Pseudo-Kernel Method in U-statistic Variance Estimation with Large Kernel Size

Qing Wang[1] and Bruce Lindsay[2]

[1]*Department of Mathematics, Wellesley College, Wellesley, MA*
[2]*Department of Statistics, Pennsylvania State University, University Park, PA*

Supplementary Materials

The following supplementary materials include:
S1. Derivation of the asymptotic kernel function of degree two based on Kullback-Leibler distance.
S2. Proof of Theorem 2, the consistency of the pseudo-kernel variance estimator of degree two.
S3. Expression of jackknife variance estimator using a pseudo-kernel of degree one.
S4. Proof of Theorem 3, the third-order unbiasedness of a pseudo-kernel variance estimator of degree three.

S1 Derivation of the asymptotic kernel of degree two

Let θ^* be the true value of the parameter θ in $\Theta \subseteq \mathcal{R}^p$. Assume f is twice-differentiable at point θ^* and define $\hat{\theta} := \hat{\theta}(\mathcal{X}_m)$. According to Taylor series, the log-likelihood function can be expanded around θ^* as follows:

$$\log f_{\hat{\theta}}(x) = \log f_{\theta^*}(x) + (\hat{\theta} - \theta^*)^T \frac{\partial}{\partial \theta} \log f_\theta(x)|_{\theta=\theta^*} + Re,$$

where the remainder term $Re = o(\|\hat{\theta}-\theta^*\|)$ and $\|\cdot\|$ represents the Euclidean norm.

Under the regularity conditions C1–C5 (Lehmann (2004)) shown in Appendix A1, the Maximum Likelihood (ML) estimator $\hat{\theta}$ of θ^* is consistent, i.e. $\hat{\theta} \xrightarrow{P} \theta^*$ as $m \to \infty$. Denote the Fisher score, i.e. the first derivative of the log-likelihood, evaluated at θ^* as $u(x)$. Then,

$$\log f_{\hat{\theta}}(x) = \log f_{\theta^*}(x) + (\hat{\theta} - \theta^*)^T u(x) + o(\|\hat{\theta} - \theta^*\|).$$

Given a training sample $\mathcal{X}_m$ of size m, write the joint log-likelihood as

$$l(\theta) = l(\theta|\mathcal{X}_m) = \sum_{x_j \in \mathcal{X}_m} \log f_\theta(x_j).$$

We have

$$l'(\theta) = l'(\theta^*) + l''(\theta^*)(\theta - \theta^*) + o(\|\theta - \theta^*\|).$$

Let $H(\theta) = l''(\theta) = \sum_{x_j \in \mathcal{X}_m} \frac{\partial^2}{\partial\theta\partial\theta^T} \log f_\theta(x_j)$ be the $p \times p$ Hessian matrix. Because the ML estimator $\hat{\theta}$ is the root of $l'(\theta) = 0$ and $l'(\theta^*) = \sum_{x_j \in \mathcal{X}_m} u(x_j)$, we have

$$l'(\theta^*) + H(\theta^*)(\hat{\theta} - \theta^*) = 0 + o(\|\hat{\theta} - \theta^*\|).$$

This implies

$$\hat{\theta} - \theta^* \approx -H^{-1}(\theta^*) \sum_{x_j \in \mathcal{X}_m} u(x_j)$$

with an error of order $o(\|\hat{\theta} - \theta^*\|)$. Let $\mathbf{I}(\theta^*)$ be the Fisher Information matrix evaluated at θ^*, defined as

$$\mathbf{I}(\theta^*) = -E\left[\frac{\partial^2}{\partial\theta\partial\theta^T} \log f_\theta(X)\right] |_{\theta=\theta^*} .$$

Because $X_1, \ldots, X_m$ are independent and identically distributed with finite variance, by the Law of Large Numbers as $m \to \infty$

$$\frac{1}{m} \sum_{X_j \in \mathcal{X}_m} \frac{\partial^2}{\partial\theta\partial\theta^T} \log f_\theta(X_j) \xrightarrow{a.s.} -\mathbf{I}(\theta).$$

Therefore, $\frac{1}{m}H(\theta^*) \xrightarrow{a.s.} -\mathbf{I}(\theta^*)$, which implies that $H^{-1}(\theta^*) = -(1/m)\mathbf{I}^{-1}(\theta^*) + o(1/m)$. We then have

$$-H^{-1}(\theta^*) \sum_{x_j \in \mathcal{X}_m} u(x_j) = \mathbf{I}(\theta^*)^{-1} \left(\frac{1}{m} \sum_{x_j \in \mathcal{X}_m} u(x_j)\right) + o(1).$$

For large enough m

$$\log f_{\hat{\theta}}(x_i) \approx \log f_{\theta^*}(x_i) + \left(\frac{1}{m} \sum_{x_j \in \mathcal{X}_m} u(x_j)\right)^T \mathbf{I}(\theta^*)^{-1} u(x_i).$$

Therefore, under Kullback-Leibler distance the symmetric kernel function $\phi(\mathcal{X}_{m+1})$ defined in Equation (2.1) has the following approximation:

$$\begin{aligned}
&\phi(X_1,\ldots,X_{m+1})\\
=&\ -\frac{1}{m+1}\sum_{X_i\in\mathcal{X}_{m+1}}\left[\log f_{\theta^*}(X_i)+\left(\frac{1}{m}\sum_{X_j\in\mathcal{X}_m^{(-i)}}u(X_j)\right)^T\mathbf{I}(\theta^*)^{-1}u(X_i)\right]+o(\|\hat{\theta}-\theta^*\|)\\
\approx&\ -\binom{m+1}{2}^{-1}\sum_{1\le i<j\le m+1}\left\{\frac{1}{4}\left[\log f_{\theta^*}(X_i)+\log f_{\theta^*}(X_j)\right]+u(X_j)^T\mathbf{I}(\theta^*)^{-1}u(X_i)\right\}
\end{aligned}$$

The error of the approximation is of order $o(\|\hat{\theta}-\theta^*\|)$.

When $\hat{\theta}$ is $\sqrt{m}$-consistent, $\|\hat{\theta}-\theta^*\| = O_p(1/\sqrt{m})$. In this case, the error of the approximation is of order $O_p(1/\sqrt{m})$. The detailed proof for this statement is shown below:

Denote the approximation error as error $= o(\|\hat{\theta}-\theta^*\|)$ and assume $\hat{\theta}$ is $\sqrt{m}$-consistent for θ^*. We have

- $\frac{|\text{error}|}{\|\hat{\theta}-\theta\|} = o(1)$. That is, for any $\epsilon > 0$, there exists an integer M such that for $m \ge M$ we have $\frac{|\text{error}|}{\|\hat{\theta}-\theta\|} \le \epsilon$.
- $\|\hat{\theta}-\theta^*\| = O_p(1/\sqrt{m})$. That is, for any $\epsilon > 0$, there exists C_ϵ and M_ϵ (both depend on ϵ) such that for $m \ge M_\epsilon$ we have
$$P\left(\frac{\|\hat{\theta}-\theta^*\|}{1/\sqrt{m}} > C_\epsilon\right) \le \epsilon.$$

For any $\epsilon > 0$, let $M_\epsilon^* = \max\{M, M_\epsilon\}$ and $C_\epsilon^* = \epsilon C_\epsilon$. If $m \ge M_\epsilon^*$, then

(i). $\frac{|\text{error}|}{\|\hat{\theta}-\theta\|} \le \epsilon$

(ii). $P\left(\frac{\|\hat{\theta}-\theta^*\|}{1/\sqrt{m}} > C_\epsilon\right) \le \epsilon$

From (i) we have
$$\frac{|\text{error}|}{1/\sqrt{m}} \le \frac{\epsilon\|\hat{\theta}-\theta^*\|}{1/\sqrt{m}}.$$

Let $A = \left\{\frac{|\text{error}|}{1/\sqrt{m}} > C_\epsilon^*\right\}$ and $B = \left\{\frac{\epsilon\|\hat{\theta}-\theta^*\|}{1/\sqrt{m}} > C_\epsilon^*\right\}$. It's easy to see that $A \subseteq B$. Therefore,

$$P\left(\frac{\text{error}}{1/\sqrt{m}} > C_\epsilon^*\right) \le P\left(\frac{\epsilon\|\hat{\theta}-\theta^*\|}{1/\sqrt{m}} > C_\epsilon^*\right) = P\left(\frac{\|\hat{\theta}-\theta^*\|}{1/\sqrt{m}} > C_\epsilon\right) \le \epsilon.$$

The last inequality holds because of (ii). As a result, if $\hat{\theta}$ is $\sqrt{m}$-consistent, the error of the approximation satisfies error $= O_p(1/\sqrt{m})$.

S2 Proof of Theorem 2

Let $\gamma = \text{Var}(U_n)$. By the Chebyshev's inequality, for any $\epsilon > 0$,

$$\begin{aligned} P(\mid \hat{V}_{\text{PS}} - \gamma \mid \geq \epsilon) &\leq \epsilon^{-2} E[(\hat{V}_{\text{PS}} - \gamma)^2] \\ &= \epsilon^{-2}[\text{bias}^2(\hat{V}_{\text{PS}}) + \text{Var}(\hat{V}_{\text{PS}})]. \end{aligned}$$

Deviation for the bias is as follows:
From Theorem 1 $\hat{V}_{\text{PS}}$ is second-order unbiased so that $E(\hat{V}_{\text{PS}}) = \binom{k}{1}^2 \binom{n}{1}^{-1} \delta_1^2 + \binom{k}{2}^2 \binom{n}{2}^{-1} \delta_2^2 + o(n^{-2})$, where δ_c^2's $(1 \leq c \leq k)$ are the variances of the orthogonal terms in the Hoeffding decomposition of U_n as defined in Appendix A2. Thus,

$$\text{bias}(\hat{V}_{\text{PS}}) = o(n^{-2}).$$

Deviation for the variance is as follows:
Let $h^{(c)}(x_1, \ldots, x_c)$ $(1 \leq c \leq k)$ be the orthogonal terms in the Hoeffding decomposition of the original U-statistic U_n. From the proof of Theorem 1 (Appendix A2)

$$\phi_{\text{PS}}(x_i, x_j) = \theta + \frac{k}{2}[h^{(1)}(x_i) + h^{(1)}(x_j)] + \frac{k(k-1)}{2} h^{(2)}(x_i, x_j) + \text{remainder},$$

where the remainder involves $h^{(c)}$ terms for $3 \leq c \leq k$ and the leading term in the remainder is $\{k(k-1)(k-2)/[2(n-2)]\} \sum_{l \neq i \text{ and } l \neq j} h^{(3)}(x_i, x_j, x_l)$.

Let $\phi_0(x_i, x_j) = (k/2)[h^{(1)}(x_i) + h^{(1)}(x_j)] + [k(k-1)/2] h^{(2)}(x_i, x_j)$, a symmetric kernel function of degree two. Write $\phi_{\text{PS}}(x_i, x_j) = \theta + \phi_0(x_i, x_j) +$ remainder. Therefore, $U_{\text{PS}} = \theta + U_0 +$ remainder, where

$$U_{\text{PS}} = \binom{n}{2}^{-1} \sum_{1 \leq i < j \leq n} \phi_{\text{PS}}(x_i, x_j), \text{ and } U_0 = \binom{n}{2}^{-1} \sum_{1 \leq i < j \leq n} \phi_0(x_i, x_j).$$

It is easy to show that U_0 can be equivalently written as

$$U_0 = \binom{k}{1} \binom{n}{1}^{-1} \sum_{i=1}^{n} h^{(1)}(x_i) + \binom{k}{2} \binom{n}{2}^{-1} \sum_{1 \leq i < j \leq n} h^{(2)}(x_i, x_j).$$

Next, we show that the following equation holds:

$$\hat{V}_{\text{PS}} = \hat{V}_0 + o(n^{-2}).$$

From the derivation of the pesudo-kernel function ϕ_{PS} in the Proof of Theorem 1 (Appendix A2), we have

$$\begin{aligned}\phi_{\text{PS}}(x_i, x_j) &= \theta + \frac{k}{2}\left[h^{(1)}(x_i) + h^{(1)}(x_j)\right] + \binom{k}{2} h^{(2)}(x_i, x_j) \\ &+ \frac{k(k-1)(k-2)}{2(n-2)} \sum_{l \neq i \text{ or } j} \left[h^{(3)}(x_i, x_j, x_l) + O(n^{-1})\right] + \text{remainder},\end{aligned}$$

where $O(n^{-1})$ is of order n^{-1} and only depends on $h^{(3)}$, and the remainder terms involve $h^{(c)}$ $(4 \leq c \leq k)$ and are orthogonal to each other and to the previous terms.

By the definition of U_{PS} we have

$$\begin{aligned}U_{\text{PS}} &= \binom{n}{2} \sum_{1 \leq i < j \leq n} \phi_{\text{PS}}(x_i, x_j) \\ &= \theta + \binom{k}{1}\binom{n}{1}^{-1} \sum_{i=1}^{n} h^{(1)}(x_i) + \binom{k}{2}\binom{n}{2}^{-1} \sum_{1 \leq i < j \leq n} h^{(2)}(x_i, x_j) \\ &+ \binom{k}{3}\binom{n}{3}^{-1} \sum_{1 \leq i < j \leq n} \sum_{l \neq i \text{ or } j} \left[h^{(3)}(x_i, x_j, x_l) + O(n^{-1})\right] + \text{remainder},\end{aligned}$$

where the remainder terms are orthogonal to the previous terms.

Denote the Hoeffding orthogonal terms of U_{PS} as $h^{*(c)}$ $(1 \leq c \leq k)$. We have

$$\begin{aligned}h^{*(1)}(x_1) &= h^{(1)}(x_1), \\ h^{*(2)}(x_1, x_2) &= h^{(2)}(x_1, x_2), \\ \text{and } h^{*(3)}(x_1, x_2, x_3) &= h^{(3)}(x_1, x_2, x_3) + O(n^{-1}).\end{aligned}$$

In general, $h^{*(c)}(x_1, \ldots, x_c) \neq h^{(c)}(x_1, \ldots, x_c)$ for $3 \leq c \leq k$.

Let δ_c^{*2} be the variance of $h^{*(c)}$. Then,

$$\begin{aligned}\mathrm{Var}(U_{\mathrm{PS}}) &= \sum_{c=1}^{k} \binom{k}{c}^2 \binom{n}{c}^{-1} \delta_c^{*2} \\ &= \binom{k}{1}^2 \binom{n}{1}^{-1} \delta_1^2 + \binom{k}{2}^2 \binom{n}{2}^{-1} \delta_2^2 + \sum_{c=3}^{k} \binom{k}{c}^2 \binom{n}{c}^{-1} \delta_c^{*2}.\end{aligned}$$

In addition, recall that

$$\mathrm{Var}(U_0) = \binom{k}{1}^2 \binom{n}{1}^{-1} \delta_1^2 + \binom{k}{2}^2 \binom{n}{2}^{-1} \delta_2^2.$$

As discussed in Section 1 of the manuscript, the unbiased U-statistic variance estimator as defined in Equation (1.3) (Wang and Lindsay (2014)) is equivalent to the unbiased variance estimator in Maesono (1998), where the latter estimates the variance of each Hoeffiding orthogonal term unbiasedly. Let $\hat{\delta}_t^2$ be the unbiased estimator for δ_t^2 $(t = 1, 2)$ and let $\hat{\delta}_c^{*2}$ be the unbiased estimator for δ_c^{*2} $(3 \le c \le k)$. Since the unbiased U-statistic variance estimator has a U-statistic representation itself (Wang and Lindsay (2014)), it is the best unbiased variance estimator and can be written as a function of the order statistics. Recall that the set of order statistics is the complete sufficient statistics in nonparametric inference (Fraser (1954)). By Lehmann-Scheffe theorem, the best unbiased variance estimators for U_{PS} and U_0 can be written uniquely as

$$\begin{aligned}\hat{V}_{\mathrm{PS}} &= \binom{k}{1}^2 \binom{n}{1}^{-1} \hat{\delta}_1^2 + \binom{k}{2}^2 \binom{n}{2}^{-1} \hat{\delta}_2^2 + \sum_{c=3}^{k} \binom{k}{c}^2 \binom{n}{c}^{-1} \hat{\delta}_c^{*2}, \\ \hat{V}_0 &= \binom{k}{1}^2 \binom{n}{1}^{-1} \hat{\delta}_1^2 + \binom{k}{2}^2 \binom{n}{2}^{-1} \hat{\delta}_2^2,\end{aligned}$$

which implies $\hat{V}_{\mathrm{PS}} = \hat{V}_0 + o(n^{-2})$.

Wang and Lindsay (2014) point out in their Remark 3 that the unbiased variance estimator of a U-statistic can be written in the form of a U-statistic itself with a kernel function of order $2k$. Since the degree for ϕ_0 is $k = 2$, we can write

$$\hat{V}_0 = \binom{n}{4}^{-1} \sum_{1 \le i_1 < \cdots < i_4 \le n} \psi(x_{i_1}, \ldots, x_{i_4}).$$

The definition for the kernel function $\psi(x_{i_1}, \ldots, x_{i_4})$ is given below:
Let S_m be a data subset of size m ($m \in \mathbb{N}^+$). The kernel function $\psi(x_{i_1}, \ldots, x_{i_4})$ for $\hat{V}_0$ is defined by

$$\psi(S_4) = \psi_4(S_4) - \psi_0(S_4),$$

where

$$\begin{aligned} \psi_0(S_4) &= \frac{\binom{n}{4}}{\binom{n}{2}\binom{n-2}{2}} \sum_{S_{2,a}, S_{2,b} \subset S_4} \phi_0(S_{2,a})\phi_0(S_{2,b}) I\{S_{2,a} \cap S_{2,b} = \emptyset\}, \\ \psi_4(S_4) &= \frac{\binom{n}{4}}{\binom{n}{2}^2} \sum_{S_{2,a}, S_{2,b} \subset S_4} \phi_0(S_{2,a})\phi_0(S_{2,b})\omega(a,b), \end{aligned}$$

$\omega(a,b) = 1/n(a,b)$, and $n(a,b) = \binom{n-(4-c)}{c}$ with c being the number of overlaps between $S_{2,a}$ and $S_{2,b}$. More details on the derivation of the U-statistic representation of the unbiased variance estimator $\hat{V}_u$ can be found in Wang (2012).

As a result, the variance of $\hat{V}_0$ can be expressed explicitly as shown in Hoeffding (1948). Let $h^{*(c)}$ ($1 \leq c \leq 4$) be the orthogonal terms in the Hoeffding decomposition of $\hat{V}_0$, and denote δ_c^{*2} as the variance of $h^{*(c)}$ ($1 \leq c \leq 4$). We have

$$\mathrm{Var}(\hat{V}_0) = \sum_{c=1}^{4} \binom{4}{c}^2 \binom{n}{c}^{-1} \delta_c^{*2} = \frac{16}{n}\delta_1^{*2} + o(1/n).$$

Assume the original U-statistic U_n is well defined so that $\delta_1^2 = \mathrm{Var}[h^{(1)}(X_1)] < \infty$. We then have $\delta_1^{*2} = \mathrm{Var}\{E[\psi(x_1, \ldots, x_4) \mid x_1 = X_1]\} < \infty$. Therefore,

$$\mathrm{Var}(\hat{V}_{\mathrm{PS}}) = \mathrm{Var}\left[\hat{V}_0 + o(n^{-2})\right] = \frac{16}{n}\delta_1^{*2} + o(1/n).$$

Based on the above deviations for the bias and variance of $\hat{V}_{\mathrm{PS}}$, for any $\epsilon > 0$

$$P(\mid \hat{V}_{\mathrm{PS}} - \gamma \mid > \epsilon) \leq \epsilon^{-2}\left(o(1/n^4) + \frac{16}{n}\delta_1^{*2} + o(1/n)\right)$$

Let $M = \left\{16\left([\frac{2\delta_1^{*2}}{\epsilon^2}] + 1\right)\right\}^2$, where the notation $[\frac{2\delta_1^{*2}}{\epsilon^2}]$ represents the integer part of $\frac{2\delta_1^{*2}}{\epsilon^2}$. Then, for any $n \geq M$ we have

$$P(\mid \hat{V}_{\mathrm{PS}} - \gamma \mid > \epsilon) \leq \frac{\epsilon^2}{64\delta_1^{*2}} + o(\epsilon^2).$$

Therefore, $P(|\hat{V}_{\text{PS}} - \gamma| > \epsilon) \to 0$ as $n \to \infty$. That is, $\hat{V}_{\text{PS}}$ is consistent for $\text{Var}(U_n)$.

S3 Expression of jackknife variance estimator using pseudo-kernel of degree one

Consider a degree-one pseudo-kernel function defined as

$$\begin{aligned}\phi_{\text{PS}}(x_i) &= \binom{n}{1} U(x_1, \ldots, x_n) - \binom{n-1}{1} U(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n) \\ &= nU_n - (n-1)U_{n-1}^{(-i)}.\end{aligned}$$

Let $\bar{\phi}_{\text{PS}} = (1/n)\sum_{i=1}^{n} \phi_{\text{PS}}(X_i)$. The jackknife variance estimator is defined as $\hat{V}_J = \frac{1}{n(n-1)} \sum_{i=1}^{n} (\phi_{\text{PS}}(X_i) - \bar{\phi}_{\text{PS}})^2$. It can be re-expressed in the following way:

$$\begin{aligned}\hat{V}_J &= \frac{1}{n(n-1)} \sum_{i=1}^{n} \left(\phi_{\text{PS}}(X_i)^2 - \frac{2}{n}\phi_{\text{PS}}(X_i) \sum_{j=1}^{n} \phi_{\text{PS}}(X_j) + \frac{1}{n^2} \sum_{i=1}^{n}\sum_{j=1}^{n} \phi_{\text{PS}}(X_i)\phi_{\text{PS}}(X_j) \right) \\ &= \frac{1}{n(n-1)} \left(\sum_{i=1}^{n} \phi_{\text{PS}}(X_i)^2 - \frac{1}{n} \sum_{i=1}^{n}\sum_{j=1}^{n} \phi_{\text{PS}}(X_i)\phi_{\text{PS}}(X_j) \right) \\ &= \frac{1}{n} \sum_{1 \le i,j \le n} \phi_{\text{PS}}(X_i)\phi_{\text{PS}}(X_j) - \frac{1}{n(n-1)} \sum_{i \neq j} \phi_{\text{PS}}(X_i)\phi_{\text{PS}}(X_j) \\ &= (\bar{\phi}_{\text{PS}})^2 - \frac{1}{n(n-1)} \sum_{i \neq j} \phi_{\text{PS}}(X_i)\phi_{\text{PS}}(X_j) \\ &= \hat{V}_u \text{ the unbiased variance estimator } Q(m) - Q(0)\end{aligned}$$

This result agrees with Comment 1 in Efron and Stein (1981), i.e. when the statistic is a linear functional (assuming the true kernel of the U-statistic is of order one) the jackknife variance estimator is unbiased.

S4 Proof of Theorem 3

Using the orthogonal terms in Hoeffding decomposition (Hoeffding (1948)) as defined in Appendix A2, we have

$$\begin{aligned}
U_n &= \theta + \binom{n}{k}^{-1} \sum_{c=1}^{k} \binom{n-c}{k-c} \sum_{(n,c)} h^{(c)}(x_{\nu_1}, \ldots, x_{\nu_c}) \\
U_{n-1}^{(-i_1)} &= \theta + \binom{n-1}{k}^{-1} \sum_{c=1}^{k} \binom{n-1-c}{k-c} \sum_{(n-1^{(-i_1)},c)} h^{(c)}(x_{\nu_1}, \ldots, x_{\nu_c}) \\
U_{n-2}^{(-i_1,-i_2)} &= \theta + \binom{n-2}{k}^{-1} \sum_{c=1}^{k} \binom{n-2-c}{k-c} \sum_{(n-2^{(-i_1,-i_2)},c)} h^{(c)}(x_{\nu_1}, \ldots, x_{\nu_c}) \\
U_{n-3}^{(-i_1,-i_2,-i_3)} &= \theta + \binom{n-3}{k}^{-1} \sum_{c=1}^{k} \binom{n-3-c}{k-c} \sum_{(n-3^{(-i_1,-i_2,-i_3)},c)} h^{(c)}(x_{\nu_1}, \ldots, x_{\nu_c})
\end{aligned}$$

The notation $\sum_{(n,c)}$ indicates the summation is taken over all subsets of size c taken out of $\mathcal{X}_n$. The pseudo-kernel of degree three can be expressed as

$$\begin{aligned}
\phi_{\mathrm{PS}}(x_{i_1}, x_{i_2}, x_{i_3}) &= \binom{n}{3}\left[\theta + \binom{n}{k}^{-1} \sum_{j=1}^{k} \binom{n-j}{k-j} \sum_{(n,j)} h^{(j)}(x_{\nu_1}, \ldots, x_{\nu_j})\right] \\
&- \binom{n-1}{3}\left[\theta + \binom{n-1}{k}^{-1} \sum_{c=1}^{k} \binom{n-1-c}{k-c} \sum_{(n-1^{(-i_1)},c)} h^{(c)}(x_{\nu_1}, \ldots, x_{\nu_c})\right] \\
&- \binom{n-1}{3}\left[\theta + \binom{n-1}{k}^{-1} \sum_{c=1}^{k} \binom{n-1-c}{k-c} \sum_{(n-1^{(-i_2)},c)} h^{(c)}(x_{\nu_1}, \ldots, x_{\nu_c})\right] \\
&- \binom{n-1}{3}\left[\theta + \binom{n-1}{k}^{-1} \sum_{c=1}^{k} \binom{n-1-c}{k-c} \sum_{(n-1^{(-i_3)},c)} h^{(c)}(x_{\nu_1}, \ldots, x_{\nu_c})\right]
\end{aligned}$$

$$
\begin{aligned}
+ & \binom{n-2}{3}\left[\theta+\binom{n-2}{k}^{-1}\sum_{c=1}^{k}\binom{n-2-c}{k-c}\sum_{(n-2^{(-i_1,-i_2)},c)} h^{(c)}(x_{\nu_1},\ldots,x_{\nu_c})\right]\\
+ & \binom{n-2}{3}\left[\theta+\binom{n-2}{k}^{-1}\sum_{c=1}^{k}\binom{n-2-c}{k-c}\sum_{(n-2^{(-i_1,-i_3)},c)} h^{(c)}(x_{\nu_1},\ldots,x_{\nu_c})\right]\\
+ & \binom{n-2}{3}\left[\theta+\binom{n-2}{k}^{-1}\sum_{c=1}^{k}\binom{n-2-c}{k-c}\sum_{(n-2^{(-i_2,-i_3)},c)} h^{(c)}(x_{\nu_1},\ldots,x_{\nu_c})\right]\\
- & \binom{n-3}{3}\left[\theta+\binom{n-3}{k}^{-1}\sum_{c=1}^{k}\binom{n-3-c}{k-c}\sum_{(n-3^{(-i_1,-i_2,-i_3)},c)} h^{(c)}(x_{\nu_1},\ldots,x_{\nu_c})\right]
\end{aligned}
$$

Following straightforward but tedious algebra simplifications, we can re-express $\phi_{\text{PS}}(x_{i_1},x_{i_2},x_{i_3})$ as

$$
\begin{aligned}
\phi_{\text{PS}}(x_{i_1},x_{i_2},x_{i_3}) &= \theta+\binom{k}{1}(1/3)[h^{(1)}(x_{i_1})+h^{(1)}(x_{i_2})+h^{(1)}(x_{i_3})]\\
&+ \binom{k}{2}(1/3)[h^{(2)}(x_{i_1},x_{i_2})+h^{(2)}(x_{i_1},x_{i_3})+h^{(2)}(x_{i_2},x_{i_3})]\\
&+ \binom{k}{3}h^{(3)}(x_{i_1},x_{i_2},x_{i_3})+\text{remainder}
\end{aligned}
$$

where the reminder term only depends on $h^{(c)}$ for $4 \le c \le k$.

As a result, we have

$$
\begin{aligned}
U_{\text{PS}} &= \binom{n}{2}^{-1}\sum_{1\le i_1<i_2<i_3\le n}\phi_{\text{PS}}(x_{i_1},x_{i_2},x_{i_3})\\
&= \theta+\frac{k}{n}\sum_{i=1}^{n}h^{(1)}(x_i)+\binom{n}{k}^{-1}\binom{n-2}{k-2}\sum_{(n,2)}h^{(2)}(x_{i_1},x_{i_2})\\
&+ \binom{n}{k}^{-1}\binom{n-3}{k-3}\sum_{(n,3)}h^{(3)}(x_{i_1},x_{i_2},x_{i_3})+\text{remainder.}
\end{aligned}
$$

From here it's easy to show that the variance of the remainder terms is of order n^{-4}. Thus, $\hat{V}_{\text{PS}}$ is third-order unbiased.