Mathematical Statistics I
Assignment 2 - Solutions

1. (i) First we write the joint probability density function of $(X_1, X_2, \ldots, X_n)$,

$$f(x_1, x_2, \ldots x_n; \theta) = \prod_{i=1}^{n} \frac{1}{\mathrm{B}(a,b)} x_i^{a-1}(1-x_i)^{b-1} = \mathrm{B}(a,b)^{-n} \left(\prod_{i=1}^{n} x_i\right)^{a-1} \left(\prod_{i=1}^{n}(1-x_i)\right)^{b-1} \tag{1}$$

where $\theta = (a, b)$ and we used the independence assumption in obtaining the first equality. Let $u(X_1, \ldots, X_n) = \left(\prod_{i=1}^{n} X_i, \prod_{i=1}^{n}(1-X_i)\right)$, we can rewrite (1) as

$$f(x_1, x_2, \ldots x_n; \theta) = k_1 \left[u\left(x_1, x_2, \ldots, x_n\right); \theta\right] k_2\left(x_1, x_2, \ldots, x_n\right)$$

where $k_1 \left[u\left(x_1, x_2, \ldots, x_n\right); \theta\right] = \mathrm{B}(a,b)^{-n} \left(\prod_{i=1}^{n} x_i\right)^{a} \left(\prod_{i=1}^{n}(1-x_i)\right)^{b}$ and $k_2\left(x_1, x_2, \ldots, x_n\right) = 1/\prod_{i=1}^{n} x_i(1-x_i)$. Hence by factorization theorem $u(X_1, \ldots, X_n)$ is a sufficient statistic for θ.
(ii) By repeating the steps we went through in the previous section with $a = b = \theta$ and $u(X_1, \ldots, X_n) = \prod_{i=1}^{n} X_i\,(1-X_i)$ we have,

$$f(x_1, x_2, \ldots x_n; \theta) = k_1 \left[u\left(x_1, x_2, \ldots, x_n\right); \theta\right] k_2\left(x_1, x_2, \ldots, x_n\right)$$

where $k_1 \left[u\left(x_1, x_2, \ldots, x_n\right); \theta\right] = \mathrm{B}(\theta,\theta)^{-n} \left(\prod_{i=1}^{n} x_i(1-x_i)\right)^{\theta}$ and $k_2\left(x_1, x_2, \ldots, x_n\right) = 1/\prod_{i=1}^{n} x_i(1-x_i)$. Hence by factorization theorem $u(X_1, \ldots, X_n)$ is a sufficient statistic for θ.

2. First we write the joint probability density function of $(X_1, X_2, \ldots, X_n)$,

$$f(x_1, x_2, \ldots x_n; \theta) = \prod_{i=1}^{n} a(\theta) h(x_i) I(\theta_1 < x_i < \theta_2) = a^n(\theta) \left(\prod_{i=1}^{n} h(x_i)\right) I(\theta_1 < x_{(1)} < x_{(n)} < \theta_2) \tag{2}$$

where $\theta = (\theta_1, \theta_2)$ and we used the independence assumption in obtaining the first equality. Let $u(X_1, \ldots, X_n) = \left(X_{(1)}, X_{(n)}\right)$, we can rewrite (2) as

$$f(x_1, x_2, \ldots x_n; \theta) = k_1 \left[u\left(x_1, x_2, \ldots, x_n\right); \theta\right] k_2\left(x_1, x_2, \ldots, x_n\right),$$

where $k_1 \left[u\left(x_1, x_2, \ldots, x_n\right); \theta\right] = a^n(\theta) I(\theta_1 < x_{(1)} < x_{(n)} < \theta_2)$ and $k_2\left(x_1, x_2, \ldots, x_n\right) = \prod_{i=1}^{n} h(x_i)$. Hence by factorization theorem $u(X_1, \ldots, X_n)$ is a sufficient statistic for θ. Note that with $a(\theta) = (\theta_2 - \theta_1)^{-1}$ and $h(x) = 1$ for all x, $f_\theta(x)$ is the density of $U\left(\theta_1, \theta_2\right)$ family of distributions.

3. (i) First we write the joint probability density function of $(X_1, X_2, \ldots, X_n)$,

$$f(x_1, x_2, \ldots x_n; \theta) = \prod_{i=1}^{n} \frac{1}{\theta_2 - \theta_1} I(\theta_1 < x_i < \theta_2) = (\theta_2 - \theta_1)^{-n} I(\theta_1 < x_{(1)} < x_{(n)} < \theta_2), \tag{3}$$

where $\theta = (\theta_1, \theta_2)$ and we used the independence assumption in obtaining the first equality. Let $u(X_1, \ldots, X_n) = \left(X_{(1)}, X_{(n)}\right)$, we can rewrite (3) as

$$f(x_1, x_2, \ldots x_n; \theta) = k_1 \left[u\left(x_1, x_2, \ldots, x_n\right); \theta\right] k_2 \left(x_1, x_2, \ldots, x_n\right)$$

where $k_1 \left[u\left(x_1, x_2, \ldots, x_n\right); \theta\right] = (\theta_2 - \theta_1)^{-n} I(\theta_1 < x_{(1)} < x_{(n)} < \theta_2)$ and $k_2 \left(x_1, x_2, \ldots, x_n\right) = 1$. Hence by factorization theorem $u(X_1, \ldots, X_n)$ is a sufficient statistic for θ.
(ii) Let $Y = \left(X_{(1)}, X_{(n)}\right)$, the density of Y is,

$$f_Y(x, y; \theta) = n(n-1)\frac{(y-x)^{n-2}}{(\theta_2 - \theta_1)^n} I(\theta_1 < x < y < \theta_2),$$

Since the quotient,

$$\frac{f\left(x_1, x_2, \ldots, x_n; \theta\right)}{f_Y\left(x_{(1)}, x_{(n)}; \theta\right)} = \frac{1}{n(n-1)(x_{(n)} - x_{(1)})^{n-2}}$$

does not depend upon θ, by definition, Y is a sufficient statistic for θ.

4. Note that X is a random variable with distribution function $F((x-a)/b)$ if and only if there exists a random variable Z with distribution function $F(z)$ and $X = bZ + a$. Hence we can work with a random sample $Z_1, \ldots, Z_n$ from distribution function $F(x)$ (corresponding to $a = 0$ and $b = 1$) where $X_1 = bZ_1 + a, \ldots, X_n = bZ_n + a$.
(i) The distribution function of the statistic $Y_i = \left(X_1 - X_i\right)/b, i = 2, \ldots, n$ is,

$$F_{Y_i}(y; a, b) = \mathrm{P}\left(\left(X_1 - X_i\right)/b \le y\right) = \mathrm{P}\left(Z_1 - Z_i \le y\right).$$

The last probability does not depend on a and b because the distribution of $Z_1, \ldots, Z_n$ does not depend on a and b. Thus, the distribution function of Y_i does not depend on a and b and, hence, Y_i is an ancillary statistic.
(ii) The distribution function of the statistic $Y_i = \left(X_1 - a\right)/\left(X_i - a\right), i = 2, \ldots, n$ is,

$$F_{Y_i}(y; a, b) = \mathrm{P}\left(\left(X_1 - a\right)/\left(X_i - a\right) \le y\right) = \mathrm{P}\left(Z_1/Z_i \le y\right).$$

The last probability does not depend on a and b because the distribution of $Z_1, \ldots, Z_n$ does not depend on a and b. Thus, the distribution function of Y_i does not depend on a and b and, hence, Y_i is an ancillary statistic.
(iii) The distribution function of the statistic $Y_i = \left(X_1 - X_i\right)/\left(X_2 - X_i\right), i = 3, \ldots, n$ is,

$$F_{Y_i}(y; a, b) = \mathrm{P}\left(\left(X_1 - X_i\right)/\left(X_2 - X_i\right) \le y\right) = \mathrm{P}\left(\left(Z_1 - Z_i\right)/\left(Z_2 - Z_i\right) \le y\right).$$

The last probability does not depend on a and b because the distribution of $Z_1, \ldots, Z_n$ does not depend on a and b. Thus, the distribution function of Y_i does not depend on a and b and, hence, Y_i is an ancillary statistic.

5. (i) First we write the joint probability density function of $(X_1, X_2, \ldots, X_n)$,

$$f(x_1, x_2, \ldots x_n; \theta) = \prod_{i=1}^{n} \frac{1}{\theta} I(x_i < \theta) = \theta^{-n} I(x_{(n)} < \theta), \tag{4}$$

where we used the independence assumption in obtaining the first equality. Let $u(X_1, \ldots, X_n) = X_{(n)}$, we can rewrite (4) as

$$f(x_1, x_2, \ldots x_n; \theta) = k_1 \left[u\left(x_1, x_2, \ldots, x_n\right); \theta\right] k_2 \left(x_1, x_2, \ldots, x_n\right)$$

where $k_1\left[u\left(x_1, x_2, \ldots, x_n\right); \theta\right]=\theta^{-n} I(x_{(n)}<\theta)$ and $k_2\left(x_1, x_2, \ldots, x_n\right)=1$. Hence by factorization theorem $u(X_1, \ldots, X_n)$ is a sufficient statistic for θ.
Now let $T=X_{(n)}$, the density of T is,

$$f_T(x)=n \frac{x^{n-1}}{\theta^n} I(x<\theta).$$

Since the quotient,

$$\frac{f\left(x_1, x_2, \ldots, x_n; \theta\right)}{f_T\left(x_{(n)}; \theta\right)}=\frac{1}{n(x_{(n)})^{n-1}}.$$

does not depend upon θ, by definition, T is a sufficient statistic for θ.
(ii) First we write the joint probability density function of $(X_1, X_2, \ldots, X_n)$,

$$f(x_1, x_2, \ldots x_n; \theta)=\prod_{i=1}^n \frac{1}{b} e^{-(x-a)/b} I(x_i>a)=b^{-n} \exp \left(\frac{1}{b} \sum_{i=1}^n(x_i-a)\right) I(x_{(1)}>a), \tag{5}$$

where $\theta=(a, b)$ and we used the independence assumption in obtaining the first equality. Let $u(X_1, \ldots, X_n)=\left(X_{(1)}, \sum_{i=1}^n X_i\right)$, we can rewrite (5) as

$$f(x_1, x_2, \ldots x_n; \theta)=k_1\left[u\left(x_1, x_2, \ldots, x_n\right); \theta\right] k_2\left(x_1, x_2, \ldots, x_n\right)$$

where $k_1\left[u\left(x_1, x_2, \ldots, x_n\right); \theta\right]=b^{-n} \exp \left(\frac{1}{b} \sum_{i=1}^n(x_i-a)\right) I(x_{(1)}>a)$ and $k_2\left(x_1, x_2, \ldots, x_n\right)=1$. Hence by factorization theorem $u(X_1, \ldots, X_n)$ is a sufficient statistic for θ.