}^n (x_i-\theta)}\exp\left(-\sum_{i=1}^n e^{-(x_i-\theta)}\right)}{e^{-\sum_{i=1}^n (y_i-\theta)}\exp\left(-\sum_{i=1}^n e^{-(y_i-\theta)}\right)} = e^{-\sum_{i=1}^n (x_i-y_i)}\exp\left(-\sum_{i=1}^n\left(e^{-(x_i-\theta)} - e^{-(y_i-\theta)}\right)\right)$$

This is constant as a function of θ if and only if x and y have the same order statistics. Therefore, the order statistics $T = \left(X_{(1)}, \ldots, X_{(n)}\right)$ are minimal sufficient for θ. Notice that θ is a location parameter, thus, the range $R = X_{(n)} - X_{(1)}$ is ancillary, and its expectation does not depend on θ. So $R - \mathrm{E}(R)$ is a function of the sufficient statistic whose value is always zero, hence T is not complete.
(c) Note that,

$$f(x_1, x_2, \ldots x_n; \theta) = \prod_{i=1}^n \binom{2}{x_i}\theta^{x_i}(1-\theta)^{2-x_i} = \theta^{\sum_{i=1}^n x_i}(1-\theta)^{2n-\sum_{i=1}^n x_i}\prod_{i=1}^n\binom{2}{x_i},$$

which can be rewritten as an exponential family form,

$$f(x_1, x_2, \ldots x_n; \theta) = (1-\theta)^{2n}\left[\prod_{i=1}^n\binom{2}{x_i}\right]\exp\left(\log\left(\frac{\theta}{1-\theta}\right)\sum_{i=1}^n x_i\right),$$

considering that the parameter space contains an open set in $\mathbb{R}$, $\sum_{i=1}^n X_i$ is a complete sufficient statistic.

4. (a) X is obviously sufficient because it is the whole data. To check completeness, calculate

$$\mathrm{E}[g(X)] = \frac{\theta}{2}g(-1) + (1-\theta)g(0) + \frac{\theta}{2}g(1).$$

If $g(-1) = g(1)$ and $g(0) = 0$, then $\mathrm{E}[g(X)] = 0$ for all θ, but $g(x)$ need not be identically 0 . So the family is not complete.
(b) Note that $f(x;\theta)$ depends on θ only through the value of $|x|$, hence $|X|$ is sufficient by the factorization theorem. The distribution of $|X|$ is Bernoulli, because $\mathrm{P}(|X| = 0) = 1-\theta$ and $\mathrm{P}(|X| = 1) = \theta$. Now we prove that binomial family of distributions (Bernoulli is a special case) is complete. Suppose that T has a binomial(n, p) distribution, $0 < p < 1$. Let g be a function such that $\mathrm{E}_p[g(T)] = 0$. Then

$$\begin{aligned}0 = \mathrm{E}_p[g(T)] &= \sum_{t=0}^n g(t)\binom{n}{t}p^t(1-p)^{n-t} \\ &= (1-p)^n\sum_{t=0}^n g(t)\binom{n}{t}\left(\frac{p}{1-p}\right)^t\end{aligned}$$

for all $p, 0 < p < 1$. The factor $(1-p)^n$ is not 0 for any p in this range. Thus it must be that

$$0 = \sum_{t=0}^n g(t)\binom{n}{t}\left(\frac{p}{1-p}\right)^t = \sum_{t=0}^n g(t)\binom{n}{t}r^t$$

for all $r, 0 < r < \infty$. But the last expression is a polynomial of degree n in r, where the coefficient of r^t is $g(t)\binom{n}{t}$. For the polynomial to be 0 for all r, each coefficient must be 0. Since none of the $\binom{n}{t}$ terms is 0, this implies that $g(t)=0$ for $t=0,1,\ldots,n$. Since T takes on the values $0,1,\ldots,n$ with probability 1 , this yields that $P_p(g(T)=0)=1$ for all p, the desired conclusion. Hence, T is a complete statistic.

(c) We can rewrite $f(x;\theta)$ as,

$$f(x;\theta)=(1-\theta)\exp\left(|x|\log[\theta/(2(1-\theta)]\right)$$

which is the form of an exponential family.

5. Note that for $\lambda=0, \mathrm{E}\left[g(X)\right]=g(0)$ and for $\lambda=1$ we have,

$$\mathrm{E}\left[g(X)\right]=e^{-1}g(0)+e^{-1}\sum_{x=1}^{\infty}\frac{g(x)}{x!}.$$

Let $g(0)=0$ and $\sum_{x=1}^{\infty}\frac{g(x)}{x!}=0$, so $\mathrm{E}\left[g(X)\right]=0$ for $\lambda=0,1$, but $g(x)$ is not identically zero. (For example, take $g(0)=0$, $g(1)=1, g(2)=-2, g(x)=0$ for $x\geq 3$.)