6262 HOMEWORK 1

1. Preliminary Results

1.1. **Some basic facts.** If we have some set $A \subset \Omega$, we define the indicator function a_A by

$$\mathbb{1}_A(\omega) = \begin{cases} a & \omega \in A \\ 0, & \omega \notin A. \end{cases}$$

For instance if $\Omega = \mathbb{R}$, the real line, we have that $\mathbb{1}_{[a,b]}(x) = 1$ if and only if $x \in [a, b]$. Notice that this is also equal to $\mathbb{1}_{(-\infty,b]}\mathbb{1}_{[a,\infty)}$.

If X is a random variable, the cdf is given by

$$F_X(x) = \mathbb{P}(X \leq x).$$

This contains most of the properties of the individual random variable.

If X is continuous, then the pdf (the probability density function) is given by

$$f_X(x) = F_X'(x) \tag{1.1}$$

at "almost all points". By almost all points we admit that there might be some points on the real line where F_X is not differentiable, though this is negligible from the point of view of measure theory (i.e. if of Lebegue measure 0).

If X is a random variable, then for any function $\phi : \mathbb{R} \to \mathbb{R}$, we have

$$\mathbb{E}[\phi(X)] = \begin{cases} \sum_x \phi(x)p_X(x), & X \text{ discrete with pmf } p_X \\ \int \phi(x)f_X(x)dx, & \text{if X is continuous with density } f_X. \end{cases} \tag{1.2}$$

For two random variables (X, Y), then for a function in two variables

$$\mathbb{E}[\phi(X, Y)] = \begin{cases} \sum_{x,y} xy\mathbb{P}_{X,Y}(x, y), & (X, Y) \text{ discrete with joint pmf } p_{X,Y} \\ \int xyf_{X,Y}(x, y)dxdy, & (X, Y) \text{ is continuous with joint density } f_{X,Y}. \end{cases} \tag{1.3}$$

1.2. **Independence of Normal variables.** If (X, Y) is a two dimensional normal distribution, then X and Y are independent if and only if the correlation between them is 0. In other words, X and Y are independent iff $Cov(X, Y) = 0$, or alternatively, $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$.

In addition, if (X, Y) is a normal vector, then $(aX + bY, cX + dY)$ is also a normal vector.

1.3. **Conditional Expectation.** Given two random variables X and Y which are square integrable, the definition of the conditional expectation of X given Y is interpreted as the function $\phi(Y)$ such that

$$\mathbb{E}[(X - \phi(Y))^2] \tag{1.4}$$

is minimized over all possible choices of the function ϕ such that $\phi(Y)$ remains square integrable. We write $\mathbb{E}[X|Y] = \phi(Y)$ or in more statistical slang $\mathbb{E}[X|Y = y] = \phi(y)$, though this could be a bit confusing.

We saw in class that the characterization of $\phi(Y)$ is given by the following equality

$$\mathbb{E}[X\psi(Y)] = \mathbb{E}[\phi(Y)\psi(Y)] \tag{1.5}$$

for any other choice of the function ψ.

As properties of the conditional expectation show the following.

1.4. **Unbiased estimators and MVUE.** For a family of distributions $f(x;\theta)$, a statistics $T = u(X_1, X_2, \dots, X_n)$ based on a n samples $X_1, X_2, \dots, X_n$ is called an unbiased estimator of θ if

$$\mathbb{E}_\theta[T] = \theta$$

for any choice of θ. We say that T is an MVUE if for any other unbiased estimator U,

$$Var(T) \le Var(U) \text{ for all } \theta.$$

In other words, an unbiased estimator predicts correctly the parameter θ in average. The condition of MVUE, gives the unbiased estimator with the smallest possible error in the estimation of θ, where here by error we really mean the L^2 norm of the difference $T - \theta$.

1.5. **Sufficient statistic and basic results.**

Definition 1. *Given $f(x;\theta)$ a family of densities, and a sample $X_1, X_2, \dots, X_n$ a sample, a statistic $T = u(X_1, X_2, \dots, X_n)$ is called* sufficient *if*

$$\frac{f(x_1;\theta)f(x_2;\theta)\dots f(x_n;\theta)}{f_T(u(x_1,x_2,\dots,x_n);\theta)} = H(x_1,x_2,\dots,x_n). \tag{1.6}$$

This is the definition from Hogg's book. This is a little confusing if both terms in the ratio on the left are zero. The alternative form of (1.6)

$$f(x_1;\theta)f(x_2;\theta)\dots f(x_n;\theta) = f_T(u(x_1,x_2,\dots,x_n);\theta)H(x_1,x_2,\dots,x_n). \tag{1.7}$$

Notice that in this form, H is a function which does not depend on θ.

Theorem 2 (Factorization Theorem). *$T = u(x_1, x_2, \dots, x_n)$ is a sufficient statistics if and only if*

$$f(x_1;\theta)f(x_2;\theta)\dots f(x_n;\theta) = k_1(u(x_1,x_2,\dots,x_n);\theta)k_2(x_1,x_2,\dots,x_n),$$

for some function $k_1(t;\theta)$ and $k_2(x_1,x_2,\dots,x_n)$, where k_2 does not depend on θ.

Theorem 3 (Rao-Blackwell). *If T is a sufficient statistic and Y is an unbiased estimator which is not a function of T, then $U = \mathbb{E}[Y|T] = \phi(T)$ is also an unbiased estimator with*

$$Var(U) < Var(T).$$

In particular, Y is not an MVUE.

The notion of complete statistic is the following.

Definition 4. *We say that a statistic T is complete if $\mathbb{E}_\theta[\phi(T)] = 0$ for all θ implies that $\phi(T) \equiv 0$.*

In other words, $\phi(T)$ is completely determined by the expectation of $\mathbb{E}_\theta[\phi(T)]$. To see this, if we have two functions ϕ_1 and ϕ_2 with the same expectations, then by taking the difference and using the above definition leads to the conclusion that $\phi_1(T) = \phi_2(T)$, which means essentially identifiability.

Theorem 5 (Lehman-Schaffé). *If T is a sufficient and complete such that for some function ϕ, $\phi(T)$ is unbiased, then this is the only unbiased estimator of this form (i.e., function of T and unbiased). In particular this is actually the MVUE.*

2. Exponential family

A family of pmfs/pdfs $f(x:\theta)$ is called a regular exponential family if

$$f(x;\theta) = e^{p(\theta)K(x)+S(x)+q(\theta)} \text{ for } x \in \mathcal{S} \tag{2.1}$$

and here $\mathcal{S}$ does not depend on the parameter θ and the function K is not constant on the set $\mathcal{S}$.

Theorem 6. *If $f(x;\theta)$ is a regular exponential family and $X_1, X_2, \dots, X_n$ is a sample from $f(x;\theta)$ then*

$$T = \sum_{k=1}^{n} K(X_k)$$

is a sufficient and complete statistic. In addition, if we can find a function ϕ, such that $\phi(T)$ is unbiased, then $\phi(T)$ is the unique MVUE.

3. Homework + Solutions

Problem 1. (1) *If $Y = 0$, then $E[X|Y] = \mathbb{E}[X]$. Interpret this.*
(2) *The function ϕ may not be unique, however $\phi(Y)$ is uniquely defined.*
(3) *If X and Y are independent, then $\mathbb{E}[X|Y] = \mathbb{E}[X]$. Interpret this.*
(4) *If $Z = h(Y)$ for another function h, then $\mathbb{E}[ZX|Y] = h(Y)\phi(Y)$. Written alternatively, $\mathbb{E}[ZX|Y] = Z\mathbb{E}[X|Y]$.*
(5) *If $X = g(Y)$, then $\mathbb{E}[X|Y] = g(Y)$.*
(6) *Argue that if $\mathbb{E}[X|Y] = \phi(Y)$ and $\mathbb{E}[Z|Y] = \psi(Y)$, then $\mathbb{E}[X+Z|Y] = \phi(Y) + \psi(Y)$.*

Solution. (1) The functions of the variable Y is simply a real number, this means that we need to look for the best number μ which minimizez the quadratic loss function

$$\mathbb{E}[(X-\mu)^2].$$

We did in class that this is the mean. In fact it is easy to see by simply differentiating with respect to μ and setting this to be equal to 0.

(2) The uniqueness of $\phi(Y)$ is based on the fact that if we have two such representations, $\mathbb{E}[X|Y] = \phi_1(Y)$ and $\mathbb{E}[X|Y] = \phi_2(Y)$ then by property (1.5), then

$$\mathbb{E}[X\psi(Y)] = \mathbb{E}[\phi_1(Y)\psi(Y)] = \mathbb{E}[\phi_2(Y)\psi(Y)]$$

for any choice of ψ. This implies in turn that

$$\mathbb{E}[(\phi_1(Y) - \phi_2(Y))\psi(Y)] = 0$$

for all choices of ψ, which mean that $\phi_1(Y) = \phi_2(Y)$.

The non-uniqueness of ϕ_1 and ϕ_2 are given by the fact that for instance $Y \geq 0$, then we can take for instance $\phi_1(x) = x$ and $\phi_2(x) = x^+ = \max(x, 0)$, are two different functions with the property that they coincide on the positive real line but are not equal on the whole real line.

(3) If X is independent of Y, then the conditioning with respect to Y should behave like conditioning on constant random variables. Formally we have that if $\mathbb{E}[X|Y] = \phi(Y)$, then by (1.5)

$$\mathbb{E}[X\psi(Y)] = \mathbb{E}[\phi(Y)\psi(Y)]$$

now on the other hand, because X and Y are independent,

$$\mathbb{E}[X\psi(Y)] = \mathbb{E}[X]\mathbb{E}[\psi(Y)] = \mathbb{E}[\mathbb{E}[X]\psi(Y)]$$

thus the constant $\mathbb{E}[X]$ plays the role of the function of $\phi(Y)$ in (1.5) for $\mathbb{E}[X|Y]$.

(4) This follows from (1.5) because
$$\mathbb{E}[XZ\psi(Y)] = \mathbb{E}[Xh(Y)\psi(Y)] = \mathbb{E}[\phi(Y)h(Y)\psi(Y)]$$
we used again (1.5) with $\psi(Y)$ replaced by $h(Y)\psi(Y)$. This means that we can now argue that $\phi(Y)h(Y) = \mathbb{E}[ZX|Y]$ again by utilizing (1.5).

(5) This is easy because using the definition from (1.4), the choice of $\phi(Y) = g(Y)$ gives $\mathbb{E}[(X-\phi(Y))^2] = 0$ which is the minimum possible value. This means that $\mathbb{E}[g(Y)|Y] = g(Y)$.

(6) This follows again using the property (1.5) we can write
$$\mathbb{E}[X\zeta(Y)] + \mathbb{E}[Z\zeta(Y)] = \mathbb{E}[\phi(Y)\zeta(Y)] + \mathbb{E}[\psi(Y)\zeta(Y)] = \mathbb{E}[(\phi(Y)+\psi(Y))\zeta(Y)]$$
for any choice of $\zeta(Y)$. In particular again by the characterization from (1.5) applied to $X+Z$ and $\phi(Y)+\psi(Y)$ gives that $\mathbb{E}[X+Z|Y] = \phi(Y)+\psi(Y)$.

□

Problem 2. *If the pair (X,Y) has a joint pmf or pdf, show that*

$$\mathbb{E}[X|Y=y] = \phi(y) \text{ where } \phi(y) = \begin{cases} \sum_x x\frac{p_{X,Y}(x,y)}{p_Y(y)} & (X,Y) \text{ are discrete with pmf } p_{X,Y} \\ \int x\frac{f_{X,Y}(x,y)}{f_Y(y)}dx & (X,Y) \text{ have joint pdf } f_{X,Y}. \end{cases} \tag{3.1}$$

Solution. This is pretty straightforward. We only need to show that for any choice of the function ψ, we have from (1.3) that
$$\mathbb{E}[X\psi(Y)] = \sum_{x,y} x\psi(y)p_{X,Y}(x,y) = \sum_y \psi(y)\sum_x x\frac{p_{X,Y}(x,y)}{p_Y(y)}p_Y(y) = \sum_y \psi(y)\phi(y)p_Y(y) = \mathbb{E}[\phi(Y)\psi(Y)]$$
where in the last equation we used (1.2) for the variable Y and the function $\phi(y)\psi(y)$.

A similar proof works for the case of continuous distributions, just replacing the summation with the integration. □

Problem 3. *Compute the following conditional expectations:*

(1) $\mathbb{E}[X|X^2]$ *if* $X \sim Exp(\lambda)$.
(2) $\mathbb{E}[X|X^3]$ *if* $X \sim N(0,2)$.
(3) $\mathbb{E}[X-2X^4|X^2]$ *for* $X \sim N(0,2)$.
(4) *Give an example of two different functions ϕ_1 and ϕ_2 such that $\mathbb{E}[X|X^2] = \phi_1(X^2)$ and also $\mathbb{E}[X|X^2] = \phi_2(X^2)$. Are $\phi_1(X^2)$ and $\phi_2(X^2)$ equal?*
(5) $\mathbb{E}[\cos(X)|X^2]$ *if* $X \sim N(0,1)$.
(6) $\mathbb{E}[X|Y]$ *if X, Y are iid* $N(0,1)$.
(7) $\mathbb{E}[X+X^2|X^4]$ *for* $X \sim N(0,1)$.
(8) $\mathbb{E}[X|X+2Y]$ *if X, Y are iid* $N(0,1)$.

Solution. (1) Using the fact that $X \geq 0$, we actually have that $X = \sqrt{X^2}$ and thus using the first problem, part 5) we get that $\mathbb{E}[X|X^2] = X$.

(2) Similar to the first part, we have that $X = (X^3)^{1/3}$, thus $\mathbb{E}[X|X^3] = X$.

(3) Using the first problem, the additivity from part 6) we get that
$$\mathbb{E}[X-2X^4|X^2] = \mathbb{E}[X|X^2] - 2\mathbb{E}[X^4|X^2] = -2X^4$$
because we showed in class that $\mathbb{E}[X|X^2] = 0$. This can be redone here in a different way, namely observing that we can write $X = Z|X|$ where Z is a Bernoulli ± 1 with equal probabilities. In addition Z and $|X|$ can be assumed independent. Therefore using a combination of the first problem, more precisely, the third and the fifth, we get that $\mathbb{E}[X|X^2] = \mathbb{E}[Z|X||X^2] = |X|\mathbb{E}[Z|X^2] = |X|\mathbb{E}[Z] = 0$.

(4) This is basically the comment from problem 1, part 2. Take for instance X to be exponential and we have $\mathbb{E}[X|X^2] = X = \phi_1(X)$ with $\phi_1(x) = x$ on one hand, and also $\mathbb{E}[X|X^2] = \phi_2(X)$ with $\phi_2(x) = \max(x, 0)$.

(5) The key is that $\cos(x) = \cos(|x|) = \cos(\sqrt{x^2})$, thus

$$\mathbb{E}[\cos(X)|X^2] = \mathbb{E}[\cos(\sqrt{X^2})|X^2] = \cos(\sqrt{X^2}) = \cos(|X|) = \cos(X)$$

where we used the first problem, part 5) to write $\cos(X)$ as a function of X^2.

(6) If X, Y are independent, we get everything from Problem 1, part 3). Thus $\mathbb{E}[X|Y] = \mathbb{E}[X] = 0$.

(7) $\mathbb{E}[X + X^2|X^4] = \mathbb{E}[X|X^4] + \mathbb{E}[X^2|X^4] = X^2$, because $\mathbb{E}[X|X^4] = 0$, with an argument like we did in class for $\mathbb{E}[X|X^2]$ or just above.

(8) This is more complicated. There are two ways of doing it. One is to use Problem 2 and find the joint density of $(X, X + 2Y)$ and then use that formula. The other is based on writing X as a sum of a function of $X + 2Y$ and another piece which will be independent of $X + 2Y$. The easier and faster way is the latter.

We can use the facts outlined at the beginning of this section to write

$$X = a(X + 2Y) + b(2X - Y)$$

and try to figure out the constants a, b such this writing is valid. Equating the coefficient of X and Y gives $a + 2b = 1$ and $2a - b = 0$, or $b = 2a$ and $5a = 1$, which means that $a = 1/5$ while $b = 2/5$. The key is now that

$$(X + 2Y, 2X - Y)$$

is a normal vector with independent components because the covariance $Cov(X+2Y, 2X-Y) = 2Cov(X, X) + 3Cov(Y, X) - 2Cov(Y, Y) = 2(Var(X) - Var(Y)) = 0$. Now, using the first Problem 1, we get

$$\mathbb{E}[X|X+2Y] = \mathbb{E}[(X+2Y)/5|X+2Y]+(2/5)\mathbb{E}[(2X-Y)|X+2Y] = (X+2Y)/5+\mathbb{E}[2X-Y] = (X+2Y)/5.$$

□

Problem 4. (1) *If (X, Y) are iid uniform on $(0, 1)$, find $\mathbb{E}[X|X + Y]$. Can you explain?*

(2) *If (X, Y) are uniform on $0 < x < y < 1$, find $\mathbb{E}[X|Y]$. Interpret this as the joint random variable is uniform on the 2 dimensional domain determined by $\{(x, y) : 0 < x < 1, 0 < y < 1, x < y\}$.*

(3) *Assume (X, Y) take the values $(0, 1), (2, 3), (3, 4), (2, 4), (4, 1)$ with equal probability. Compute $\mathbb{E}[X|Y]$.*

Solution. (1) The key here is the symmetry between X and Y. Thus we can use the second problem to argue that $\mathbb{E}[X|X + Y] = \phi(X + Y)$ and $\mathbb{E}[Y|X + Y] = \phi(X + Y)$. Now adding the two together, we obtain $\mathbb{E}[X+Y|X+Y] = 2\phi(X+Y)$. Therefore, $\phi(X+Y) = (X + Y)/2$, so $\mathbb{E}[X|X + Y] = (X + Y)/2$.

Another solution would be to compute the joint density of $(X, X + Y)$ and then use the second Problem. This is possible, however is very long because we have to also use a change of variables in two dimensions.

(2) For this part we really have to use Problem 2. In the first place, the meaning of the problem is that the point in the plane, (X, Y) moves inside the region $0 < x < y < 1$ which is just the triangle in the unit square below the diagonal. Thus, we have

$$f_{X,Y}(x, y) = 2 \text{ on the domain } 0 < x < y < 1.$$

This is because being uniform implies that the density is constant, say c. This constant integrated over the whole domain must be equal to 1, thus we get $c \times (1/2) = 1$, which gives $c = 2$. Therefore,

$$f_Y(y) = \int f_{X,Y}(x,y)dx = 2y \text{ for } 0 < y < 1.$$

Thus

$$\phi(y) = \int x \frac{f_{X,Y}(x,y)}{f_Y(y)} dx = \int_0^y x/y dx = y/2$$

and thus $\mathbb{E}[X|Y] = Y/2$.

(3) For this we again can use the second Problem to compute the function ϕ. We get that Y has marginal given by $1, 3, 4$ with probabilities $2/5$, $1/5$ and $2/5$. What we need to find now is

$$\phi(y) = \sum_x x p_{X,Y}(x,y)/p_Y(y) = \begin{cases} 2 & y = 1 \\ 2 & y = 3 \\ 5/2 & y = 4 \end{cases} \tag{3.2}$$

and $\mathbb{E}[X|Y] = \phi(Y)$.

□

Remark 7. *There is a point in the last part of this problem. We computed the conditional expectation with the formula from (3.1). However the point is that the variable Y is discrete and thus can be written in the form*

$$Y = \sum_k \alpha_k \mathbb{1}_{A_k} \tag{3.3}$$

for some partition $\{A_k\}_k$ of the sample space. Here, for an event A (a subset in the sample space), we set

$$\mathbb{1}_A(\omega) = \begin{cases} 1 & \omega \in A \\ 0, & \textit{otherwise}\ . \end{cases}$$

In our case at hand we have

$$Y = \mathbb{1}_{\{Y=1\}} + 3\mathbb{1}_{\{Y=3\}} + 4\mathbb{1}_{\{Y=4\}}$$

Notice that if we write everything explicitly, we would have that the sample space where both X and Y are defined is given by $\Omega = \{(0,1),(2,3),(3,4),(2,4),(4,1)\}$ with the uniform probability on it. Thus the subsets $\{Y = 1\} = \{(0,1),(4,1)\}$, $\{Y = 3\} = \{((2,3))\}$ and $\{Y = 4\} = \{(3,4),(2,4)\}$. The point is that now a variable of the form $\phi(Y)$ is perfectly defined by prescribing the values $\phi(1), \phi(3)$ and $\phi(4)$. Therefore we would have the following writing

$$\phi(Y) = \phi(1)\mathbb{1}_{\{Y=1\}} + \phi(3)\mathbb{1}_{\{Y=3\}} + \phi(4)\mathbb{1}_{\{Y=4\}}.$$

This explains why the function above, in (3.2) is defined in this way. The conditional expectation is thus perfectly well defined in terms of the function Y.

As a more general lesson, we have the following conclusion. If Y is discrete and written in the form (3.3), we then have

$$\phi(Y) = \sum_k \phi(\alpha_k)\mathbb{1}_{A_k}.$$

This explains why the function ϕ has to be specified only at the values α_k and it does not really matter how is defined at any other points.

This situation appears also in Problem 6 below.

Problem 5. *If (X,Y) have joint pdf given by $f_{X,Y}(x,y) = 12(2x + y^2)/7$ on the set $0 < x < y < 1$ and 0 otherwise, find $\mathbb{E}[X|Y]$.*

Solution. Here we use Problem 2. The function we need to compute is

$$\phi(y) = \int_0^y x f_{X,Y}(x,y)/f_Y(y)dx$$

for $0 < y < 1$.

For the computation of the density of Y we need to use the fact that

$$f_Y(y) = \int_0^y f_{X,Y}(x,y)dx = \frac{12}{7}(y^2 + y^3) = \frac{12}{7}y^2(y+1).$$

On the other hand the other integral

$$\int_0^y x f_{X,Y}(x,y)dx = \frac{2}{7}(3y^4 + 4y^3)$$

Thus we get

$$\phi(y) = \frac{\int_0^y x f_{X,Y} dx}{f_Y(y)} = \frac{y(3y+4)}{6(y+1)}.$$

□

Problem 6. *Flip a coin until the head comes up and let X be the number of flips. Compute $\mathbb{E}[X|\cos(\pi X/2)]$.*

Solution. First we need to notice that $\cos(\pi X/2)$ takes only values $-1, 0, 1$. Thus we can use the comments in the Remark 7 for more explanation. Our sample space is $\Omega = \{1, 2, 3, \dots, \}$. In fact we can write this as

$$\cos(\pi y/2) = 0 \times \mathbb{1}_A(y) + (-1) \times \mathbb{1}_B(y) + 1 \times \mathbb{1}_C(y)$$

where $A = \{1, 3, 5, 7, 9, 11, 13, \dots\}$, $B = \{2, 6, 10, \dots\}$ and $C = \{4, 8, 12, \dots\}$ and $\mathbb{1}_A(y) = 1$ if $y \in A$ and 0 otherwise. Notice that the sets A, B, C form a partition of the whole space.

Thus we need to compute the conditional expectation $\mathbb{E}[X|\cos(\pi X/2)] = \phi(Y)$ where ϕ has to be specified at three values, namely, $0, -1, 1$. Thus it suffices to study the value at the points $-1, 0, 1$ because

$$\phi(Y) = \phi(0)\mathbb{1}_A + \phi(-1)\mathbb{1}_B + \phi(1)\mathbb{1}_C.$$

Now from the definition (1.5), we need to check that

$$\mathbb{E}[X\psi(\cos(\pi X/2))] = \mathbb{E}[\phi(Y)\psi(Y)].$$

for any function ψ. Taking $\psi(y) = \mathbb{1}_A(y)$, and noticing that the right hand side becomes $\phi(0)\mathbb{P}(A)$, we obtain the value of $\phi(0)$ which is in fact given by

$$\phi(0)\mathbb{P}(A) = \mathbb{E}[X\psi(\cos(\pi X/2))] = \mathbb{E}[X\mathbb{1}_A(y)] = \sum_{x \in A} x\mathbb{P}(X = x) = 1/2 + 3/2^3 + 5/2^5 + 7/2^7 + 9/2^9 + \dots$$

$$= \sum_{k=0}^{\infty} (2k+1)/2^{2k+1} = 10/9.$$

Since, $\mathbb{P}(A) = 1/2 + 1/2^3 + 1/2^5 + \dots = 2/3$, we obtain

$$\phi(0) = \frac{5}{3}.$$

In a similar fashion, for $\psi(y) = \mathbb{1}_B(y)$, we get

$$\phi(1)\mathbb{P}(B) = \mathbb{E}[X\psi(\cos(\pi X/2))] = \mathbb{E}[X\mathbb{1}_B(y)] = \sum_{x \in B} x\mathbb{P}(X = x) = 2/2^2 + 6/2^6 + 10/2^{10} + \dots$$

$$= \sum_{k=0}^{\infty} (4k+2)/2^{4k+2} = 136/225.$$

Since $\mathbb{P}(B) = 1/2^2 + 1/2^6 + 1/2^10 + \cdots = 4/15$, then we get that

$$\phi(1) = 34/15.$$

Finally, for $\phi = \mathbb{1}_C(y)$ we finally get

$$\phi(1)\mathbb{P}(C) = \mathbb{E}[X\psi(\cos(\pi X/2))] = \mathbb{E}[X\mathbb{1}_C(y)] = \sum_{x\in B} x\mathbb{P}(X=x) = 4/2^4 + 8/2^8 + 12/2^12 + \dots$$
$$= \sum_{k=1}^{\infty}(4k)/2^{4k} = 64/225$$

and $\mathbb{P}(C) = 1/2^4 + 1/2^8 + \cdots = 1/15$, from which we get

$$\phi(1) = 64/15.$$

Thus,

$$\mathbb{E}[X|\cos(\pi X/2)] = \frac{5}{3}\mathbb{1}_A(Y) + \frac{34}{15}\mathbb{1}_B(Y) + \frac{64}{15}\mathbb{1}_C(Y).$$

Thus if we define the function

$$\phi(y) = \begin{cases} 5/3 & y=0 \\ 34/15 & y=-1 \\ 64/15 & y=1 \end{cases}$$

Then we can write

$$\mathbb{E}[X|\cos(\pi X/2)] = \phi(\cos(\pi X/2)).$$

Notice that we used here the identity

$$\sum_{k=0}^{\infty} k\rho^k = \rho/(1-\rho)^2$$

to compute the various sums above.

□

Problem 7. *Let $\mathbb{E}[X|Y] = \phi(Y)$ and using (1.5) show that $\mathbb{E}[X] = \mathbb{E}[\phi(Y)]$ and*

$$Var(X) = Var(\phi(Y)) + \mathbb{E}[(X-\phi(Y))^2]. \tag{3.4}$$

In particular argue that $Var(X) \geq Var(\phi(Y))$ with equality if and only if $X = \phi(Y)$ or that X is a function of Y.

Solution. We worked the idea in class. Denote $\mathbb{E}[X|Y] = \phi(Y)$ and notice in the first palce that $\mathbb{E}[\phi(Y)] = \mathbb{E}[X]$ as one can see from (1.5) for $\psi(y) = 1$ Here are some details. Write

$$\begin{aligned} Var(X) &= \mathbb{E}[(X-\mathbb{E}[X])^2] = \mathbb{E}[(X-\phi(Y)+\phi(Y)-\mathbb{E}[X])^2] \\ &= \mathbb{E}[(X-\phi(Y))^2] + \mathbb{E}[(\phi(Y)-\mathbb{E}[X])^2] + 2\mathbb{E}[(X-\phi(Y))(\phi(Y)-\mathbb{E}[X])] \end{aligned} \tag{3.5}$$

Now, using (1.5) with $\psi(y) = \phi(y) - \mathbb{E}[X]$, we get that $\mathbb{E}[X\psi(Y)] = \mathbb{E}[\phi(Y)\psi(Y)]$ and this means that $\mathbb{E}[(X-\phi(Y))\psi(Y)] = 0$, thus the last term in (3.5) is 0.

To finish the rest, we notice that $\mathbb{E}[(\phi(Y)-\mathbb{E}[X])^2] = \mathbb{E}[(\phi(Y)-\mathbb{E}[\phi(Y)])^2] = Var(\phi(Y))$. Putting all together, we get the relation (3.4).

□

Problem 8. *Show that if $X_1, X_2, \dots, X_n$ is a sample and $T = u(x_1, x_2, \dots, x_n)$ is a sufficient statistic, then for any one-to-one map $g : \mathbb{R} \to \mathbb{R}$, $\bar{T} = g(T)$ is also a sufficient statistic.*

Solution. We discussed this in class, however here are some details. This is based on the factorization Theorem 3 and the writing

$$\begin{aligned} f(x_1;\theta)f(x_2;\theta)\dots f(x_n;\theta) &= k_1(u(x_1,x_2,x\dots,x_n);\theta)k_2(x_1,x_2,\dots,x_n) \\ &= k_1(g^{-1}(g(u(x_1,x_2,x\dots,x_n)));\theta)k_2(x_1,x_2,\dots,x_n) \\ &= \tilde{k}_1(g(u(x_1,x_2,x\dots,x_n));\theta)k_2(x_1,x_2,\dots,x_n) \end{aligned}$$

where $\tilde{k}_1(t;\theta) = k_1(g^{-1}(t);\theta)$. This shows, again from the factorization theorem that $\bar{T}$ is also a sufficient statistic.

□

Problem 9. *Assume $X_1, X_2, \dots, X_n$ is a sample with density*

$$f(x;\theta) = \begin{cases} \theta x^{\theta-1}, & 0<x<1 \\ 0 & \textit{otherwise}. \end{cases}$$

(1) *Show that $T = \prod_{k=1}^n X_k$ is a sufficient statistic. Also show that $T = \sum_{k=1}^n \ln(X_k)$ is also a sufficient statistic.*
(2) *Show that $U = \sum_{k=1}^n X_i$ IS NOT a sufficient statistic.*

Solution. (1) The fastest way of seeing this is to realize that $f(x;\theta)$ is a regular exponential family because we can write it as

$$f(x;\theta) = \begin{cases} e^{(\theta-1)\ln(x)+\ln(\theta)} & 0<x<1 \\ 0, & \text{otherwise}, \end{cases}$$

thus it is of the form (2.1) with $\mathcal{S} = (0,1)$, $p(\theta) = \theta - 1$, $K(x) = \ln(x)$, $S(x) = 0$ and $q(\theta) = \ln(\theta)$. This implies that $V = \sum_{k=1}^n \ln(X_k)$ is a sufficient and complete statistic. Now combining this with Problem 8 with the function $g(x) = e^x$, gives that $T = g(V) = e^{\sum_{k=1}^n \ln(X_k)} = e^{\ln(\prod_{k=1}^n X_k)} = \prod_{k=1}^n X_k$ is also a sufficient statistic. This answers the first part of the problem.

There is a second argument to the fact that T is a sufficient statistic and this is based on the factorization theorem 2. Essentially we have that

$$f(x_1;\theta)f(x_2;\theta)\dots f(x_n;\theta) = \theta^n \left(\prod_{k=1}^n x_k\right)^{\theta-1}$$

which means with $k_1(y;\theta) = \theta^n y^{\theta-1}$ for $0<y<1$ and $k_2(x_1,\dots,x_n) = 1$, we have the conclusion that T is a sufficient statistic. To show that $\bar{T}$ is a sufficient statistic, we just need to use Problem 8

(2) To show that U is not a sufficient statistic we need to show that we can not use for instance the factorization theorem. We assume by contradiction that U is a sufficient statistic and thus

$$f(x_1;\theta)f(x_2;\theta)\dots f(x_n;\theta) = \theta^n \left(\prod_{k=1}^n x_k\right)^{\theta-1} = k_1(\sum_{k=1}^n x_k;\theta)k_2(x_1,x_2,\dots,\theta_n)$$

for all choices of $x_1, x_2, \dots, x_n \in (0,1)$ and $\theta > 0$. This is hard to believe at first site, because on the left hand side we have a function of the product, while on the other hand we have a function of the sum and θ. To show though that this is not possible, we can argue with the fact that

$$k_2(x_1, x_2, \dots, x_n) = \frac{f(x_1;\theta)f(x_2;\theta)\dots f(x_n;\theta)}{k_1(\sum_{k=1}^n x_k;\theta)}$$

Since the right hand side does not depend on θ, we can pick one as a reference, say for instance $\theta = 1$. Thus we would have that

$$\frac{f(x_1;\theta)f(x_2;\theta)\dots f(x_n;\theta)}{k_1(\sum_{k=1}^n x_k;\theta)} = \frac{f(x_1;\theta)f(x_2;1)\dots f(x_n;1)}{k_1(\sum_{k=1}^n x_k;1)} = \frac{1}{k_1(\sum_{k=1}^n x_k;1)}.$$

In particular this means that

$$\theta^n \left(\prod_{k=1}^n x_k\right)^{\theta-1} = \frac{k_1(\sum_{k=1}^n x_k;\theta)}{k_1(\sum_{k=1}^n x_k;1)}.$$

At this point, θ no longer play any role, the game here is between the sum and the products of x's. For instance, if we take $\theta = 2$, we obtain that

$$2^n x_1 x_2 \dots x_n = \frac{k_1(\sum_{k=1}^n x_k;2)}{k_1(\sum_{k=1}^n x_k;1)} \text{ for all } 0 < x_1, x_2, \dots, x_n < 1.$$

This is contradictory because for instance if we take $x_1 = x_2 = \cdots = x_n = 1/n$, the sum is 1 and the product is $1/n^n$. However if we change this slightly, by taking $x_1 = x_2 = \cdots = x_{n-2} = 1/n$ and $x_{n-1} = 1/n - \epsilon$ and $x_n = 1/n + \epsilon$, the sum will be the same, however the product becomes something different, namely, $(1 - n^2\epsilon^2)/n^n$ which gives different values for different values of ϵ.

□

Problem 10. *Let $X_1, X_2, \dots, X_n$ be a sample from a uniform distribution on $[0,\theta]$. Show that $T = \max\{X_1, X_2, \dots, X_n\}$ is a sufficient statistic. Do this using the definition and also using the factorization theorem.*

Is this a complete statistic? Why or why not?

Solution. Before we jump into the details of the solution, notice that we can rewrite

$$f(x;\theta) = \frac{1}{\theta}\mathbb{1}_{[0,\theta]}(x) = \frac{1}{\theta}\mathbb{1}_{[0,\infty)}(x)\mathbb{1}_{(-\infty,\theta]}(x).$$

The first approach is using the definition. In the first place we need to find the density of the T. This is usually done by first computing the cdf (the cumulative distribution function) and then taking the derivative as guaranteed by (1.1).

Now, take $x \in (0,\theta)$. We have

$$\begin{aligned}\mathbb{P}(T \le x) &= \mathbb{P}(\max\{X_1, X_2, \dots, X_n\} \le x) = \mathbb{P}(X_1 \le x, X_2 \le x, \dots, X_n \le x) \\ &= \mathbb{P}(X_1 \le x)\mathbb{P}(X_2 \le x)\dots\mathbb{P}(X_n \le x) = (x/\theta)^n.\end{aligned}$$

Of course we have that $F_T(x) = 0$ if $x \le 0$ and $F_T(x) = 1$ for $x > \theta$. Taking the derivative, we get that

$$f_T(x) = n\frac{x^{n-1}}{\theta^n} \text{ for } 0 < x < \theta \text{ and } 0 \text{ otherwise.} \tag{3.6}$$

which we can write as

$$f_T(x;\theta) = \frac{n}{\theta^n}x^{n-1}\mathbb{1}_{(-\infty,\theta]}(x)\mathbb{1}_{[0,\infty)}(x) \tag{3.7}$$

For any $x_1, x_2, \ldots, x_n \in (0, \theta)$,

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

The left hand side becomes 0 if for instance one $x_k < 0$. The left hand side is also 0 if one of the x's becomes $> \theta$. In this case what seems to be happening is that the ratio $\frac{f(x_1;\theta)f(x_2;\theta)\ldots f(x_n;\theta)}{f_T(\max\{x_1,x_2,\ldots,x_n\};\theta)}$ in fact depends on θ. This is because the ratio is ill defined for some values of x's. This is why I added a clarification in the Definition 1 between the equation (1.6) and (1.7). If we write according to (1.7), then we have a better form

(3.8)

$$\begin{aligned}
f(x_1;\theta)f(x_2;\theta)\ldots f(x_n;\theta) &= \mathbb{1}_{[0,\theta]}(x_1)\mathbb{1}_{[0,\theta]}(x_2)\ldots\mathbb{1}_{[0,\theta]}(x_n)\\
&= \frac{1}{\theta^n}\mathbb{1}_{(-\infty,\theta]}(x_1)\mathbb{1}_{(-\infty,\theta]}(x_2)\ldots\mathbb{1}_{(-\infty,\theta]}(x_n)\mathbb{1}_{[0,\infty)}(x_1)\mathbb{1}_{[0,\infty)}(x_2)\ldots\mathbb{1}_{[0,\infty)}(x_n)\\
&\overset{(*)}{=} \frac{1}{\theta^n}\mathbb{1}_{(-\infty,\theta]}(\max\{x_1,x_2,\ldots,x_n\})\mathbb{1}_{[0,\infty)}(\min\{x_1,x_2,\ldots,x_n\})\\
&\overset{(**)}{=} f_T(\max\{x_1,x_2,\ldots,x_n\};\theta)\frac{\mathbb{1}_{[0,\infty)}(\min\{x_1,x_2,\ldots,x_n\})}{n\max\{x_1,x_2,\ldots,x_n\}^{n-1}}.
\end{aligned}$$

where the equality (*) is justified by the fact that all x_k are $\le \theta$ if and only if $\max\{x_1, x_2, \ldots, x_n\} \le \theta$ while all x_k are non-negative if and only if $\min\{x_1, x_2, \ldots, x_n\}$ is non-negative. Also equality (**) is justified by (3.7). Thus we can take $H(x_1, x_2, \ldots, x_n) = \frac{\mathbb{1}_{[0,\infty)}(\min\{x_1,x_2,\ldots,x_n\})}{n\max\{x_1,x_2,\ldots,x_n\}^{n-1}}$ which now it does not depend at all on θ.

The second argument uses the factorization theorem as follows and the sequence of equalities in (3.9) only up until (including) equality (*). In other words, we write

$$\begin{aligned}
f(x_1;\theta)f(x_2;\theta)\ldots f(x_n;\theta) &= \frac{1}{\theta^n}\mathbb{1}_{(-\infty,\theta]}(\max\{x_1,x_2,\ldots,x_n\})\mathbb{1}_{[0,\infty)}(\min\{x_1,x_2,\ldots,x_n\})\\
&= k_1(\max\{x_1,x_2,\ldots,x_n\};\theta)k_2(x_1,x_2,\ldots,x_n)
\end{aligned}$$

and from factorization Theorem 2, we get the conclusion that T is a sufficient statistic.

To check completeness we need to check that if $\mathbb{E}[\phi(T)] = 0$ for any θ, then $\phi(T)$ is identically 0. One way of doing this is to write using (1.2) and (3.6) that

$$\mathbb{E}[\phi(T)] = \int_0^\theta \phi(x)nx^{n-1}dx/\theta^n = 0$$

for any θ. This in particular implies that

$$\int_0^\theta \phi(x)x^{n-1}dx = 0$$

for any $\theta > 0$. In particular, we get by differentiation with respect to θ that

$$\phi(\theta)\theta^{n-1} = 0$$

for $\theta > 0$ which then implies $\phi(\theta) = 0$ for any $\theta > 0$. Thus T is also a complete statistic.

□

Problem 11. *Find a sufficient statistic for a sample from $Beta(\theta, 2)$. Recall that the density of $Beta(\alpha, \beta)$ has density $\frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}$ where $B(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$ and $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}e^{-x}dx$.*

Solution. The Beta distribution is given by density

$$f(x;\theta) = \theta(\theta+1)x^{\theta-1}(1-x)\mathbb{1}_{[0,1]}(x)$$

in other words, the density is $\theta(\theta+1)x^{\theta-1}(1-x)$ for $x \in [0,1]$ and is 0 otherwise. This can also be obtained by integrating over $[0,1]$, the main function $x^{theta-1}(1-x)$. The shortcut to the problem is to notice that this is written as

$$f(x;\theta) = e^{\ln(\theta(\theta+1)x^{\theta-1}(1-x))} = e^{(\theta-1)\ln(x)+\ln(1-x)+\ln(\theta(\theta+1))}$$

for $0 < x < 1$ and 0 otherwise. Thus, this is a regular exponential family and consequently, a complete and sufficient statistic based on a sample of size n is $T = \sum_{k=1}^n \ln(x_k) = \ln(\prod_{k=1}^n x_k)$. This is also the MVUE for the parameter θ.

□

Problem 12. *Find a sufficient statistics for a sample from a Bernoulli distribution with parameter θ. Is this statistic sufficient? Can you find the MVUE for the sample? Hint: write the pmf as $f_X(x;\theta) = \theta^x(1-\theta)^{1-x}$.*

Solution. We discussed this in class. For completeness, just notice that

$$f(x;\theta) = e^{\ln(\theta)x+\ln(1-\theta)(1-x)} = e^{\ln(\theta/(1-\theta))x+\ln(1-\theta)}$$

and consequently, for a sample of size n, we have a complete and sufficient statistics given by $T = \sum_{k=1}^n X_k$. This is not unbiased, but a simple calculation gives that

$$\mathbb{E}[T] = n\theta$$

therefore a quick fix is that T/n is an unbiased estimator and according to the combination of Rao-Blackwell and Lehman-Shaffe's Theorems 3 and 5, we get that T/n is the MVUE.

□

Problem 13. *If $X_1, X_2, \dots, X_n$ is a sample from the $Poisson(\lambda)$, then $\sum_{k=1}^n X_k$ is a sufficient statistic. Find a MVUE? Justify your answer.*

Solution. We did this in class, and is based on the fact that the Poisson is a regular exponential family, with

$$f(x;\theta) = e^{\ln(\theta)x-\ln(x!)-\theta}$$

and thus a complete and sufficient statistic is $T = \sum_{k=1}^n X_k$. Again, as in Bernoulli case from Problem 12, $\mathbb{E}[T] = n\theta$ and consequently, T/n is the MVUE for θ.

□

Problem 14. *Assume that $X_1, X_2, \dots, X_n$ is a sample from the density*

$$f(x;\theta) = \begin{cases} e^{-(x-\theta)}, & x > \theta \\ 0, & otherwise. \end{cases}.$$

(1) *Is $f(x;\theta)$ a regular exponential family?*
(2) *Find a sufficient statistic for θ.*

Solution. (1) This is not a regular exponential family because the support of each $f(x;\theta)$ depends on θ, thus, it does not satisfy the definition in Definition 2.1 because we have

$$f(x;\theta) = \begin{cases} e^{-x-\theta}, & x > \theta \\ 0, & otherwise. \end{cases}$$

This looks as an exponential family, except that the support (the set where the density is non-zero) depends on θ. This is not okay with the definition.

(2) Intuitively, a sufficient statistic is a statistic which contains almost all information about θ. One observation is that if we have a sample $x_1, x_2, \dots, x_n$, then $\theta < \min\{x_1, x_2, \dots, x_n\}$. This suggest something about the min.

To find a sufficient statistic we try to use the factorization property to unravel the structure. First we write

$$f(x;\theta) = \mathbb{1}_{[\theta,\infty)}(x)e^{-(x-\theta)}$$

We take the n sample and the likelihood function to write

$$\begin{aligned} f(x_1;\theta)f(x_2;\theta)\dots f(x_n;\theta) &= \mathbb{1}_{[\theta,\infty)}(x_1)e^{-(x_1-\theta)}\mathbb{1}_{[\theta,\infty)}(x_2)e^{-(x_2-\theta)}\dots \mathbb{1}_{[\theta,\infty)}(x_n)e^{-(x_n-\theta)} \\ &= \mathbb{1}_{[\theta,\infty)}(\min\{x_1, x_2, \dots, x_n\})e^{n\theta}e^{-(x_1+x_2+\dots+x_n)} \\ &= k_1(\min\{x_1, x_2, \dots, x_n\};\theta)k_2(x_1, x_2, \dots, x_n) \end{aligned}$$

where $k_1(y;\theta) = \mathbb{1}_{[\theta,\infty)}(y)e^{n\theta}$ and $k_2(x_1, x_2, \dots, x_n) = e^{-(x_1+x_2+\dots+x_n)}$.

These satisfy the condition from the factorization Theorem 2 and thus guarantees that $T = \min\{X_1, X_2, \dots, X_n\}$ is a sufficient statistic.

□

Problem 15. *If X is a single sample from $N(0,\theta)$, $\theta > 0$ then X is a sufficient but not complete statistic for θ. Can you give an example of a sufficient and complete statistic?*

Solution. Interestingly,

$$f(x;\theta) = e^{x^2/(2\theta)-(1/2)\ln(2\pi\theta)}.$$

Thus this is a regular exponential family. This gives X^2 as a complete and sufficient statistic for a single sample. However this does not show why X is a sufficient statistic because X is not a function of X^2.

On the other hand, we can write

$$f(x;\theta) = k_1(x;\theta)k_2(x)$$

with $k_1(x;\theta) = e^{x^2/(2\theta)-(1/2)\ln(2\pi\theta)}$ and $k_2(x) = 1$. This shows now, from the factorization Theorem 2 that $T = X$ is a sufficient statistic.

X itself is not complete because we can find a function ϕ such that $\mathbb{E}[\phi(X)] = 0$ for all θ, without having ϕ identically 0. Indeed, the most natural function we can think of is an odd function. The simplest of them is $\phi(x) = x$ and indeed we have $\mathbb{E}[X] = 0$ for any θ. Thus X is not complete.

□

Problem 16. *Show that the family $N(\theta,\theta)$ for $\theta > 0$ is a regular exponential family, but $N(\theta,\theta^2)$ is not.*

Solution. Clearly for $N(\theta,\theta)$ we have

$$f(x;\theta) = \frac{1}{\sqrt{2\pi\theta}}e^{-(x-\theta)/(2\theta)} = e^{-x^2/(2\theta)+x-\theta/2-(1/2)\ln(2\pi\theta)}$$

and this is a regular family with $K(x) = x^2$, $p(\theta) = -1/(2\theta)$, $S(x) = x$, $q(\theta) = -\theta/2 - (1/2)\ln(2\pi\theta)$. From this we get that a complete and sufficient statistic based on a sample $X_1, X_2, \dots, X_n$ is $\sum_{k=1}^n X_k^2$.

The other family, namely $N(\theta,\theta^2)$ has

$$f(x;\theta) = \frac{1}{\sqrt{2\pi\theta^2}}e^{-(x-\theta)/(2\theta^2)} = e^{-x^2/(2\theta^2)+x\theta-1/2-(1/2)\ln(2\pi\theta^2)}.$$

It is not possible to write $-x^2/(2\theta^2) + x\theta = p(\theta)K(x)$ because we would have for $\theta = 1$, $p(1)K(x) = -x^2/2 + x$, while for $\theta = -1$, $p(-1)K(x) = -x^2/2 - x$. This means that $p(1)$ and

$p(-1)$ would be non-zero and subtracting the two equalities would give $(p(1)-p(0))K(x) = 2x$, thus K must be of the form $K(x) = ax$. Plugging this back into $p(1)K(x) = -x^2/2 + x$, shows that $p(1)ax = -x^2/2 + x$ which is not possible.

□

Problem 17. *Let $f(x;\theta)$ for θ positive integer be the uniform distribution on $\{1, 2, 3, \ldots, \theta\}$. Take a sample $X_1, X_2, \ldots, X_n$.*

(1) *Set $T = \max\{X_1, X_2, \ldots, X_n\}$. Show that T is a sufficient statistic.*
(2) *Show that T is also a complete statistic.*
(3) *Prove that $U = \frac{T^{n+1}-(T-1)^{n+1}}{T^n-(T-1)^n}$ is the unique MVUE of θ.*

Solution. (1) This is very similar to the Problem 10 and we use here the factorization theorem. First write

$$f(x;\theta) = \frac{1}{\theta}\mathbb{1}_{[0,\theta]}(x)$$

and then for an n sample $x_1, x_2, \ldots, x_n$ we have

$$\begin{aligned} f(x_1;\theta)f(x_2;\theta)\ldots f(x_n\ \theta) &= \frac{1}{\theta^n}\mathbb{1}_{[0,\theta]}(x_1)\mathbb{1}_{[0,\theta]}(x_2)\ldots\mathbb{1}_{[0,\theta]}(x_n) \\ &= \frac{1}{\theta^n}\mathbb{1}_{(-\infty,\theta]}(x_1)\mathbb{1}_{(-\infty,\theta]}(x_2)\ldots\mathbb{1}_{(-\infty,\theta]}(x_n)\mathbb{1}_{[0,\infty)}(x_1)\mathbb{1}_{[0,\infty)}(x_1)\ldots\mathbb{1}_{[0,\infty)}(x_n) \\ &= \frac{1}{\theta^n}\mathbb{1}_{(-\infty,\theta]}(\max\{x_1, x_2, \ldots, x_n\})\mathbb{1}_{[0,\infty)}(\min\{x_1, x_2, \ldots, x_n\}). \end{aligned}$$

Thus from factorization Theorem 2 with $k_1(x;\theta) = \mathbb{1}_{(-infty,\theta]}(x)/theta^n$ and $k_2(x_1, x_2, \ldots, x_n) = \mathbb{1}_{[0,\infty)}(\min\{x_1, x_2, \ldots, x_n\})$ we conclude that $T = \max\{X_1, X_2, \ldots, X_n\}$ is a sufficient statistic.

(2) The completeness is a little different. We need to first compute the distribution of T. This is done by first computing the cumulative function and then extrating the pmf. Thus, for $t \in \{0, 1, 2, \ldots, \theta\}$

$$\begin{aligned} F_T(t) &= \mathbb{P}(\max\{X_1, X_2, \ldots, X_n\} \le t) = \mathbb{P}(X_1 \le t, X_2 \le t, \ldots, X_n \le t) \\ &= \mathbb{P}(X_1 \le t)\mathbb{P}(X_2 \le t)\ldots\mathbb{P}(X_n \le t) \\ &= \left(\frac{t}{\theta}\right)^n \end{aligned}$$

Thus

$$\mathbb{P}(T = t) = \mathbb{P}(T \le t) - \mathbb{P}(T \le t-1) = \frac{t^n - (t-1)^n}{\theta^n} \text{ for } t \in \{1, 2, \ldots, \theta\}. \tag{3.9}$$

To see if the estimator T is complete, we need to check that if ϕ is a function, then (from (1.2))

$$\mathbb{E}[\phi(T)] = \sum_{k=1}^{\theta}\phi(k)\mathbb{P}(T = k) = \sum_{k=1}^{\theta}\phi(k)\frac{k^n - (k-1)^n}{\theta^n}$$

If this is 0 for any choice of θ, this means that $\sum_{k=1}^{\theta}\phi(k)(k^n - (k-1)^n) = 0$, thus, for $\theta = 1$ we get $\phi(1) = 0$, for $\theta = 2$, we get that $\phi(2)(2^n - 1) = 0$ and thus $\phi(2) = 0$. In general if we subtract the equalities obtained for θ and $\theta - 1$, we get

$$\phi(\theta)(\theta^n - (\theta-1)^n) = 0$$

which gives $\phi(\theta) = 0$ for any $\theta = 1, 2, \ldots$. In particular this means that ϕ is identically 0 on the positive integers, which means that T is a complete statistic.

(3) If we show that $U = \frac{T^{n+1}-(T-1)^{n+1}}{T^n-(T-1)^n}$, then the combination of Theorems 3 and 5 shows that the estimator U is the MVUE. Thus the only thing we need to show is that U is unbiased. To this end, we use (1.2) to compute

$$\mathbb{E}[U] = \mathbb{E}[\phi(T)] = \sum_{t=1}^{\theta} \frac{t^{n+1}-(t-1)^{n+1}}{t^n-(t-1)^n} \mathbb{P}(T=t)$$

where $\phi(t) = \frac{t^{n+1}-(t-1)^{n+1}}{t^n-(t-1)^n}$. Now using (3.9), we have

$$\begin{aligned}\mathbb{E}[U] &= \sum_{t=1}^{\theta} \frac{t^{n+1}-(t-1)^{n+1}}{\cancel{t^n-(t-1)^n}} \frac{\cancel{t^n-(t-1)^n}}{\theta^n} \\ &= \frac{1}{\theta^n} \sum_{t=1}^{\theta} (t^{n+1}-(t-1)^{n+1}) \\ &= \frac{1}{\theta^n} \theta^{n+1} = \theta\end{aligned}$$

where the last sum is a telescoping sum and thus it simplifies to the last term. Consequently, U is an unbiased estimator and thus unique MVUE.

$\square$