1. (5pts) Let $\delta$ be the UMVUE of $g(P)$ for a model $\mathcal{P}_1$, where $\mathcal{P}_0 \subset \mathcal{P}_1 \subset \mathcal{P}_2$. Explain whether or not $\delta$ is necessarily unbiased or the UMVUE for $\mathcal{P}_0$ and/or $\mathcal{P}_2$.

Solution: If $P \in \mathcal{P}_0 \subset \mathcal{P}_1$ then $P$ is in $\mathcal{P}_1$, so unbiasedness for $\mathcal{P}_1$ implies $E_P[\delta] = g(P)$ for all $P \in \mathcal{P}_0$. Alternatively, let $U_1$ be the set of estimators that are unbiased for $g(P)$ for all $P \in \mathcal{P}_1$, and let $U_0$ be the unbiased estimators under $\mathcal{P}_0$. Then $U_1 \subset U_0$ so if something is unbiased for $\mathcal{P}_1$ (i.e. in $U_1$) then it is unbiased for $\mathcal{P}_0$ (in $U_0$). However, the UMVUE for $\mathcal{P}_0$ may lie in $U_0$ but not in $U_1$.

An estimator that is unbiased for $\mathcal{P}_1$ is not necessarily unbiased for $\mathcal{P}_2$, and it follows that $\delta$ is not necessarily UMVUE for $\mathcal{P}_2$ either.

2. (10pts) Let $\delta_1$ and $\delta_2$ have finite variances for each $\theta \in \Theta$ and suppose they are the UMVUEs for $g_1(\theta)$ and $g_2(\theta)$, respectively. Show that $a_1\delta_1 + a_2\delta_2$ is the UMVUE for $a_1g_1(\theta) + a_2g_2(\theta)$, where $a_1$ and $a_2$ are known.

Solution: Recall that an estimator with finite variance is UMVUE for its expectation iff it is uncorrelated with every unbiased estimator of zero (LC Theorem 2.1.7). The estimator $\delta = a_1\delta_1 + a_2\delta_2$ has finite variance, and for an unbiased estimator $U$ of zero we have

$$
\text{Cov}[\delta, U] = E[\delta U] = a_1E[\delta_1 U] + a_2E[\delta_2 U] = 0
$$

since both $\delta_1$ and $\delta_2$ are UMVUE for their expectations and so uncorrelated with $U$.

Several people tried to solve this problem with the variance inequality on $\delta$ and then on another estimator $\delta' = a_1\delta_1' + a_2\delta_2'$. With this approach, you can show that the bounds for $\delta$ are lower than the bounds for $\delta'$ based purely on $\delta_1$ and $\delta_2$ being UMVUEs. However, this approach will not work for two reasons: An ordering on the bounds doesn’t imply the variances themselves are ordered, and we need to consider estimators $\delta'$ that are not linear combinations of estimators of $g_1(\theta)$ and $g_2(\theta)$.

A few people argued that since $\delta_1$ and $\delta_2$ are UMVUE, they must be functions of a complete sufficient statistic. Then $\delta$ will be a function of the c.s.s., and is unbiased for $a_1g(\theta) + a_2g(\theta)$. This reasoning is flawed because there exist UMVUEs in situations
where there is not a c.s.s. (see L. Bondesson (1983), “On Uniformly Minimum Variance Unbiased Estimation when no Complete Sufficient Statistics Exist”). If there is a c.s.s., then a UMVUE will be a function of it, but a UMVUE can exist without a c.s.s.

3. (15pts) Let $X_1, \ldots, X_n \sim \text{i.i.d.}$ with $E[|X_i|] < \infty$. Show formally that the sample mean $\bar{X}$ is a version of $E[X_1|X(1), \ldots, X(n)]$.

SOLUTION: For this problem you need to show that $\int_A \bar{X} \, dP = \int_A X_1 \, dP$ for every $A \in \sigma(X(1), \ldots, X(n))$. Do this by starting with $\bar{X}$:

$$
\int_A \bar{X} \, dP = \frac{1}{n} \sum_{i=1}^{n} \int_A X_i \, dP
= \frac{1}{n} \sum_{i=1}^{n} E[X_1 1_A]
= \frac{1}{n} \sum_{i=1}^{n} E[X_1 1_A] \text{ because the } X_i \text{ s are i.i.d.}
= E[X_1 1_A] = \int_A X_1 \, dP.
$$

You need to be careful about the second to last line, as it only generally holds for $A \in \sigma(X(1), \ldots, X(n))$. For $A$ outside of this $\sigma$-algebra, $E[X_1 1_A]$ is not necessarily the same as $E[X_i 1_A]$ for arbitrary sets $A$ even if the $X_i$’s are i.i.d. For example if $A = \{X_2 < 3\}$ then the result doesn’t hold.

4. (20pts) Let $X_1, X_2$ be i.i.d. $f(x|\theta)$ where

$$
f(x|\theta) = \begin{cases} 
3x^2/\theta^3 & \text{if } 0 < x < \theta, \\
0 & \text{otherwise.}
\end{cases}
$$

(a) For $\theta \in \Theta = \mathbb{R}^+$, find a complete sufficient statistic for $\theta$ based on $X_1$ and $X_2$.

SOLUTION: The factorization criteria shows that $T = X_{(2)} = \max\{X_1, X_2\}$ is sufficient, and Lehmann and Scheffe’s theorem shows it is minimal. To show that it is complete, suppose $E_\theta[g(T)]$ for all $\theta > 0$, i.e.

$$
\int_0^\theta g(t)p(t|\theta) \, dt = 0 \ \forall \ \theta > 0,
$$
where \( p(t|\theta) = 6t^5/\theta^6 \) is the density of \( T \). Letting \( g_+ \) and \( g_- \) be the positive and negative parts of \( g = g_+ - g_- \), we have

\[
\int_0^\theta g_+(t)t^5 \, dt = \int_0^\theta g_-(t)t^5 \, \forall \theta > 0.
\]

It was sufficient for the exam to make reference to the result in Lehmann (or my supplementary notes) that this implies \( g_+ = g_- \) a.e., so that \( g = 0 \) a.e.

Alternatively, you can take derivatives:

\[
\frac{d}{d\theta} \int_0^\theta g(t)t^5 \, dt = \frac{d}{dt}0
\]

\[
g(\theta)\theta^5 = 0
\]

\[
g(\theta) = 0 \text{ for almost all } \theta > 0 \Rightarrow g = 0 \text{ a.e.}
\]

Thus \( E_\theta[g(T)] = 0 \) for all \( \theta \) implies \( g = 0 \) a.e., so \( T \) is complete.

(b) Find the UMVUE estimator for \( \theta \).

**SOLUTION:** The expectation of \( T \) is \( 6\theta/7 \), so the UMVUE is \( 7T/6 \).

(c) Determine if the UMVUE is admissible for estimating \( \theta \) with squared error loss (Hint: Consider the class of scale multiples of the UMVUE).

**SOLUTION:** Considering the class of estimators \( cT \), the minimum risk (variance plus squared bias) is attained when \( c = 8/7 \), so \( 7T/6 \) is not admissible even in this specific class.