1.

Briefly describe any advantages of using sufficient, minimally sufficient, and/or complete statistics.

The main advantage is efficiency. Sufficient statistics allow for us to leverage as little of the data as possible to meet our goals. Using sufficient statistics gives us confidence that we haven't lost any crucial information for estimating parameters along the way. When we move down to minimal or complete statistics we have confidence that the statistic we're using is superior to other sufficient statistics. In particular, if we find a complete statistic for our use case we know we can't find something better that is also unbiased.

6.1.15

Let $X_1, \cdots, X_n$ be iid $N(\theta, a\theta^2)$ where a is a constant and θ.

A.

Show that the parameter space does not contain a two dimensional open set.

For this to be an open set we would need to be able to fully explore the parameter space. We can't do that here as the values of μ and σ^2 are completely linked to each other. What we have here is actually a parabolic line. If we were to plot this we would be restricted to the values on the parabola, we can't explore the values above and below it. Thus, we do not have an open set.

B.

Show that the statistic $T = \left(\bar{X}, S^2\right)$ is a sufficient statistic for θ, but the family of distributions is not complete.

For this we'll be leveraging the factorization theorem. I'll primarily be following the example in the book on page 224 as they do the bulk of the algebra work for me. From there I'll be using example 6.2.9 as a template. Our goal is to rearrange the exponent so that we can get both $\bar{x}$ and S^2 in there. For reference,

$$S^2 = \frac{1}{n-1}\sum (x_i - \bar{x})^2$$

$$\begin{aligned}
f(\vec{x} \mid \theta, a\theta^2) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi a\theta^2}} \exp\left(-\frac{(x_i - \theta)^2}{2a\theta^2}\right) \\
&= (2\pi\theta^2)^{-n/2} \exp\left(-\frac{1}{2a\theta^2}\sum_{i=1}^{n}(x_i - \theta)^2\right) \\
&= (2\pi\theta^2)^{-n/2} \exp\left(-\frac{1}{2a\theta^2}\sum_{i=1}^{n}(x_i + \bar{x} - \bar{x} - \theta)^2\right) \\
&= (2\pi\theta^2)^{-n/2} \exp\left(-\frac{1}{2a\theta^2}\left(n(\bar{x} - \theta)^2 + \sum_{i=1}^{n}(x_i - \bar{x})^2\right)\right) \\
&= (2\pi\theta^2)^{-n/2} \exp\left(-\frac{1}{2a\theta^2}\left(n(\bar{x} - \theta)^2 + \frac{n-1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2\right)\right) \\
&= (2\pi\theta^2)^{-n/2} \exp\left(-\frac{1}{2a\theta^2}\left(n(\bar{x} - \theta)^2 + (n-1)S^2\right)\right) \\
&= (2\pi\theta^2)^{-n/2} \exp\left(-\frac{n}{2a\theta^2}(\bar{x} - \theta)^2\right) \cdot \exp\left(-\frac{n-1}{2a\theta^2}S^2\right) \cdot 1
\end{aligned}$$

From this, we have:

$$g(T(\vec{x}) \mid \theta) = (2\pi\theta^2)^{-n/2} \exp\left(-\frac{n}{2a\theta^2}(\bar{x}-\theta)^2\right) \cdot \exp\left(-\frac{n-1}{2a\theta^2}S^2\right)$$
$$h(\vec{x}) = 1$$

So, by the factorization theorem, $T(\vec{x}) = (\bar{x}, S^2)$ is a sufficient statistic for $\vec{\theta}$.

To show that the family of distributions is not complete let us examine the definition of completeness.

Definition 6.2.21: Let $f(t \mid \theta)$ be a family of pdfs or pmfs for a statistic $T(\vec{x})$. The family of probability distributions is called **complete** if:

$$E(g(T)) = 0 \; \forall\theta \implies P(g(T) = 0) = 1 \; \forall\theta$$

It's hard to explain exactly what this means without an example, so I'll save that for the conclusion of this problem. Let's directly examine the expectation of the function of the statistic provided as a hint in the problem:

$$g(\bar{x}, S^2) = \left(\frac{n}{a+n}\right)\bar{X}^2 - \frac{S^2}{a}$$

$$\begin{aligned}
E\left[\left(\frac{n}{a+n}\right)\bar{X}^2 - \frac{S^2}{a}\right] &= \frac{n}{a+n}E[\bar{x}^2] - \frac{1}{a}E[S^2] \\
&= \frac{n}{a+n}(Var(\bar{x}) + (E[\bar{x}])^2) - \frac{1}{a}\cdot a\theta^2 \\
&= \frac{n}{a+n}\left(\frac{a\theta^2}{n} + \theta^2\right) - \theta^2 && (Var(\bar{x}) = \sigma^2/n) \\
&= \frac{a\theta^2}{a+n} + \frac{n\theta^2}{a+n} - \theta^2 \\
&= \frac{\theta^2(a+n)}{a+n} - \theta^2 \\
&= \theta^2 - \theta^2 \\
&= 0
\end{aligned}$$

Why this is proof that the family is **not complete** is that, while this expectation is always zero, $g(\bar{x}, S^2)$ is not 0 following from this. We can come up with a ton of situations where that linear combination of $\bar{x}$ and S^2 are not 0 and yet the expectation is still always 0. The implication does not hold, thus we do not have completeness.

6.20

For each of the following pdfs let $X_1, \cdots, X_n$ be iid observations. Find a complete sufficient statistic, or show that one does not exist.

A

$$f(x \mid \theta) = \frac{2x}{\theta^2} I_{(0,\theta)}(x) I_{(0,\infty)}(\theta)$$

Our order of operations here will be to find a sufficient statistic and then verify that it's complete. So, let's start by finding the joint distribution. The big thing to note here is that x is bounded by θ. This will alter the indicator function in the joint pdf once we pass it through the product.

$$\begin{aligned}
f(\vec{x} \mid \theta) &= \prod_{i=1}^{n} \frac{2x_i}{\theta^2} I_{(0,\theta)}(x_i) I_{(0,\infty)}(\theta) \\
&= \left(\frac{2x}{\theta^2}\right)^n I_{(0,\theta)}(x_{(n)}) && \text{(Distribute the product)} \\
&= x^n \left(\frac{2}{\theta^2}\right)^n I_{(0,\theta)}(x_{(n)}) && \text{(Rearrange)}
\end{aligned}$$

Which gives us:

$$\begin{aligned}
g(T(\vec{x}) \mid \theta) &= \left(\frac{2}{\theta^2}\right)^n I_{(0,\theta)}(x_{(n)}) \\
h(x) &= x^n \\
T(\vec{X}) &= x_{(n)}
\end{aligned}$$

Thus, by the factorization theorem, $T(\vec{X}) = x_{(n)}$ is a sufficient statistic for θ.

To verify that this is complete we need to take its expected value. To do that we need a couple things. First, the pdf of the max order statistic. Second, the cdf of the pdf provided.

The pdf of the order statistics simplifies down greatly for the max. We get:

$$\begin{aligned}
f_{x_{(n)}}(x) &= \frac{n!}{(n-1)!} f_x(x)(F_x(x))^{n-1} \\
&= n f_x(x)(F_x(x))^{n-1}
\end{aligned}$$

Now for the cdf.

$$
\begin{aligned}
F_X(x) &= \int_{t=0}^{t=x} f_X(t)dt \\
&= \int_{t=0}^{t=x} \frac{dt}{\theta^2} dt \\
&= \left[\frac{dt^2}{2\theta^2} \right]_{t=0}^{t=x} \\
&= \frac{x^2}{\theta^2}
\end{aligned}
$$

Plugging this into the pdf for the max gives us:

$$
\begin{aligned}
f_{x_{(n)}}(x) &= n f_x(x)(F_x(x))^{n-1} \\
&= n \frac{2x}{\theta^2} \left(\frac{x^2}{\theta^2} \right)^{n-1} \\
&= \frac{2xn}{\theta^2} \left(\frac{x}{\theta} \right)^{2n-2} \\
&= x^{2n-1} 2n \theta^{-2n} \\
&= \frac{2nx^{2n-1}}{\theta^{2n}}
\end{aligned}
$$

Now finally we can look at the expected value of some function of the max.

$$
\begin{aligned}
E[g(x_{(n)})] &= \int_{x=0}^{x=\theta} g(x) \frac{2nx^{2n-1}}{\theta^{2n}} dx \\
&= \frac{2n}{\theta^2} \int g(x) x^{2n-1} dx
\end{aligned}
$$

Some notes. $\frac{2n}{\theta^2}$ will always be nonzero due to the bounds on both θ and n. Similarly, x's bounds also force x^{2n-1} to be non-zero on its support. The only way for the integrand to evaluate to 0 is for $g(x) = 0$ for all θ. Thus, $T(\vec{x}) = X_{(n)}$ is complete.