MATH524 – Spring 2025

Problem Set: Week 5

1. (**Rao-Blackwell property of U-statistics**) Prove the following property of U-statistics:

$$\mathbb{E}[\hat{U}_n^2] \leq \mathbb{E}[(h(X_1, \ldots, X_r) - \theta)^2]$$

Solution: The inequality is proven by simplifying and applying Jensen's inequality:

$$\begin{aligned}
\mathbb{E}[\hat{U}_n^2] &= \mathbb{E}\left[(\mathbb{E}[h(X_1, \ldots, X_r) - \theta | \{X_i\}])^2\right] \\
&\leq \mathbb{E}\left[(\mathbb{E}[(h(X_1, \ldots, X_r) - \theta)^2 | \{X_i\}]\right] \\
&= \mathbb{E}[(h(X_1, \ldots, X_r) - \theta)^2].
\end{aligned}$$

This is known as the Rao-Blackwell theorem which states that the mean squared-error of any estimator can be reduced by conditioning it on a sufficient statistic of the data.

2. (**Sampling variance of variance**) Recall that for the variance estimation

$$h(x_1, x_2) = \frac{1}{2}(x_1 - x_2)^2$$

Using the following questions we will derive the variance of the sample variance.

(a) Determine the functional form of $\tilde{h}_1(x_1)$

Solution:

$$\tilde{h}_1(x_1) = \mathbb{E}[h(X_1, X_2)|X_1 = x_1] - \sigma^2 = \frac{1}{2}((X_1 - \mu)^2 - \sigma^2)$$

(b) Compute $\mathbb{E}[\tilde{h}_1^2]$.

Solution:

$$\mathbb{E}[\tilde{h}_1^2] = \frac{1}{4}\mathrm{Var}((X_1 - \mu)^2) = \frac{1}{4}(\mathbb{E}[(X - \mu)^4] - \sigma^4)$$

(c) Compute $\mathbb{E}[h^2]$.

Solution:

$$\begin{aligned}
\mathbb{E}[h^2] &= \frac{1}{4}\mathbb{E}[((X_1 - \mu) - (X_2 - \mu))^4] \\
&= \frac{1}{4}\sum_{j=0}^{4}\binom{4}{j}\mathbb{E}[(X_1 - \mu)^j]\mathbb{E}[(X_2 - \mu)^{4-j}] \\
&= \frac{1}{4}(2\mathbb{E}[(X - \mu)^4] + 6\sigma^4).
\end{aligned}$$

Note that the second equality comes from the expansion of the fourth power inside the expectation. The final equality follows from the fact that $\mathbb{E}[(X - \mu)] = 0$.

(d) Compute $\mathrm{Var}(U_n)$. Note that this is equivalent to $\mathrm{Var}(s_n^2)$.

Solution: Using the Hoeffding decomposition,

$$\begin{aligned}
\text{Var}(s_n^2) &= \binom{n}{2}^{-1} \left(2(n-2)\mathbb{E}[\tilde{h}_1^2] + (\mathbb{E}[h^2] - \theta^2)\right) \\
&= \frac{2}{n(n-1)} \left(2(n-1)\mathbb{E}[\tilde{h}_1^2] - \mathbb{E}[\tilde{h}_1^2] + (\mathbb{E}[h^2] - \theta^2)\right) \\
&= \frac{4\mathbb{E}[\tilde{h}_1^2]}{n} - \frac{4\mathbb{E}[\tilde{h}_1^2]}{n(n-1)} + \frac{2(\mathbb{E}[h^2] - \theta^2)}{n(n-1)} \\
&= \frac{\mu_4 - \sigma^4}{n} + \frac{2\sigma^4}{n(n-1)},
\end{aligned}$$

where we use the notation $\mu_4 = \mathbb{E}[(X-\mu)^4]$ is used to denote the centralized moment.

(e) Determine the asymptotic distribution of $\sqrt{n}(U_n - \theta)$.

Solution: We start by first taking the variance from the previous part, and noting that the second term is of order $O(n^{-2})$. This means the second term is of higher-order even when the U-statistic is scaled by $\sqrt{n}$, and vanishes in the asymptotic limit. Then, using the asymptotic normality theorem for U-statistice, we can verify that all the conditions of the statement hold for the variance estimator. Thus,

$$\sqrt{n}(U_n - \sigma^2) \xrightarrow{d} \mathbb{N}(0, 4\mathbb{E}[h_1^2]).$$

We can identify the exact variance from the solution to part (b):

$$\mathbb{E}[h_1^2] = \frac{\mu_4 + \sigma^2}{4}.$$

3. (**Degenerate U-statistics**) Consider the sample variance function, as in Question 1.

$$h(x_1, x_2) = \frac{1}{2}(x_1 - x_2)^2$$

(a) Show that the kernel of the U-statistics is first-order degenerate if $\mathbb{E}[(X - \mathbb{E}[X])^4] = \mathbb{E}[(X - \mathbb{E}[X])^2]^2$.

Hint: *It may be useful to consider the variance derived in part (e) from the first question.*

Solution: From Question 1, part (e), we know that the variance of the asymptotic distribution is $\mu_4 - \sigma^4$. The U-statistic will be first-order degenerate precisely if this variance is equal to 0. This directly gives the condition from the question since μ_4 is the centralized fourth moment and σ^2 is the centralized second moment.

(b) Show that if the U-statistic is degenerate then X must be a discrete random variable and find its density function.

Solution: We start by observing that $\mu_4 = \sigma^4$ can only hold if $\mathbb{E}[(X - \mathbb{E}[X])^2]$ is some constant c. WLOG, we can start by setting $\mathbb{E}[X] = 0$. Then, if we define the distribution of X as

$$X = \begin{cases} -\sigma & w.p.\ 1/2 \\ \sigma & w.p.\ 1/2, \end{cases}$$

for $\sigma = \sqrt{\mu_2}$, we get that $\mu_4 = \sigma^4$ and thus U will be first-order degenerate.

(c) Using the asymptotic normality result for degenerate U-statistics[1] determine the limiting distribution of

$$n(s_n^2 - \sigma^2).$$

[1] Use may wish to use the following fact with proof: The eigenvalues for h_2 in this case are 0 and $-\sigma^2$. The corresponding eigendecomposition is given by $h_2(x_1, x_2) = -\sigma^2 \frac{x_1}{\sigma} \frac{x_2}{\sigma}$.

Solution: Start by evaluating all of the necessary conditional expectations.

$$\mathbb{E}[(X_1 - X_2)^2/2] = \sigma^2.$$

Furthermore,

$$\tilde{h}_1(x_1) = \mathbb{E}[(x_1 - X_2)^2/2|X_1 = x_1] - \sigma^2 = -\sigma^2.$$

By this fact,

$$\begin{aligned} h_2(x_1, x_2) &= \frac{1}{2}(x_1 - x_2)^2 - \mathbb{E}[(x_1 - X_2)^2/2] - \mathbb{E}[(X_1 - x_2)^2/2] + \sigma^2 \\ &= \frac{1}{2}(x_1 - x_2)^2 - \sigma^2. \end{aligned}$$

It can be verified that the eigen values of this function are 0 and $-\sigma^2$. Thus, the eigen decomposition is given by

$$h_2(x_1, x_2) = -\sigma^2 \frac{x_1}{\sigma}\frac{x_2}{\sigma}.$$

Now we can simply apply the theorem for asymptotic convergence of degenerate U-statistics from Chapter 4 to get that

$$n(s_n^2 - \mu_2) \xrightarrow{d} -\sigma^2(Z^2 - 1).$$

4. (**Degenerate moment estimation**) Lets take another look at the squared-moment estimation problem from Example 8 in Chapter 4. Recall that for $\theta = \mathbb{E}[X]^2 = 0$, the U-statistic with a second-order kernel,

$$U = \frac{1}{n(n-1)} \sum_j \sum_{i \neq j} X_i X_j$$

has a degenerate approximation under the standard analysis tools. Now suppose that $\sigma_2^2 = \text{Var}(X_1, X_2) = \sigma^4 > 0$. This means that U is first-order degenerate. Show that we can still establish the following distributional approximation:

$$nU_n \xrightarrow{d} \sigma^2(Z^2 - 1),$$

where $Z \sim \mathbb{N}(0, 1)$. ***Hint:*** *You may choose to prove this by decomposing the U-statistic double-sum into two separate single-summations and then analyze each term individually.*

Solution: Following the hint, we can investigate the double-sum. Start by considering the case for $j = 1$:

$$\sum_{i \neq j} X_i X_j = \sum_{i=2}^{n} X_i X_1 = X_1 \sum_{i=2}^{n} = X_1(\sum_{i=1}^{n} X_i - X_1) = X_1 \sum_{i=1}^{n} X_i - X_1^2.$$

But this argument can be repeated for all possible values of j (all X_i are i.i.d.) and added together. Thus,

$$\sum_{i \neq j} X_i X_j = \left(\sum_{i=1}^{n} X_i\right)^2 - \sum_{i=1}^{n} X_i^2.$$

Plugging this decomposition back into the definition of U,

$$\begin{aligned} U = \frac{1}{n(n-1)} \sum_{i<j} X_i X_j &= \frac{1}{n(n-1)} \left(\left(\sum_{i=1}^{n} X_i\right)^2 - \sum_{i=1}^{n} X_i^2 \right) \\ &= \frac{1}{(n-1)} \left(\left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i\right)^2 - \frac{1}{n} \sum_{i=1}^{n} X_i^2 \right). \end{aligned}$$

Now, both terms can be analysed separately. For the first term, we can directly apply the CLT:

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i \xrightarrow{d} \mathbb{N}(0, \sigma^2) =: \sigma Z.$$

For the second term, by the LLN,

$$\frac{1}{n} \sum_{i=1}^{n} X_i^2 \xrightarrow{p} \sigma^2.$$

Then, by Slutsky's Theorem, we can put both terms together,

$$\left(\left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i \right)^2 - \frac{1}{n} \sum_{i=1}^{n} X_i^2 \right) \xrightarrow{d} (\sigma Z)^2 - \sigma^2 = \sigma^2 (Z^2 - 1).$$

5. (**Consistency of empirical CDF**) Using the SLLN, show tht $F_n(x) \xrightarrow{a.s.} F(x)$ for all $x \in \mathbb{R}$. **Solution**: If we consider the random variable $Y_i = \mathbf{1}(X_i \leq x)$. We know Y_i are i.i.d. Furthermore,

$$\mathbb{E}[|Y_i|] = \mathbb{E}[|\mathbf{1}(X_i \leq x)|] = \mathbb{E}[\mathbf{1}(X_i \leq x)] = \mathbb{P}(X_i \leq x) = F(x) < \infty.$$

Then, we can apply Kolmogorov's SLLN (see Chapter 1 notes for the full statement) directly and get the desired result.