CB-6.0: To prove the minimal sufficient statistic \( T(X) = (X_{(n)}, X_{(1)}) \) is not complete, we want to find \( g[T(X)] \) such that \( E g[T(X)] = 0 \) for all \( \theta \), but \( g[T(X)] \) ≠ 0. In example 6.2.17, it is shown that \( R = X_{(n)} - X_{(1)} \) follows distribution \( \text{beta}(n - 1, 2) \), which does not depend on \( \theta \). Then, \( ER = (n - 1)/(n + 1) \) does not depend on \( \theta \). Thus, we can define \( g[T(X)] = g[X_{(n)}, X_{(1)}] = X_{(n)} - X_{(1)} - (n - 1)/(n + 1) \) which is a nonzero function and has expected value equal to 0. So, \( T(X) = (X_{(n)}, X_{(1)}) \) is not complete.

CB-6.16: (a) The density of the multinomial vector is given by

\[
f(x|\theta) = \frac{n!}{x_1!x_2!x_3!x_4!} (\frac{1}{2} + \frac{\theta}{4})^{x_1} (\frac{1 - \theta}{4})^{(x_2 + x_3)} (\frac{\theta}{4})^{x_4}
\]

where \( n = x_1 + x_2 + x_3 + x_4 \) and \( \theta \) is fixed. And the density can be further written as

\[
f(x|\theta) = \frac{n!}{x_1!x_2!x_3!x_4!} \exp\{n \log(\frac{\theta}{4})\} \exp\{x_1 \log(\frac{2 + \theta}{\theta}) + (x_2 + x_3) \log(\frac{1 - \theta}{\theta})\}
\]

So, the density belongs to the exponential family. It follows that \( d = 1 \) and \( k = 2 \). Thus, it is a curved exponential family since \( d \leq k \).

(b) By Theorem 6.2.10, it follows that \( T(X) = (X_1, X_2 + X_3) \) is a sufficient statistics for \( \theta \).

(c) Let \( x \) and \( y \) be two sample points, then

\[
\frac{f(x|\theta)}{f(y|\theta)} = \frac{n}{y_1!y_2!y_3!y_4!} \left(\frac{2 + \theta}{\theta}\right)^{(x_1 - y_1)} \left(\frac{1 - \theta}{\theta}\right)^{(x_2 + x_3) - (y_2 + y_3)}
\]

which is a constant of \( \theta \) if and only if \( x_1 = y_1 \) and \( x_2 + x_3 = y_2 + y_3 \). First, if \( x_1 = y_1 \) and \( x_2 + x_3 = y_2 + y_3 \), then \( \frac{f(x|\theta)}{f(y|\theta)} = \frac{y_1!y_2!y_3!y_4!}{x_1!x_2!x_3!x_4!} \) is just a function of data. Second, we show the only if part by contradiction. For example, if \( x_1 \neq y_1 \), then \( \frac{f(x|\theta)}{f(y|\theta)} = \frac{y_1!y_2!y_3!y_4!}{x_1!x_2!x_3!x_4!} (\frac{2 + \theta}{\theta})^a \), where \( a \) is non-zero constant and is equivalent to \( x_1 - y_1 \). In this case, the ratio is a function of \( \theta \). Thus, the assumption of \( x_1 \neq y_1 \) does not
hold. Similarly, we can show that if $x_2 + x_3 \neq y_2 + y_3$, the ratio is also a function of \( \theta \). Thus, by Theorem 6.2.13 and (b), \( T(X) = (X_1, X_2 + X_3) \) is a minimal sufficient statistics for \( \theta \).

CB-6.17: The pmf is \( f(x|\theta) = \theta(1 - \theta)^{x-1} \) which belongs to the exponential family. Thus, by Theorem 6.2.10, it follows that \( T(X) = \sum_{i=1}^n X_i \) is a sufficient statistic for \( \theta \). The mgf of \( X \) is \( M_X(t) = \frac{pe^t}{1-(1-p)e^t} \). Define \( Y = \sum_{i=1}^n x_i - n \), then the mgf of \( Y \) is \( M_Y(t) = e^{nt} \prod e^{x_i} = (\frac{p}{1-(1-p)e^t})^n \) which is exactly the mgf for negative binomial \( (n, \theta) \). Thus, \( \sum_{i=1}^n X_i - n \) follows negative binomial \( (n, \theta) \). By Theorem 6.2.25, it follows that \( T(X) = \sum_{i=1}^n X_i \) is a complete statistic for \( \theta \). So, the family distribution of \( \sum X_i \) is complete.

CB-6.23: Let \( x \) and \( y \) be two sample points, then

\[
\frac{f(x|\theta)}{f(y|\theta)} = \frac{\theta^{-n}I_{(x_1/2, x_1)}(\theta)}{\theta^{-n}I_{(y_1/2, y_1)}(\theta)} = \frac{I_{(x_1)}(\theta)}{I_{(y_1)}(\theta)}
\]

which is constant of \( \theta \) if and only if \( x_1 = y_1 \) and \( x_1 = y_1 \). First, if \( x_1 = y_1 \) and \( x_1 = y_1 \), then the ratio equal to constant 1 when \( x_1 < \theta < y_1 \). Second, show the only if part by contradiction. For example, if \( x_1 \neq y_1 \) and \( x_1 \leq y_1 \), then we can get

\[
\frac{f(x|\theta)}{f(y|\theta)} = \begin{cases} 
1 & \text{if } x_1/2 \leq \theta \leq x_1 \\
0 & \text{if } x_1 \leq \theta \leq y_1 
\end{cases}
\]

So, the ratio is no longer a constant function of \( \theta \). Similarly, we can show the ratio is not a constant under other situations when the assumption of \( x_1 = y_1 \) and \( x_1 = y_1 \) is not true. Thus, \( T(X) = (X_1, X_1) \) is a minimal sufficient statistic for \( \theta \). \( \text{unif}(\theta, 2\theta) \) is a scale family with standard pdf \( f(z) \sim \text{unif}(1,2) \). So if \( Z_1, Z_2, \cdots, Z_n \) is a random sample from \( \text{unif}(1,2) \), then \( X_1 = \theta Z_1, X_2 = \theta Z_2, \cdots, X_n = \theta Z_n \) is a random sample from \( \text{unif}(\theta, 2\theta) \) and \( X_1 = \theta Z_1 \) and \( X_1 = \theta Z_1 \). So the distribution of \( X_1/X_1 = Z_1/Z_1 \) does not depend on \( \theta \) and the function of the sufficient statistic \( g(T(X)) = X_1/X_1 \) is ancillary. Thus, as in Exercise 6.10, \( (X_1, X_1) \) is not complete.

CB-6.30: (a)Let \( x \) and \( y \) be two sample points, then

\[
\frac{f(x|\mu)}{f(y|\mu)} = \frac{\epsilon^{\sum(x_i-\mu)}I(x_1 \geq \mu)}{\epsilon^{\sum(y_i-\mu)}I(y_1 \geq \mu)} = \epsilon^{\sum y_i - \sum x_i} \frac{I(x_1 \geq \mu)}{I(y_1 \geq \mu)}
\]

is a constant of \( \mu \) if and only if \( x_1 = y_1 \). First, if \( x_1 = y_1 \), then the ratio is \( \epsilon^{\sum y_i - \sum x_i} \) which is a constant function of \( \mu \). Second, show only if part by contradiction. If \( x_1 \neq y_1 \) and \( x_1 \leq y_1 \), then we can get

\[
\frac{f(x|\mu)}{f(y|\mu)} = \begin{cases} 
1 & \text{if } \mu \leq x_1 \\
0 & \text{if } x_1 \leq \mu \leq y_1 
\end{cases}
\]
So, the ratio is no longer a constant function of $\mu$. Similarly, the ratio is not a constant function of $\mu$ when $x(1) \geq y(1)$. Thus, by Theorem 6.2.13, $X(1)$ is a minimal sufficient statistic. Define $Y = X(1)$, by Theorem 5.4.4, the pdf of $Y$ is

$$f_Y(y) = n\exp\{-n(y - \mu)\}; y \geq \mu$$

Then, $E_\mu g(Y) = \int_\mu^\infty g(y)ne^{-n(y-\mu)}dy = e^{n\mu} \int_\mu^\infty g(y)ne^{-ny}dy$. Since $ne^{n\mu} \geq 0$ for all $\mu$, if $E_\mu g(Y) = 0$ for all $\mu$ then $\int_\mu^\infty g(y)e^{-ny}dy = 0$ for all $\mu$. And it follows that

$$0 = \frac{d}{d\mu} \left[ \int_\mu^\infty g(y)e^{-ny}dy \right] = -g(\mu)e^{-n\mu}$$

This implies $g(\mu) = 0$ for all $\mu$. Thus, $Y = X(1)$ is a complete sufficient statistic.

(b) Since $X(1)$ is a complete and minimal sufficient statistic as shown in (a), by Basu’s Theorem, we need to show $S^2$ is an ancillary statistic in order to show the independence of $S^2$ and $X(1)$. Note that $f(x|\mu)$ is a location family. So we can write $X_i = Z_i + \mu$ where $Z_1, \ldots, Z_n$ is a sample from $f(x|0)$. Then, we can get

$$S^2 = \frac{1}{n-1} \sum (X_i - \bar{X})^2 = \frac{1}{n-1} \sum (Z_i - \bar{Z})^2$$

which is a function of only $Z_1, \ldots, Z_n$. Thus, the distribution of $S^2$ does not depend on $\mu$. So, $S^2$ is ancillary and independent of $X(1)$. 

©Sujit Ghosh, NCSU Statistics