BIAS, CONSISTENCY, AND ALTERNATIVE PERSPECTIVES OF THE INFINITESIMAL JACKKNIFE

Wei Peng, Lucas Mentch*, and Len Stefanski*

University of Pittsburgh and North Carolina State University*

Supplementary Material

A. IJ via OLS Linear Regression

In what follows, we outline in detail the connection between OLS linear regression and the infinitesimal jackknife. In particular, we show how the infinitesimal jackknife estimator of variance of bagged estimates derived recently by Efron [12] can equivalently be obtained via a straightforward linear regression of the bootstrap estimates on their respective sampling weights. We begin with some general preliminary results for a general resampling setup and then transition into specific findings for the bootstrap regime.

Remark A.1. Appendix A should be read as a standalone section. In particular, because our goal is to cast everything in a familiar regression context, the notation used here differs slightly in some instances from that utilized in the main text and in the remaining appendices.

Suppose we have a sample $Z_1, ..., Z_n \sim P$ with realized observations $\boldsymbol{z} = (z_1, ..., z_n)$ from which we construct an estimator $\hat{\theta} = s(\boldsymbol{z})$ for some parameter of interest θ. Let $\mathcal{D}_n = [Z_1 \cdots Z_n]^T$ denote the original data matrix.

Consider a general resampling setup and let $\boldsymbol{z}_1^*, ..., \boldsymbol{z}_B^*$ denote B resamples of the original data that are used to construct the corresponding estimates $\hat{\theta}_1, ..., \hat{\theta}_B$. Let $w_b = (w_{b,1}, ..., w_{b,n})^T$ denote the associated resampling weights that count the number of times each observation (row) in $\mathcal{D}_n$ appears in each resample. That is, $w_{b,j} = c$ indicates that the j^{th} sample (row) of $\mathcal{D}_n$ appears exactly c times in the b^{th} resample. Denote the average across these resampled estimates by

$$\tilde{\theta} = \frac{1}{B}\sum_{b=1}^{B}\hat{\theta}_b$$

so that $\tilde{\theta}$ corresponds to the standard bagged estimate of θ whenever bootstrapping is the particular kind of resampling employed. Our primary goal in the following subsection is to derive a closed form estimate for $\text{Var}(\tilde{\theta})$ in the bootstrap regime. We begin by deriving some more general preliminary results.

First note that by the law of total variance, we can write

$$\begin{aligned}\text{Var}(\tilde{\theta}) &= \mathbb{E}\left[\text{Var}\left(\tilde{\theta}|\mathcal{D}_n\right)\right] + \text{Var}\left[\mathbb{E}\left(\tilde{\theta}|\mathcal{D}_n\right)\right] \\ &\overset{B}{\approx} \text{Var}\left[\mathbb{E}\left(\tilde{\theta}|\mathcal{D}_n\right)\right]\end{aligned}$$

since $\text{Var}\left(\tilde{\theta}|\mathcal{D}_n\right) \to 0$ as $B \to \infty$. Further, we have that

$$\mathbb{E}\left(\tilde{\theta}|\mathcal{D}_n\right) = \frac{1}{B}\sum_{b=1}^{B}\mathbb{E}\left(\hat{\theta}_b\middle|\mathcal{D}_n\right)$$

Let $\gamma_b = \mathbb{E}\left(\hat{\theta}_b\middle|\mathcal{D}_n\right)$. Since $\theta_1, \ldots, \theta_B$ are identically distributed conditional on $\mathcal{D}_n$, we have

$$\text{Var}(\tilde{\theta}) \approx \text{Var}(\gamma_b) \tag{A.1}$$

We close this Section with a final key observation: in this setup, conditional on $\mathcal{D}_n$, for each resample $b = 1, ..., B$ we can write

$$\hat{\theta}_b = g(w_b)$$

for some (unknown) function g. Thus, in order to investigate the properties of the resampled estimates, we need only understand how g depends on w_b.

The Bootstrap Setting

We now narrow our focus to the bootstrap regime where B equally-weighted resamples of size n are independently taken from the rows of $\mathcal{D}_n$ with replacement so that each weight vector w_b is thus distributed as $Multinomial(n; \frac{1}{n}, ..., \frac{1}{n})$. Now note that conditional on $\mathcal{D}_n$, for each resample $b = 1, ..., B$ we can write

$$\hat{\theta}_b = g(w_b)$$

so that $g(1_n)$ gives the estimate based on all original observations $\hat{\theta}$. Importantly, this means that in order to investigate the properties of the bootstrap estimates, we need only understand how g depends on the weights w_b. For each bootstrap replicate b, we can write

$$\hat{\theta}_b - \hat{\theta} = g(w_b) - g(1_n).$$

Now, if we assume that g is differentiable, then a first-order Taylor approximation to $\hat{\theta}_b - \hat{\theta}$ is given by

$$g(1_n)^T(w_b - 1_n).$$

Absent this differentiability assumption, we could alternatively consider modeling the underlying relationship g linearly via

$$\hat{\theta}_b - \hat{\theta} = \beta^T(w_b - 1_n) + \epsilon_b \qquad \text{for } b = 1, ..., B. \tag{A.2}$$

Taking this approach, the ordinary least squares estimate for β is given by

$$\begin{aligned}\hat{\beta}_{\text{OLS}} = \hat{\beta} &= \arg\min_{\beta} \sum_{b=1}^{B} \left(\hat{\theta}_b - \hat{\theta} - \beta^T(w_b - 1_n)\right)^2 \\ &= \left(\frac{1}{B-1}X^TX\right)^{-1}\left(\frac{1}{B-1}X^TY\right) \end{aligned} \tag{A.3}$$

where $Y = (\hat{\theta}_1 - \hat{\theta}, ..., \hat{\theta}_B - \hat{\theta})^T$ and $X = ((w_1 - 1_n)^T, ..., (w_B - 1_n)^T)^T$ correspond to the (centered) bootstrap estimates and weights, respectively.

Recall from (A.1) that

$$\text{Var}(\tilde{\theta}) \approx \text{Var}(\gamma_b)$$

where $\gamma_b = \mathbb{E}\left(\hat{\theta}_b \middle| \mathcal{D}_n\right)$. The operation of $\mathbb{E}(\cdot|\mathcal{D}_n)$ serves to smooth $\hat{\theta}_b$, as does the linear approximation in Eq. (A.2). We thus use $\hat{\gamma}_b = \hat{\theta} + \beta^T(w_b - 1_n)$ for $b = 1, \ldots B$ as approximations of γ_b. Now $\text{Var}(\gamma_b)$ can be estimated by the sample variance of $\hat{\gamma}_1, \ldots, \hat{\gamma}_B$.

Further, note that since $w_1, ..., w_B \overset{iid}{\sim} Multinomial(n; \frac{1}{n}, ..., \frac{1}{n})$, each $w_{b,i} \sim Binomial(n, \frac{1}{n})$ so that $\mathbb{E}(w_{b,i}) = 1$ for all $b = 1, ..., B$ and all $i = 1, ..., n$. Thus, estimating the variance of $\tilde{\theta}$ with the sample variance of $\{\hat{\gamma}_1, ..., \hat{\gamma}_B\}$, we have

$$\begin{aligned}\widehat{\text{Var}}(\tilde{\theta}) &\approx \frac{1}{B-1}\sum_{b=1}^{B}\left(\hat{\beta}^T(w_b - 1)\right)^2 \\ &= \hat{\beta}^T\left[\frac{1}{B-1}\sum_{b=1}^{B}(w_b - 1)(w_b - 1)^T\right]\hat{\beta}.\end{aligned}$$

Looking at the middle term, we have

$$\begin{aligned}\left[\frac{1}{B-1}\sum_{b=1}^{B}(w_b-1)(w_b-1)^T\right] &\overset{B}{\approx} \mathbb{E}\left((w_1-1)(w_1-1)^T\right)\\ &= I_n - \frac{1}{n}1_{n\times n}\\ &\overset{n}{\approx} I_n\end{aligned}$$

and thus,

$$\widehat{\mathrm{Var}}(\tilde{\theta}) \approx \hat{\beta}^T I_n \hat{\beta} = \sum_{j=1}^{n} \hat{\beta}_j^2. \tag{A.4}$$

Thus, in order to produce our estimate for $\mathrm{Var}(\tilde{\theta})$, it remains only to work out the solution to $\hat{\beta}$ given in (A.3).

Let's begin by considering the expectation of the inverse of the first term in (A.3). Observe that for $i \neq j$

$$\begin{aligned}\left[\mathbb{E}\left(\frac{1}{B-1}X^TX\right)\right]_{i,j} &= \frac{1}{B-1}\mathbb{E}\left(\sum_{k=1}^{B}(w_{k,i}-1)(w_{k,j}-1)\right)\\ &= \frac{1}{B-1}\sum_{k=1}^{B}\mathbb{E}(w_{k,i}-1)(w_{k,j}-1)\\ &= \frac{1}{B-1}\sum_{k=1}^{B}\mathrm{Cov}(w_{k,i},w_{k,j})\\ &= \frac{B}{B-1}(-n)\left(\frac{1}{n}\right)\left(\frac{1}{n}\right)\\ &= \left(\frac{B}{B-1}\right)\frac{-1}{n}\end{aligned}$$

where the third equality comes from the fact that each $w_{b,i} \sim Binomial(n, \frac{1}{n})$ and the fourth and fifth equalities follow from $w_1, ..., w_B \overset{iid}{\sim} Multinomial(n; \frac{1}{n}, ..., \frac{1}{n})$. For the diagonal elements

$(i = j)$, the covariance terms above become variance terms so that

$$\frac{B}{B-1}\text{Cov}(w_{1,i}, w_{1,j}) = \frac{B}{B-1}\text{Var}(w_{1,i}) = \left(\frac{B}{B-1}\right)\frac{n-1}{n}$$

and thus, in matrix form, we have

$$\mathbb{E}\left(\frac{1}{B-1}X^T X\right) = -\frac{1}{n}1_{n\times n} + I_n.$$

Finally, note that

$$\left(\frac{1}{B-1}X^T X\right) \overset{B}{\approx} \mathbb{E}\left(\frac{1}{B-1}X^T X\right) \overset{n}{\approx} I_n \tag{A.5}$$

and thus

$$\hat{\beta} = \left(\frac{1}{B-1}X^T X\right)^{-1}\left(\frac{1}{B-1}X^T Y\right) \approx I_n^{-1}\left(\frac{1}{B-1}X^T Y\right) = \frac{1}{B-1}X^T Y.$$

Now,

$$\begin{aligned}\frac{1}{B-1}X^T Y &= \frac{1}{B-1}\left((w_1 - 1_n)^T, ..., (w_B - 1_n)^T\right)\left((\hat{\theta}_1 - \hat{\theta}), ..., (\hat{\theta}_B - \hat{\theta})\right)^T \\ &= \left(\frac{1}{B-1}\sum_{b=1}^{B}(w_{b,1} - 1)(\hat{\theta}_b - \hat{\theta}),\ \dots,\ \frac{1}{B-1}\sum_{b=1}^{B}(w_{b,n} - 1)(\hat{\theta}_b - \hat{\theta})\right)^T\end{aligned}$$

so that element-wise,

$$\hat{\beta}_j = \frac{1}{B-1}\sum_{b=1}^{B}(w_{b,j} - 1)(\hat{\theta}_b - \hat{\theta})$$

and since $\mathbb{E}(w_{b,j}) = 1$, $\hat{\beta}_j$ is effectively the sample covariance of $(w_{1,j}, ..., w_{B,j})$ and $(\hat{\theta}_1, ..., \hat{\theta}_B)$.

Denoting this by sample covariance by $\widehat{\text{Cov}}_j$ and putting this together with (A.4), we have

$$\widehat{\text{Var}}(\tilde{\theta}) \approx \sum_{j=1}^{n}\hat{\beta}_j^2 = \sum_{j=1}^{n}\widehat{\text{Cov}}_j^2 \tag{A.6}$$

which coincides exactly with the infinitesimal jackknife variance estimate given by Efron in [12].

B. Proofs and Calculations for $\mathrm{IJ_B}$ (IJ for Bootstrap)

Proof of Theorem 1:

1. By definition,

$$\begin{aligned}
\mathbb{E}_*[s^* w_j^*] &= \sum_{w_1^*+\cdots+w_n^*=n} p(w_1^*,\ldots,w_n^*) s(X_1^*,\ldots,X_n^*) w_j^* \\
&= \sum_{\substack{w_j^* \geq 1 \\ w_1^*+\cdots+w_n^*=n}} \frac{(n-1)!}{w_1^* \ldots ((w_j^*-1)!) \cdots (w_n^*)!} \frac{1}{n^{n-1}} s(X_1^*,\ldots,X_n^*) \\
&= \mathbb{E}_*[s(X_1^*,\ldots,X_n^*) | X_1^* = X_j] \\
&= e_j.
\end{aligned}$$

2. Conditional on the data, knowing $X_1^*,\ldots,X_n^*$ is equivalent to knowing $w_1^*,\ldots,w_n^*$. Therefore, l^* can be also viewed as the projection of $s^* - \mathbb{E}_*[s^*]$ onto the linear space spanned by $X_1^*,\ldots,X_n^*$. Then we have

$$\begin{aligned}
l^* &= \sum_i (\mathbb{E}_*[s^*|X_i^*] - \mathbb{E}_*[s^*]) \\
&= \sum_i \sum_j (\mathbb{E}_*[s^*|X_i^* = X_j] - \mathbb{E}_*[s^*]) \mathbf{1}_{\{X_i^*=X_j\}} \\
&= \sum_i \sum_j (e_j - s_0) \mathbf{1}_{\{X_i^*=X_j\}} \\
&= \sum_j \sum_i (e_j - s_0) \mathbf{1}_{\{X_i^*=X_j\}} \\
&= \sum_j w_j^* (e_j - s_0)
\end{aligned}$$

as desired.

3. By 1 above, $\mathrm{IJ_B} = \sum_j \mathrm{Cov}_*^2(s^*, w_j^*) = \sum_j (\mathbb{E}_*[s^* w_j^*] - \mathbb{E}_*[s^*]\mathbb{E}_*[w_j^*])^2 = \sum_j (e_j - s_0)^2$. By 2, we have $\mathrm{Var}_*(l^*) = \mathrm{Var}_*(\sum_j (w_j^* - 1)(e_j - s_0)) = \sum_j (e_j - s_0)^2$. Thus, $\mathrm{Var}_*(l^*) = \mathrm{JK}_\mathrm{B}^\sharp = \mathrm{IJ_B}$.

■

Example Calculations:

Example 1: Sample Mean Consider $s = s(x_1, \ldots, x_n) = \frac{1}{n}\sum_{i=1}^n x_i$. We have

$$s^* = \frac{1}{n}\sum_{i=1}^n X_i^* \quad \text{and} \quad l^* = \sum_{i=1}^n \sum_{j=1}^n \mathbf{1}_{X_i^* = X_j}(e_j - s_0). \tag{B.7}$$

Then, $\mathbb{E}_*[s^*] = \frac{1}{n}\sum_{i=1}^n X_i$ and $\mathrm{Var}_*(l^*) = \frac{1}{n^2}\sum_{i=1}^n (X_i - \bar{X})^2$. Therefore,

$$\mathrm{Var}(\mathbb{E}_*[s^*]) = \sigma^2/n, \quad \mathbb{E}[\mathrm{Var}_*(l^*)] = (n-1)\sigma^2/n^2 \tag{B.8}$$

and thus, we have $\frac{\mathbb{E}[\mathrm{Var}_*(l^*)]}{\mathrm{Var}(\mathbb{E}_*[s^*])} = \frac{n-1}{n} \to 1$ as $n \to \infty$. In Figures 1 and 2, $X_1, ..., X_n$ follow $\mathcal{N}(0, \sigma^2)$ where $n = 100$ and $\sigma^2 = 1$. Since we know that $\mathrm{Var}(\mathbb{E}_*[s^*]) = \sigma^2/n$, an oracle estimate would be $\widehat{\sigma^2}/n$, where $\widehat{\sigma^2}$ is the sample variance. The gray dashed line denotes the true value of $\mathrm{Var}(\mathbb{E}_*[s^*])$. We find that $\widehat{\mathrm{IJ}}_\mathrm{B}^{mc}$ and $\widehat{\mathrm{IJ}}_\mathrm{B}^{whe}$ are quite close as expected and both perform well. The original $\widehat{\mathrm{IJ}}_\mathrm{B}$ seems to overestimate substantially when $B = 100$.

Example 2: Sample Variance Consider $s = \binom{n}{2}^{-1} \sum_{i<j} (x_i - x_j)^2$. We have

$$\mathbb{E}_*[s^*] = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2 \quad \text{and} \quad \mathrm{Var}_*(l^*) = \frac{1}{n^2}\sum_i \left[(X_i - \bar{X})^2 - \frac{1}{n}\sum_i (X_i - \bar{X})^2\right]^2.$$

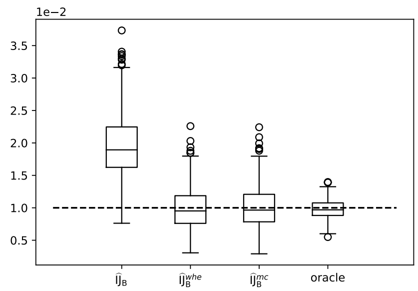

Figure 1: Performance of the infinitesimal jackknife and its bias-corrected alternatives on estimating the variance of the bagged sample mean (B=100).

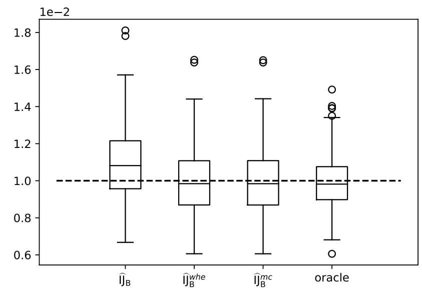

Figure 2: Performance of the infinitesimal jackknife and its bias-corrected alternatives on estimating the variance of the bagged sample mean (B=1000).

Then we have

$$\begin{aligned}\mathrm{Var}(\mathbb{E}_*[s^*]) &= \left(\frac{n-1}{n}\right)^2 \left[\frac{\mu_4}{n} - \frac{\mu_2^2}{n}\frac{n-3}{n-1}\right] \\ &= a_n\mu_4 - b_n\mu_2^2,\end{aligned} \tag{B.9}$$

where μ_i is the ith central moment of X for $i = 2, 4$. Let $\mathbf{X} = (X_1, \dots, X_n)^T$, then $\mathbb{E}[\mathrm{Var}_*(l^*)]$ can be written as $\frac{1}{n}\mathbb{E}[\mathbf{X}^T \mathrm{A}\mathbf{X}]^2$, where $\mathrm{A} = \Sigma_1 - \frac{1}{n}\sum_i \Sigma_i$, $\Sigma_i = (e_i - \frac{1}{n}\mathbf{1}_n)(e_i^T - \frac{1}{n}\mathbf{1}_n^T)$ and

$e_i = (0, \ldots, 0, 1, 0, \ldots, 0)$. After some calculation, we obtain

$$\begin{aligned}
&\mathbb{E}[\mathrm{Var}_*(l^*)] \\
&= \frac{(n-1)}{n^2}\left[\mathbb{E}[(X_1-\bar{X})^4] - \mathbb{E}[(X_1-\bar{X})^2(X_2-\bar{X})^2]\right] \\
&= \left(\frac{n-1}{n}\right)^2\left[\left(\frac{n^3-(n-1)^2}{n^2(n-1)^2} + \frac{n}{(n-1)^5}\right)\mu_4 - \left(\frac{n^2-2n+3}{(n-1)n^2} - \frac{3n^2(2n-3)}{(n-1)^5}\right)\mu_2^2\right] \\
&= a'_n\mu_4 - b'_n\mu_2^2.
\end{aligned}$$

Thus, we have

$$\begin{aligned}
\frac{a'_n}{a_n} &= 1 + \frac{n^2+n-1}{n(n-1)^2} + \frac{n^2}{(n-1)^5} = 1 + \frac{1}{n} + o\left(\frac{1}{n}\right) \\
\frac{b'_n}{b_n} &= 1 + \frac{n+3}{n(n-3)} - \frac{3n^3(2n-3)}{(n-1)^4(n-3)} = 1 - \frac{5}{n} + o\left(\frac{1}{n}\right).
\end{aligned}$$

Since $a'_n/a_n \to 1$ and $b'_n/b_n \to 1$ as $n \to \infty$, we have $\frac{\mathbb{E}[\mathrm{Var}_*(l^*)]}{\mathrm{Var}(\mathbb{E}_*[s^*])} \to 1$. $\mathrm{IJ_B}$ is therefore asymptotically unbiased for estimating the variance of the sample variance. Since the sample variance is close to a linear statistic, the result is not surprising. In Figures 3 and 4, $X_1, ..., X_n$ follow $\mathcal{N}(0, \sigma^2)$ where $n = 100$ and $\sigma^2 = 1$. Since we know $\mathrm{Var}(\mathbb{E}_*[s^*]) = 2\sigma^4/n$, an oracle estimate would be $2(\widehat{\sigma^2})^2/n$, where $\widehat{\sigma^2}$ is the sample variance. The gray dashed line denotes the true value of $\mathrm{Var}(\mathbb{E}_*[s^*])$. As in the first example, $\widehat{\mathrm{IJ}}_\mathrm{B}^{mc}$ and $\widehat{\mathrm{IJ}}_\mathrm{B}^{whe}$ are both quite close and perform well. The original $\widehat{\mathrm{IJ}}_\mathrm{B}$ again seems to suffer from overestimation when $B = 100$.

Example 3: Sample Maximum Consider $s = \max\{X_1, ..., X_n\}$, where $X_1, \ldots, X_n$ are uniformly distributed in $[0, 1]$. For the order statistics $X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)}$, we have

$$cov(X_{(i)}, X_{(j)}) = \frac{i(n-j+1)}{(n+1)^2(n+2)} \quad \text{and} \quad \mathbb{E}[X_{(i)}X_{(j)}] = \frac{i(j+1)}{(n+1)(n+2)}. \tag{B.10}$$

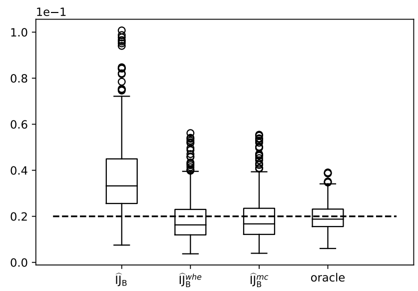

Figure 3: Performance of the infinitesimal jackknife and its bias-corrected alternatives on estimating the variance of the bagged sample variance (B=100).

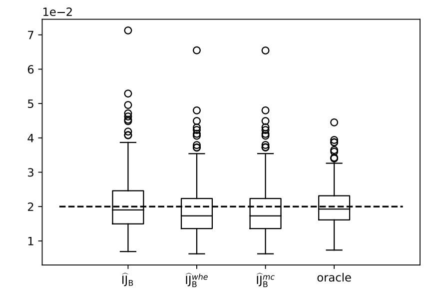

Figure 4: Performance of the infinitesimal jackknife and its bias-corrected alternatives on estimating the variance of the bagged sample variance (B=1000).

Note that

$$\mathbb{E}_*[s^*] = s_0 = \sum_{i=1}^{n} X_{(i)} p_i^n,$$

where $p_i^n = q_i^n - q_{i-1}^n$ and $q_i^n = \left(\frac{i}{n}\right)^n$ for $i = 1, \ldots, n$. Thus,

$$\mathrm{Var}(\mathbb{E}_*[s^*]) = v^T \mathrm{A} v \tag{B.11}$$

where $\mathrm{A} = cov(\mathbf{u}) = \left[\frac{i(n+1-j)}{(n+1)^2(n+2)}\right]_{ij}$ and $v = (p_1^n, \cdots, p_n^n)$. Let

$$\tilde{e}_i = \sum_{j=I+1}^{n} X_{(j)} p_j^{n-1} + X_{(i)} q_i^{n-1}, \quad \text{where } q_i^{n-1} = \left(\frac{i}{n}\right)^{n-1}.$$

We have $\mathrm{Var}_*(l^*) = \sum_{i=1}^{n} (e_i - s_0)^2 = \sum_{i=1}^{n} (\tilde{e}_i - s_0)^2$. Thus,

$$\mathbb{E}[\mathrm{Var}_*(l^*)] = \sum (v_i - v)^T \mathrm{B} (v_i - v) \tag{B.12}$$

where $\mathrm{B} = \mathbb{E}[\mathbf{u}\mathbf{u}^T] = \left[\frac{i(j+1)}{(n+1)(n+2)}\right]_{ij}$ and $v_i = (\cdots, 0, \cdots, q_i^{n-1}, \cdots, p_j^{n-1}, \cdots)$. Next, we have

$$
\begin{aligned}
v^T\mathrm{A}v == &\frac{1}{(n+1)^2(n+2)}\sum_i\sum_j p_i^n p_j^n i(n+1-j) \\
&= \frac{1}{(n+1)^2(n+2)}\left(\sum_i i\cdot p_i^n\right)\left(\sum_j (n-j+1)p_j^n\right) \\
&= \frac{1}{(n+1)^2(n+2)}\left(\sum_i i\cdot p_i^n\right)\left(n+1-\sum_i i\cdot p_i^n\right) \\
&= \frac{1}{(n+1)^2(n+2)}(n-\sum_j(\frac{j-1}{n})^n))(1+\sum_j(\frac{j-1}{n})^n))
\end{aligned}
\tag{B.13}
$$

by the fact that

$$
\sum_{i=1}^n i\cdot p_i^n = \sum_{i=1}^n iq_i^n - \sum_{i=0}^{n-1}(i+1)q_i^n = n - \sum_{i=0}^{n-1} q_i^n. \tag{B.14}
$$

Now, let $\mathbf{e}_n = [1, 2, \cdots, n]^T$. Then

$$
\begin{aligned}
(v_i - v)^T\mathrm{B}(v_i - v) &= \frac{(v_i - v)^T\mathbf{e}_n\cdot(\mathbf{e}_n^T + 1_n^T)\mathrm{V}}{(n+1)(n+2)} \\
&= \frac{(v_i - v)\mathbf{e}_n\cdot\mathbf{e}_n^T(v_i - v)}{(n+1)(n+2)} \\
&= \frac{1}{(n+1)(n+2)}\sum_i\left[(n-\sum_{j=I}^{n-1}(\frac{j}{n})^{n-1}) - (n-\sum_{j=0}^{n-1}(\frac{j}{n})^n)\right]^2 \\
&= \frac{1}{(n+1)(n+2)}\sum_i\left[\sum_{j=1}^n(\frac{j-1}{n})^n - \sum_{j=i+1}^n(\frac{j-1}{n})^{n-1}\right]^2.
\end{aligned}
\tag{B.15}
$$

In summary, we have

$$
\begin{aligned}
\frac{\mathbb{E}[\mathrm{Var}_*(l^*)]}{\mathrm{Var}(\mathbb{E}_*[s^*])} &= \frac{(n+1)\sum_i[\sum_{j=1}^n(\frac{j-1}{n})^n - \sum_{j=i+1}^n(\frac{j-1}{n})^{n-1}]^2}{(n-\sum_j(\frac{j-1}{n})^n)(1+\sum_j(\frac{j-1}{n})^n)} \\
&\to c \in [0.24, 0.25] \quad \text{as } n \to \infty.
\end{aligned}
\tag{B.16}
$$

■

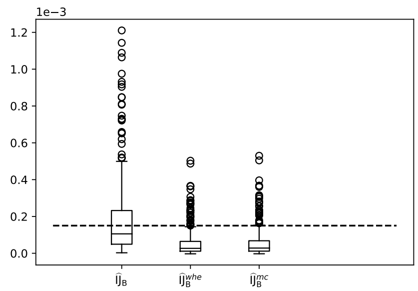

Figure 5: Performance of the infinitesimal jackknife and its bias-corrected alternatives on estimating the variance of the bagged sample maximum (B=100).

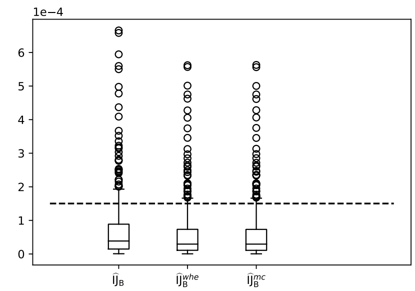

Figure 6: Performance of the infinitesimal jackknife and its bias-corrected alternatives on estimating the variance of the bagged sample maximum (B=1000).

Here we can see that $\mathrm{IJ_B}$ is underestimating of $\mathrm{Var}(\mathbb{E}_*[s^*])$ by a considerable margin. In this case, $\mathbb{E}_*[s^*]$ is not close to a linear statistic, so $\mathrm{IJ_B}$ should not be expected to perform well. In Figures 5 and 6, $X_1, \ldots, X_n$ follow Uniform(0, 1) and $n = 100$ with the dashed line corresponding to the true value of $\mathrm{Var}(\mathbb{E}_*[s^*])$. In this case, there is no obvious oracle estimator for $\mathbb{E}_*[s^*]$. Unlike the previous two examples, although $\widehat{\mathrm{IJ}}_{\mathrm{B}}^{mc}$ and $\widehat{\mathrm{IJ}}_{\mathrm{B}}^{whe}$ remain quite similar, all three estimators suffer from considerable underestimation even when $B = 1000$.

Proof of Proposition 1: Note that $\mathbb{E}[\mathrm{Var}_*(l^*)] = (n-1)\mathbb{E}[e_1^2 - e_1 e_2]$ and $\mathrm{Var}(\mathbb{E}_*[s^*]) =$

$\frac{1}{n}\text{Var}(e_1) + \frac{n-1}{n}cov(e_1, e_2)$. Let $\rho = cov(e_1, e_2)/\text{Var}(e_1)$. We have

$$\begin{aligned}
\mathbb{E}[\text{Var}_*(l^*)]/\text{Var}(\mathbb{E}_*[s^*]) &= \frac{(n-1)[\mathbb{E}[e_1^2] - \mathbb{E}^2[e_1] + \mathbb{E}[e_1]\mathbb{E}[e_2] - \mathbb{E}[e_1 e_2]]}{\frac{1}{n}\text{Var}(e_1) + \frac{n-1}{n}cov(e_1, e_2)} \\
&= \frac{(n-1)[\text{Var}(e_1) - cov(e_1, e_2)]}{\frac{1}{n}\text{Var}(e_1) + \frac{n-1}{n}cov(e_1, e_2)} \\
&= \frac{(n-1)(1-\rho)}{1/n + (n-1)/n \cdot \rho} \\
&= n\frac{1-\rho}{1/(n-1) + \rho}.
\end{aligned} \tag{B.17}$$

Let $f(\rho) = \frac{n(1-\rho)}{1/(n-1)+\rho}$. It is immediate that $f(\rho) \to 1$ if and only if $\rho = 1 - \frac{1}{n} + o\left(\frac{1}{n}\right)$. Thus, IJ_B is an asymptotically unbiased estimator of $\text{Var}(\mathbb{E}_*[s^*])$ if and only if $1-\rho = 1/n + o(1/n)$. ■

C. Proofs and Calculations for IJ_U and ps-IJ_U (IJ for U-statistics)

How does U depend on $\mathbb{P}_n$, such that $\text{U} = f(\mathbb{P}_n)$ for some f? The dependence is abstract so that the subsampling proceeds according to the probabilities determined by $\mathbb{P}_n$. Following directly from the original definition of the IJ, we arrive at the following theorem.

Theorem C.1. *The IJ estimator of the variance of a U-statistic is given by*

$$\text{IJ}_\text{U} = \frac{k^2}{n^2}\sum_{j=1}^{n}[\alpha e_j - \beta s_0]^2, \tag{C.18}$$

where $e_j = \mathbb{E}_*[s^*|X_1^* = X_j]$, $s_0 = \mathbb{E}_*[s^*]$ *and*

$$\alpha = 1 + \frac{1}{n}\left\{\frac{k-1}{2} - \frac{1}{k}\sum_{j=0}^{k-1}\frac{j^2}{(n-j)}\right\}, \quad \beta = 1 + \frac{1}{k}\sum_{j=0}^{k-1}\frac{j}{n-j}.$$

Proof. When subsampling without replacement, according to the weight of each sample, the probability of $(x_1, \dots, x_k)$ being selected is

$$\begin{cases} \sum_{i_1,\dots,i_k} \frac{\mathbb{P}_n(x_{i_1})}{1} \times \frac{\mathbb{P}_n(x_{i_2})}{1-\mathbb{P}_n(x_{i_1})} \times \cdots \times \frac{\mathbb{P}_n(x_{i_k})}{1-\sum_{j=1}^{k-1}\mathbb{P}_n(x_{i_j})}, & x_1, \dots x_k \in \mathcal{D}_n \text{ and are distinct.} \\ 0, & \text{otherwise.} \end{cases} \tag{C.19}$$

Note that any subsampling with a general re-weighting scheme can be derived similarly. Consider $f((1-\epsilon)\mathbb{P}_n + \epsilon\delta_{X_i})$ and let $\delta = 1-\epsilon$. We first provide the probability of obtaining $(x_1, \dots, x_k)$. On one hand, if $X_i \notin (x_1, x_2, \dots, x_k)$, then

$$p(x_1, x_2, \dots, x_k) = p_0 = \left[\frac{\delta}{n} \cdot \frac{\delta}{(n-\delta)} \cdots \frac{\delta}{(n-(k-1)\delta)}\right] \times k!. \tag{C.20}$$

On the other hand, if $X_i \in (x_1, \dots, x_k)$, then $p(x_1, x_2, \dots, x_k) = p_1 = \sum_{i=0}^{k-1} q_i$, where

$$\begin{aligned} q_0 &= \left[\frac{(n-(n-1)\delta)}{n} \cdot \frac{1}{n-1} \cdots \frac{1}{n-k+1}\right] \times (k-1)! \\ q_1 &= \left[\frac{\delta}{n} \cdot \frac{n-(n-1)\delta}{n-\delta} \cdot \frac{1}{n-2} \cdots \frac{1}{n-k+1}\right] \times (k-1)! \\ &\vdots \\ q_{k-1} &= \left[\frac{\delta}{n}\frac{\delta}{n-\delta} \cdots \frac{\delta}{n-(k-2)\delta} \cdot \frac{n-(n-1)\delta}{n-(k-1)\delta}\right] \times (k-1)!. \end{aligned} \tag{C.21}$$

Thus,

$$f((1-\epsilon)\mathbb{P}_n + \epsilon\delta_{X_i}) = \sum_{(n,k)} s(X_{i_1}, \dots, X_{i_k})(p_0 \mathbf{1}_{i \notin \{i_1,\dots,i_k\}} + p_1 \mathbf{1}_{i \in \{i_1,\dots,i_k\}}),$$

where the sum is taken over all $\binom{n}{k}$ of subsamples of size k. We have

$$\frac{1}{p} p_0'(\delta)|_{\delta=1} = -\left[\frac{0}{n} + \frac{1}{n-1} + \cdots + \frac{k-1}{n-(k-1)}\right] - k$$

and

$$
\begin{aligned}
\frac{1}{p}p_1' &= \frac{1}{p}\sum_{j=0}^{k-1} q_j'|_{\delta=1} \\
&= \frac{1}{k}\sum_{j=0}^{k-1}\left[(n-j-1) - \left[\frac{0}{n} + \frac{1}{n-1} + \frac{2}{n-2} + \cdots \frac{j}{n-j}\right]\right] \\
&= -\frac{1}{k}\left[\frac{0\cdot k}{n} + \frac{1\cdot(k-1)}{n-1} + \cdots + \frac{(k-1)\cdot 1}{n-(k-1)}\right] - \frac{k+1}{2} + n.
\end{aligned}
$$

Putting all together, we have

$$
\begin{aligned}
\mathrm{D}_i &= \lim_{\delta\to 1}\frac{f(\delta\mathbb{P}_n + (1-\delta)\delta_{X_i}) - f(\mathbb{P}_n)}{1-\delta} \\
&= \sum_{(n,k)} (p_0'\mathbf{1}_{w_i^*=0} + p_1'\mathbf{1}_{w_i^*=1}) s(X_{i_1},\ldots,X_{i_k}) \\
&= \sum_{(n,k)} p\left[\frac{p_0'}{p} + (\frac{p_1'}{p} - \frac{{p_0}'}{p})w_i^*\right] s(X_{i_1},\ldots,X_{i_k}) \\
&= \frac{k}{n}(\frac{p_1'}{p} - \frac{p_0'}{p})e_i + \frac{p_0'}{p}s_0,
\end{aligned}
\tag{C.22}
$$

where $p = \binom{n}{k}^{-1}$, $e_i = \mathbb{E}_*[s^*|X_1^* = X_i]$ and $s_0 = \mathbb{E}_*[s^*]$. And $*$ refers to the procedure of subsampling without replacement. Then the infinitesimal jackknife estimate for U-statistic is

$$
\begin{aligned}
\mathrm{IJ_U} &= \frac{1}{n^2}\sum_{j=1}^{n}\left[\frac{k}{n}(\frac{p_1'}{p} - \frac{p_0'}{p})e_j + \frac{p_0'}{p}s_0\right]^2 \\
&= \frac{k^2}{n^2}\sum_{j=1}^{n}\left[\frac{p_1' - p_0'}{np}e_j + \frac{p_0'}{kp}s_0\right]^2 \\
&= \frac{k^2}{n^2}\sum_{j=1}^{n}[\alpha e_j - \beta s_0]^2
\end{aligned}
\tag{C.23}
$$

where

$$
\alpha = (p_1' - p_0')/(np) = 1 + \frac{1}{n}\left\{\frac{k-1}{2} - \frac{1}{k}\sum_{j=0}^{k-1}\frac{j^2}{(n-j)}\right\}, \tag{C.24}
$$

and

$$
\beta = -p_0'/(kp) = 1 + \frac{1}{k}\sum_{j=0}^{k-1}\frac{j}{n-j}. \tag{C.25}
$$

□

To understand the bias of $\mathrm{IJ_U}$, we will use H-decomposition, setting it up by introducing following notation for kernels $s^1, \dots, s^k$ of degrees $1, \dots, k$. These kernels are defined recursively as follows

$$s^1(x_1) = s_1(x_1) \tag{C.26}$$

and

$$s^c(x_1, \dots, x_c) = s_c(x_1, x_2, ..., x_c) - \sum_{j=1}^{c} \sum_{i_1, \dots, i_j \in \{1, \dots, c\}} s^j(x_{i_1}, \dots, x_{i_j}) \tag{C.27}$$

where $s_c(x_1, \dots, x_c) = \mathbb{E}[s(x_1, \dots, x_c, X_{c+1}, \dots, X_k)] - \mathbb{E}[s]$. Let $V_j = \mathrm{Var}(s^j)$ for $j = 1, \dots, k$. Then $\mathbb{E}[\mathrm{IJ_U}]$ can be written as a linear combination of those V_j. In particular, we have the following theorem.

Theorem C.2. *Let* $\theta = \mathbb{E}[s]$ *and* $\mathrm{IJ_U}$ *be as defined in Eq.* (C.18)*. Then*

$$\mathbb{E}[\mathrm{IJ_U}] = \sum_{j=1}^{k} r_j \binom{k}{j}^2 \binom{n}{j}^{-1} V_j + \frac{k^2}{n}(\alpha - \beta)^2 \theta^2, \tag{C.28}$$

where

$$r_j = \frac{(n-k)^2}{n^2} \left[\frac{j}{1 - j/n} \alpha^2 \right] + \frac{k^2}{n}(\alpha - \beta)^2, \quad \text{for } j = 1, \dots, k. \tag{C.29}$$

Remark C.1. Note that $\mathrm{Var(U)} = \sum_{j=1}^{k} \binom{k}{j}^2 \binom{n}{j}^{-1} V_j$. If k is held fixed, then $\alpha, \beta \to 1$ and thus $r_j \to j$ for $j = 1, \dots, k$. Since in such case both $\mathrm{Var(U)}$ and $\mathbb{E}[\mathrm{IJ}_u]$ will be dominated by the V_1 term, $\mathrm{IJ_U}$ is asymptotically unbiased.

Proof of Theorem C.2: By definition,

$$\begin{aligned}
(\alpha e_1 - \beta s_0) = &-\beta\binom{n}{k}^{-1}\sum s(X_{i_1},\dots,X_{i_k};\nexists 1) \\
&+\left(\frac{\alpha\cdot(k-1)!}{(n-1)\dots(n-k+1)} - \frac{\beta\cdot(k-1)!k}{n\cdots(n-k+1)}\right)\sum s(X_{i_1},\dots,X_{i_k};\exists 1) \\
= &-(1-\frac{k}{n})\beta\binom{n-1}{k}^{-1}\sum s(X_{i_1},\dots,X_{i_k};\nexists 1) \\
&+(\alpha-\frac{k}{n}\beta)\binom{n-1}{k-1}^{-1}\sum s(X_{i_1},\dots,X_{i_k};\exists 1).
\end{aligned} \tag{C.30}$$

Note that according to H-decomposition,

$$s(x_1,\dots,x_k) = \mathbb{E}[s] + \sum_{j=1}^{k}\sum_{i_1,\dots,i_j\in\{1,\dots,j\}} s^j(x_{i_1},\dots,x_{i_j}).$$

Then

$$\begin{aligned}
(\alpha e_1 - \beta s_0) = &(\alpha-\beta)\theta - (1-\frac{k}{n})\beta\sum_{j=1}^{k}\binom{k}{j}\binom{n-1}{j}^{-1}\sum s^j(X_{i_1},\dots,X_{i_j};\nexists 1) \\
&+(\alpha-\frac{k}{n}\beta)\sum_{j=1}^{k-1}\binom{k-1}{j}\binom{n-1}{j}^{-1}\sum s^j(X_{i_1},\dots,X_{i_j};\nexists 1) \\
&+(\alpha-\frac{k}{n}\beta)\sum_{j=1}^{k}\binom{k-1}{j-1}\binom{n-1}{j-1}^{-1}\sum s^j(X_{i_1},\dots,X_{i_j};\exists 1) \\
:= &(\alpha-\beta)\theta + A_n + B_n,
\end{aligned} \tag{C.31}$$

where

$$A_n = \sum_{j=1}^{k}\left[(\alpha-\frac{k}{n}\beta)\binom{k-1}{j}\binom{n-1}{j}^{-1} - (1-\frac{k}{n})\beta\binom{k}{j}\binom{n-1}{j}^{-1}\right]\sum s^j(X_{i_1},\dots,X_{i_j};\nexists 1)$$

and

$$B_n = (\alpha-\frac{k}{n}\beta)\sum_{j=1}^{k}\binom{k-1}{j-1}\binom{n-1}{j-1}^{-1}\sum s^{(j)}(X_{i_1},\dots,X_{i_j};\exists 1).$$

Thus, we have

$$\begin{aligned}\mathbb{E}[A_n^2] &= \sum_{j=1}^{k}\left[(\frac{k}{j}-1)\alpha+(\frac{k}{n}-\frac{k}{j})\beta\right]^2\binom{k-1}{j-1}^2\binom{n-1}{j}^{-1}V_j \\ \mathbb{E}[B_n^2] &= (\alpha-\frac{k}{n}\beta)^2\sum_{j=1}^{k}\binom{k-1}{j-1}^2\binom{n-1}{j-1}^{-1}V_j,\end{aligned} \tag{C.32}$$

where $V_j = \mathrm{Var}(s^j)$. Since A_n and B_n are uncorrelated and have mean zero, we have

$$\begin{aligned}&\mathbb{E}\left[(\alpha e_1-\beta s_0)^2\right] \\ &= \mathbb{E}[A_n^2]+\mathbb{E}[B_n^2]+(\alpha-\beta)^2\theta^2 \\ &= (\alpha-\beta)^2\theta^2+\sum_{j=1}^{k}\binom{k-1}{j-1}^2\Lambda(j)V_j,\end{aligned}$$

where $\Lambda(j) = \left[\left[(\frac{k}{j}-1)\alpha+(\frac{k}{n}-\frac{k}{j})\beta\right]^2\binom{n-1}{j}^{-1}+(\alpha-\frac{k}{n}\beta)^2\binom{n-1}{j-1}^{-1}\right]$, for $j=1,\cdots,k$. Therefore,

$$\begin{aligned}\mathbb{E}[\mathrm{IJ_U}] &= \frac{k^2}{n^2}\sum\mathbb{E}[(\alpha e_j-\beta s_0)^2] \\ &= \frac{k^2}{n}\sum_{j=1}^{k}\binom{k-1}{j-1}^2\Lambda(j)V_j+\frac{k^2}{n}(\alpha-\beta)^2\theta^2.\end{aligned} \tag{C.33}$$

Recall that

$$\mathrm{Var(U)} = \sum_{j=1}^{k}\binom{k}{j}^2\binom{n}{j}^{-1}V_j. \tag{C.34}$$

We consider the ratio of the coefficient of V_j in $\mathbb{E}[\mathrm{IJ_U}]$ and that in Var(U) and obtain

$$\begin{aligned}r_j &= \frac{k^2}{n}\Lambda(j)\binom{k-1}{j-1}^2\binom{k}{j}^{-2}\binom{n}{j} \\ &= \frac{k^2}{n}\frac{j^2}{k^2}\Lambda(j)\binom{n}{j} \\ &= \frac{(n-k)^2}{n^2}\left[\frac{j}{1-j/n}\alpha^2\right]+\frac{k^2}{n}(\alpha-\beta)^2\end{aligned} \tag{C.35}$$

for $j=1,\ldots,k$. ∎

Proof of Theorem 2: The result is immediate by substituting α with 1 and β with 1 respectively in Theorem C.2. ■

Proof of Proposition 2: By definition,

$$\begin{aligned}
\mathrm{Cov}_*(s^*, w_j^*) &= \sum_{w_1^*+\cdots+w_n^*=k} p(w_1^*, \ldots, w_n^*)[s^* - s_0]w_j^* \\
&= \sum_{w_1^*+\cdots+w_n^*=k} p(w_1^*, \ldots, w_n^*)s^* w_j^* - \frac{k}{n}s_0 \\
&= \frac{k}{n} \sum_{w_j^*=1, w_1^*+\cdots+w_n^*=k} \frac{(k-1)!}{(n-1)\cdots(n-k+1)} s^* - \frac{k}{n}s_0 \qquad \text{(C.36)} \\
&= \frac{k}{n}\left[\mathbb{E}_*[s(X_1^*, \ldots, X_k^*)|X_1^* = X_j] - s_0\right] \\
&= \frac{k}{n}(e_j - s_0).
\end{aligned}$$

It follows that ps-IJ $= \sum \mathrm{Cov}_*^2(s^*, w_j^*) = \frac{k^2}{n^2}\sum(e_j - s_0)^2$. ■

To prove Theorems 3 and 4, we need to establish the following lemma.

Lemma C.1. *Suppose that* $\sum X_i^2 \xrightarrow{p} 1$, $\sum \mathbb{E}[X_i^2] \to 1$, *and* $\sum_{i=1}^n \mathbb{E}[Y_i^2] \to 0$, *then*

$$\sum [X_i + Y_i]^2 \xrightarrow{p} 1 \quad \textit{and} \quad \sum \mathbb{E}\left[(X_i + Y_i)^2\right] \to 1. \qquad \text{(C.37)}$$

Proof. Note that

$$\sum (X_i + Y_i)^2 = \sum X_i^2 + \sum Y_i^2 + 2\sum X_i Y_i. \qquad \text{(C.38)}$$

Since $\sum \mathbb{E}[Y_i^2] \to 0$, we have $\sum Y_i^2 \xrightarrow{l_1} 0$, which implies that $\sum Y_i^2 \xrightarrow{p} 0$. By Cauchy–Schwarz inequality, we have

$$\begin{aligned}\mathbb{E}\left[|\sum X_i Y_i|\right] &\le \sum \sqrt{\mathbb{E}[X_i^2]}\sqrt{\mathbb{E}[Y_i^2]} \\ &\le \sqrt{\sum \mathbb{E}[X_i^2]}\sqrt{\sum \mathbb{E}[Y_i^2]} \\ &\to 0.\end{aligned} \tag{C.39}$$

Thus, $\sum X_i Y_i \xrightarrow{l_1} 0$, which implies that $\sum X_i Y_i \xrightarrow{p} 0$. Therefore, $\sum (X_i + Y_i)^2 \xrightarrow{p} 1$ by Slutsky's lemma. Furthermore, since $\mathbb{E}[\sum X_i Y_i] \to 0$, we have $\sum \mathbb{E}[X_i + Y_i]^2 \to 1$. □

Proof of Theorem 3: For simplicity, we first ignore the extra randomness ω. According to the H-decomposition of $s(x_1, \ldots, x_k)$, we have

$$\begin{aligned}&\frac{1}{n}\sum [e_i - s_0]^2 \\ &= \frac{1}{n}\frac{(n-k)^2}{n^2}\sum_{i=1}^{n}\left[\sum_{j=1}^{k} - \binom{k-1}{j-1}\binom{n-1}{j}^{-1}\sum s^j(X_{i_1}, \ldots, X_{i_j}; \nexists i)\right. \\ &\quad \left. + \binom{k-1}{j-1}\binom{n-1}{j-1}^{-1}\sum s^j(X_{i_1}, \ldots, X_{i_j}; \exists i)\right]^2 \\ &= \frac{1}{n}\frac{(n-k)^2}{n^2}\sum_{i=1}^{n}\left[-\frac{1}{n-1}\sum_{j\neq i}^{n} s^1(X_i) + s^1(X_i) + \sum_{j=2}^{k}\right. \\ &\left. - \binom{k-1}{j-1}\binom{n-1}{j}^{-1}\sum s^j(X_{i_1}, \ldots, X_{i_j}; \nexists i) + \binom{k-1}{j-1}\binom{n-1}{j-1}^{-1}\sum s^{(j)}(X_{i_1}, \ldots, X_{i_j}; \exists i)\right]^2 \\ &= \frac{1}{n}\frac{(n-k)^2}{n^2}\sum_{i=1}^{n}\left[s^1(X_i) + \mathrm{T}_i\right]^2.\end{aligned} \tag{C.40}$$

$s^1(X_i)$ and T_i are uncorrelated and have mean 0. After some calculation, we find that

$$\mathbb{E}[(s^1(X_i))^2] = V_1$$

and

$$\mathbb{E}[\mathrm{T}_i^2] = \frac{1}{n-1}V_1 + \sum_{j=2}^{k}\binom{k-1}{j-1}^2\left[\binom{n-1}{j}^{-1} + \binom{n-1}{j-1}^{-1}\right]V_j$$

$$= \frac{1}{n-1}V_1 + \frac{n}{k^2}\sum_{j=2}^{k}\frac{j}{1-j/n}\frac{\binom{k}{j}^2}{\binom{n}{j}}V_j.$$

Then

$$\begin{aligned}\mathbb{E}[\mathrm{T}_i^2] &= \left[\frac{1}{n-1}V_1 + \sum_{j=2}^{k}\binom{k-1}{j-1}^2\binom{n-1}{j-1}^{-1}V_j\right](1+o(1))\\ &= \left[\frac{1}{n-1}V_1 + \sum_{j=2}^{k}\frac{j}{k}\binom{k-1}{j-1}\binom{n-1}{j-1}^{-1}\left[\binom{k}{j}V_j\right]\right](1+o(1))\\ &\leq \left[\frac{1}{n-1}V_1 + \frac{2}{n}\sum_{j=2}^{k}\binom{k}{j}V_j\right](1+o(1))\\ &= \left[\frac{1}{n-1}\zeta_1 + \frac{2}{n}(\zeta_k - k\zeta_1)\right](1+o(1)),\end{aligned} \tag{C.41}$$

where $\zeta_k = \mathrm{Var}(s) = \sum_{j=1}^{k}\binom{k}{j}V_j$ and $\zeta_1 = \mathrm{Var}(\mathbb{E}[s|X_1]) = V_1$. Let $\mathrm{L} = \mathbb{E}[(s^1(X_i))^2]$ and $\mathrm{R} = \mathbb{E}[\mathrm{T}_i^2]$. Since $\frac{k}{n}(\frac{\zeta_k}{k\zeta_1} - 1) \to 0$, we have

$$\mathrm{R}/\mathrm{L} \leq \left[\frac{2/n(\zeta_k - k\zeta_1)}{\zeta_1} + \frac{1}{n-1}\right](1+o(1)) \to 0. \tag{C.42}$$

Therefore, $(s^1(X_i))^2$ dominates T_i^2 and thus by Lemma C.1,

$$\begin{aligned}\frac{1}{n}\sum[e_i - s_0]^2/V_1 &\xrightarrow{p} \frac{1}{n}\frac{(n-k)^2}{n^2}\sum_{i=1}^{n}[s^{(1)}(X_i)]^2/V_1\\ &\xrightarrow{p} \frac{(n-k)^2}{n^2}\mathbb{E}[s^{(1)}(X_i)]^2/V_1\\ &\to 1.\end{aligned} \tag{C.43}$$

So, $\text{ps-IJ}_{\text{U}}/\frac{k^2}{n}V_1 = \frac{1}{n}\sum[e_i - s_0]^2/V_1 \xrightarrow{p} 1$. Observe that

$$
\begin{aligned}
1 \le \text{Var}(\text{U}_{n,k})/\frac{k^2}{n}V_1 &= \left(\frac{k^2}{n}V_1\right)^{-1}\sum_{j=1}^{k}\binom{k}{j}^2\binom{n}{j}^{-1}V_j \\
&\le 1 + \left(\frac{k^2}{n}V_1\right)^{-1}\frac{k^2}{n^2}\sum_{j=2}^{k}\binom{k}{j}V_j \\
&\le 1 + \frac{k}{n}\left(\frac{\zeta_k}{k\zeta_1} - 1\right) \\
&\to 1.
\end{aligned}
\tag{C.44}
$$

Therefore, $\text{Var}(\text{U}_{n,k})/\frac{k^2}{n}V_1 \to 1$ and thus $\text{ps-IJ}_{\text{U}}/\text{Var}(\text{U}_{n,k}) \xrightarrow{p} 1$.

For $s = s(x_1, \ldots, x_k; \omega)$, we define an extended H-decomposition by letting

$$s^1(x_1) = s_1(x_1), \tag{C.45}$$

$$s^c(x_1, \ldots, x_c) = s_c(x_1, \ldots, x_c) - \sum_{j=1}^{c}\sum_{i_1,\ldots,i_j\in\{1,\ldots,j\}} s^j(x_{i_1}, \ldots, x_{i_j}) \tag{C.46}$$

for $c = 1, \ldots, k-1$ and

$$s^k(x_1, \ldots, x_k) = s(x_1, \ldots, x_k; \omega) - \sum_{j=1}^{k-1}\sum_{i_1,\ldots,i_j\in\{1,\ldots,j\}} s^j(x_{i_1}, \ldots, x_{i_j}) \tag{C.47}$$

where $s_c(x) = \mathbb{E}[s(x_1, \ldots, x_c, X_{c+1}, X_k; \omega)] - \mathbb{E}[s]$. Then $s(x_1, \ldots, x_k; \omega) = \mathbb{E}[s] + \sum_{j=1}^{k}\sum_{i_1,\ldots,i_j\in\{1,\ldots,j\}} s^j(x_{i_1}, \ldots, x_{i_j})$ and thus for

$$\text{ps-IJ}_{\text{U}}^{\omega} = \frac{k^2}{n^2}\sum[e_i^{\omega} - s_0^{\omega}]^2, \tag{C.48}$$

it can be decomposed the same way as Eq. (C.40). Thus, we have $\text{ps-IJ}_{\text{U}}^{\omega}/\frac{k^2}{n}\zeta_{1,\omega} \xrightarrow{p} 1$. ■

Proof of Theorem 4: As above, let us first ignore the extra randomness ω for simplicity. Letting $p = N\binom{n}{k}^{-1}$, we have

$$\begin{aligned}
\hat{e}_i - \hat{s}_0 &= \frac{n}{Nk}\sum s(X_{i_1},\ldots,X_{i_k};\exists i) - \frac{1}{N}\sum s(X_{i_1},\ldots,X_{i_k}) \\
&= \frac{n}{Nk}\sum(s(X_{i_1},\ldots,X_{i_k};\exists i)-\theta) - \frac{1}{N}\sum(s(X_{i_1},\ldots,X_{i_k})-\theta) + \left(\frac{\hat{N}_i}{N_i} - \frac{\hat{N}}{N}\right)\theta \\
&= \binom{n-1}{k-1}^{-1}\sum\frac{\rho}{p}(s(X_{i_1},\ldots,X_{i_k};\exists i)-\theta) - \binom{n}{k}^{-1}\sum\frac{\rho}{p}(s(X_{i_1},\ldots,X_{i_k})-\theta) + r_i \\
&\triangleq e_i^\dagger - s_0^\dagger + r_i.
\end{aligned} \tag{C.49}$$

where $N_i = Nk/n$, $\hat{N} = \sum\rho$ and $\hat{N}_i = \sum\rho\mathbf{1}_{i\in\{i_1,\ldots,i_k\}}$.

Comparing the H-decomposition of $s^\dagger(x_1,\ldots,x_k;\rho) = \frac{\rho}{p}s(x_1,\ldots,x_k)$ and $s(x_1,\ldots,x_k)$, we have $V_j^\dagger = V_j$ for $j = 1,\ldots,k-1$ and $V_k^\dagger = V_k + \frac{1-p}{p}\zeta_k$. Similar to Eq. (C.40), we have

$$\frac{1}{n}\sum_{i=1}^{n}[e_i^\dagger - s_0^\dagger]^2 = \frac{1}{n}\frac{(n-k)^2}{n^2}\sum_{i=1}^{n}\left[s^1(X_i) + \mathrm{T}_i^\dagger\right]^2, \tag{C.50}$$

where $s^1(x) = \mathbb{E}[s(x,X_2,\ldots,X_k)]$. Note that $\mathbb{E}[(s^1(X_1))^2] = V_1^\dagger = V_1$ and

$$\begin{aligned}
\mathbb{E}[(\mathrm{T}_i^\dagger)^2] &= \frac{1}{n-1}V_1^\dagger + \frac{n}{k^2}\sum_{j=2}^{k}\frac{j}{1-j/n}\frac{\binom{k}{j}^2}{\binom{n}{j}}V_j^\dagger \\
&= \frac{1}{n-1}V_1 + \frac{n}{k^2}\sum_{j=2}^{k}\frac{j}{1-j/n}\frac{\binom{k}{j}^2}{\binom{n}{j}}V_j + \frac{n}{k^2}\frac{k}{1-k/n}\frac{1}{N}(1-p)\zeta_k \\
&:= \mathrm{R} + \mathrm{M}
\end{aligned} \tag{C.51}$$

where $\mathrm{M} = \frac{1}{1-k/n}\frac{n}{Nk}(1-p)\zeta_k$. Let $\mathrm{L} = \mathbb{E}[(s^1(X_i))^2]$. Since $\frac{k}{n}(\frac{\zeta_k}{k\zeta_1} - 1) \to 0$, we have $\mathrm{R}/\mathrm{L} \to 0$

by Eq. (C.42). Next, we have

$$\begin{aligned}
\mathrm{M/L} &= \frac{\frac{1}{1-k/n}\frac{n}{Nk}(1-p)\zeta_k}{\zeta_1} \\
&\leq \frac{1}{1-k/n} \times \frac{n}{N}\frac{\zeta_k}{k\zeta_1} \qquad \text{(C.52)} \\
&= \left[\frac{n}{N}\frac{\zeta_k}{k\zeta_1}\right] \cdot O(1) \to 0
\end{aligned}$$

for ζ_k is bounded and $\frac{n}{Nk\zeta_1} \to 0$. Thus, $\frac{1}{n}\sum_{i=1}^{n}[e_i^\dagger - s_0^\dagger]^2/V_1 \xrightarrow{p} 0$ by Lemma C.1. Note that

$$\begin{aligned}
\mathbb{E}[r_i^2] &= \mathbb{E}\left[\left(\frac{\hat{N}_i}{N_i} - 1\right) - \left(\frac{\hat{N}}{N} - 1\right)\right]^2 \theta^2 \\
&\leq 2\theta^2\left[\frac{1}{N_i}\left(1 - \frac{N}{\binom{n}{k}}\right) + \frac{1}{N}\left(1 - \frac{N}{\binom{n}{k}}\right)\right] \qquad \text{(C.53)} \\
&\leq 4\theta^2/N_i.
\end{aligned}$$

Thus, $\frac{1}{n}\sum \mathbb{E}[\sum r_i^2]/V_1 \leq 4\theta^2 \frac{n}{N}\frac{1}{kV_1} \to 0$ according to the conditions. By Lemma C.1 again, we have $\frac{1}{n}\sum_i(\hat{e}_i - \hat{s}_0)^2/V_1 \xrightarrow{p} 1$ and it follows that $\widehat{\text{ps-IJ}_\text{U}}/\frac{k^2}{n}V_1 \xrightarrow{p} 1$.

Again, the extra randomness only results in an extended version of H-decomposition. Everything above can be directly applied to $s(x_1, \ldots, x_k; \omega)$. ■

D. Higher Order Pseudo Infinitesimal Jackknife

Recall that in the context of U-statistics, $\text{Var(U)} = \sum_{j=1}^{k}\binom{k}{j}^2\binom{n}{j}^{-1}V_j$. In the final discussion provided in the main text, we noted that the preceding results largely assumed that the U-statistic was close to linear statistic so that the variance of U-statistic is dominated by its first order term k^2/nV_1 and so the problem of providing a good estimate for Var(U) can be reduced to providing a good estimate for V_1. But what if the statistic is not close to linear and the remaining terms in Var(U) are not negligible? Can we obtain an improved estimator by proposing further estimates of V_j for $j = 2, \ldots, k$? We now address these questions.

We begin by considering the second term V_2 and extend those results to all j, $3 \leq j \leq k$. Since $V_2 = \mathrm{Var}(\mathbb{E}[s|X_1, X_2] - \mathbb{E}[s|X_1] - \mathbb{E}[s|X_2] + \mathbb{E}[s])$, a natural estimate for the second order term $\binom{k}{2}^2 \binom{n}{2}^{-1} V_2$ would be

$$\left(\binom{k}{2} / \binom{n}{2}\right)^2 \sum_{i,j} [e_{ij} - e_i - e_j + s_0]^2 \tag{D.54}$$

where $e_{ij} = \mathbb{E}_*[s^*|X_1^* = X_i, X_2^* = X_j]$. Before analyzing the properties of this estimate, we first point out its connection to the ps-IJ$_\mathrm{U}$.

Proposition D.1. *Let $\mathcal{D}_n^* = (X_1^*, \ldots, X_k^*)$ be a subsample of $\mathcal{D}_n$ and $w_{ij}^* = \mathbf{1}_{X_i, X_j \in \mathcal{D}_n^*} - \frac{k}{n} \mathbf{1}_{X_i \in \mathcal{D}_n^*} - \frac{k}{n} \mathbf{1}_{X_j \in \mathcal{D}_n^*} + \frac{k(k-1)}{n(n-1)}$. Then*

$$\mathrm{Cov}_*(s^*, w_{ij}^*) = \binom{k}{2} / \binom{n}{2} (e_{ij} - e_i - e_j + s_0) \tag{D.55}$$

where $*$ refers the procedure of subsampling without replacement and $e_{ij} = \mathbb{E}_*[s^*|X_1^* = X_i, X_2^* = X_j]$. We call Eq. (D.54) *the second order pseudo-IJ estimator of U-statistics:*

$$\begin{aligned} \text{ps-IJ}_\mathrm{U}(2) &= \sum_{i,j} \mathrm{Cov}_*^2(s^*, w_{ij}^*) \\ &= \binom{k}{2}^2 / \binom{n}{2}^2 \sum_{i,j} [e_{ij} - e_i - e_j + s_0]^2. \end{aligned} \tag{D.56}$$

Proof. We have

$$
\begin{aligned}
&(e_{12} - e_1 - e_2 + s_0) \\
&= \sum_{w_1^*=1, w_2^*=1, w_1^*+\cdots+w_n^*=k} \frac{(k-2)!}{(n-2)\cdots(n-k)} s^* - \sum_{w_1^*=1, w_1^*+\cdots+w_n^*=k} \frac{(k-1)!}{(n-1)\cdots(n-k)} s^* - \\
&\quad \sum_{w_2^*=1, w_1^*+\cdots+w_n^*=k} \frac{(k-1)!}{(n-2)\cdots(n-k)} s^* + \sum_{w_1^*+\cdots+w_n^*=k} \frac{k!}{n\cdots(n-k)} s^* \\
&= \sum_{w_1^*+\cdots+w_n^*=k} \frac{n(n-1)}{k(k-1)} \binom{n}{k}^{-1} (w_1^*)(w_2^*) s^* - \sum_{w_1^*+\cdots+w_n^*=k} \frac{n}{k} \binom{n}{k}^{-1} (w_1^*) s^* - \\
&\quad \sum_{w_1^*+\cdots+w_n^*=k} \frac{n}{k} \binom{n}{k}^{-1} (w_2^*) s^* + \sum_{w_1^*+\cdots+w_n^*=k} \binom{n}{k}^{-1} s^* \\
&= \sum \binom{n}{k}^{-1} \left(\frac{n(n-1)}{k(k-1)} w_1^* w_2^* - \frac{n}{k} w_1^* - \frac{n}{k} w_2^* + 1 \right) s^* \\
&= \frac{n(n-1)}{k(k-1)} \sum \binom{n}{k}^{-1} \left(w_1^* w_2^* - \frac{k-1}{n-1} w_1^* - \frac{k-1}{n-1} w_2^* + \frac{k(k-1)}{n(n-1)} \right) s^*.
\end{aligned}
\tag{D.57}
$$

Thus,

$$
\sum_{i,j} \mathrm{Cov}_*^2(s^*, w_{ij}^*) = \left(\frac{\binom{k}{2}}{\binom{n}{2}} \right)^2 \sum_{i,j} (e_{1,2} - e_1 - e_2 + s_0)^2. \tag{D.58}
$$

□

Note that the first-order ps-IJ$_\mathrm{U}$ involves the covariance of s^* and w_j^* – the counts of how many times each original observation appears in a subsample, whereas ps-IJ$_\mathrm{U}(2)$ involves covariance the of s^* and w_{ij}^* – the count of how often each *pair* of observations appears in a subsample. In this sense then, ps-IJ$_\mathrm{U}(2)$ is a natural extension of ps-IJ$_\mathrm{U}$ and for notational convenience, we can also write ps-IJ$_\mathrm{U}$ as ps-IJ$_\mathrm{U}(1)$. Similarly, we can extend this idea to derive a general d^{th} order estimator ps-IJ$_\mathrm{U}(d)$ for $d = 1, \ldots, k$.

Corollary D.1. *For $d = 1, \ldots, k$, define*

$$\begin{aligned} \text{ps-IJ}_{\text{U}}(d) &= \sum_{(n,d)} \text{Cov}_*^2(s^*, w^*_{i_1,\ldots,i_d}) \\ &= \binom{k}{d}^2 \Big/ \binom{n}{d}^2 \sum_{(n,d)} \left[\sum_{j=0}^{d} (-1)^{d-j} \sum_{(d,j)} e_{i_1,\ldots,i_j} \right]^2 \end{aligned} \tag{D.59}$$

where $w^*_{i_1,\ldots,i_d} = \sum_{j=0}^{d} (-1)^{d-j} \frac{\binom{n-d+j}{k-d+j}}{\binom{n}{k}} \left[\sum_{(d,j)} \prod w^*_{i_j} \right]$. The expression for $w^*_{i_1,\ldots,i_d}$ is somewhat involved because we are considering subsampling without replacement. If instead we perform subsampling with replacement, then $w^*_{i_1,\ldots,i_d} = \prod (w^*_{i_j} - 1)$.

Like $\mathbb{E}[\text{ps-IJ}_{\text{U}}]$, $\mathbb{E}[\text{ps-IJ}_{\text{U}}(d)]$ is a linear combination of the V_j. Let $a_i = \binom{n-i}{k-i}^{-1}$ for $i = 0, 1, \ldots, d$ and define b_i for $i = 0, 1, \ldots, d$ by

$$b_0 = a_0$$

$$b_1 = a_1 - a_0 = a_1 - b_0$$

$$\vdots$$

$$b_d = a_d - \binom{d}{1} a_{d-1} + \binom{d}{2} a_{d-2} - \ldots a_0 = a_d - \binom{d}{1} b_{d-1} - \binom{d}{2} b_{d-2} - \cdots - b_0.$$

Additionally, let $c_i = b_i \binom{n-d}{k-i}$ and $m_i = c_{d-i}$ for $i = 0, \ldots, d$. Then for $j = 1, \ldots, k$, the coefficient of V_j in $\mathbb{E}[\text{ps-IJ}_{\text{U}}(d)]$ is $\binom{k}{d}^2/\binom{n}{d}\lambda_j(d)$, where

$$\begin{aligned}
\lambda_j(d) = &\binom{d}{0}\binom{n-d}{j-d}^{-1}\left(m_0\binom{n-d}{j-d}\right)^2 + \\
&\binom{d}{1}\binom{n-d}{j-d+1)}^{-1}\left[m_1\binom{k-d+1}{j-d+1} - m_0\binom{k-d}{j-d+1}\right]^2 + \\
&\binom{d}{2}\binom{n-d}{j-d+2}^{-1}\left[m_2\binom{k-d+2}{j-d+2} - \binom{2}{1}m_1\binom{k-d+1}{j-d+2} + m_0\binom{k-d}{j-d+2)}\right]^2 + \\
&\vdots \\
&\binom{d}{d}\binom{n-d}{j}^{-1}\left[m_d\binom{k}{j} - \binom{d}{d-1}m_{d-1}\binom{k-1}{j}\right. \\
&\quad \left. + \binom{d}{d-2}m_{d-2}\binom{k-2}{j} - \ldots \binom{d}{1}m_0\binom{k-d}{j}\right]^2.
\end{aligned}$$

Putting this all together, we have the follow result.

Proposition D.2. *Writing the* $\text{Var}(\text{U})$ *and* $\mathbb{E}[\text{ps-IJ}_{\text{U}}(d)]$ *in terms of* $V_1, \ldots, V_k$*, the ratio of the coefficient of* V_j *in* $\mathbb{E}[\text{ps-IJ}_{\text{U}}(d)]$ *to that in* $\text{Var}(\text{U})$ *is given by* $r_j(d)$*, where*

$$r_j(d) = \frac{\lambda_j(d)\binom{k}{d}^2\binom{n}{d}^{-1}}{\binom{k}{j}^2\binom{n}{j}^{-1}}, \quad j = 1, \ldots, k. \tag{D.60}$$

Furthermore, $r_j(d)$ *is monotone increasing with respect to* j*.*

Proof. We derive $\mathbb{E}[\text{ps-IJ}_{\text{U}}(2)]$ in work that follows. Expressions for $d \geq 3$ can be derived in the same spirit. Consider $e_{ij} = \mathbb{E}_*[s^*|X_1^* = X_1, X_2^* = X_2]$. We have

$$
\begin{aligned}
(e_{12} - e_1 - e_2 + s_0) = &\binom{n}{k}^{-1} \sum s(X_{i_1}, \ldots, X_{i_k}; \nexists 1, \nexists 2) + \\
&\left(\binom{n-2}{k-2}^{-1} - 2\binom{n-1}{k-1}^{-1} + \binom{n}{k}^{-1}\right) \sum s(X_{i_1}, \ldots, X_{i_k}; \exists 1, \exists 2) - \\
&\left(\binom{n-1}{k-1} - \binom{n}{k}^{-1}\right) \sum s(X_{i_1}, \ldots, X_{i_k}; \exists 1, \nexists 2) - \\
&\left(\binom{n-1}{k-1} - \binom{n}{k}^{-1}\right) \sum s(X_{i_1}, \ldots, X_{i_k}; \nexists 1 \exists 2) \\
:= &\mathrm{I} + \mathrm{II} + \mathrm{III} + \mathrm{IV}.
\end{aligned}
$$

Looking at each term individually, by H-decomposition we have

$$
\begin{aligned}
\mathrm{I} &= \binom{n}{k}^{-1} \binom{n-2}{k} \binom{n-2}{k}^{-1} \sum s(X_{i_1}, \ldots, X_{i_k}; \nexists 1, \nexists 2) \\
&= \binom{n}{k}^{-1} \binom{n-2}{k} \sum_{j=1}^{k} \binom{k}{j} \binom{n-2}{j}^{-1} \sum s^j(X_{i_1}, \ldots, X_{i_j}; \nexists 1, \nexists 2)
\end{aligned}
\tag{D.61}
$$

$$
\begin{aligned}
\mathrm{II} &= -\left(\binom{n-1}{k-1}^{-1} - \binom{n}{k}^{-1}\right) \sum s(X_{i_1}, \ldots, X_{i_k}; \exists 1, \nexists 2) \\
&= -\left(\binom{n-1}{k-1}^{-1} - \binom{n}{k}^{-1}\right) \binom{n-2}{k-1} \binom{n-2}{k-1}^{-1} \sum s(X_{i_1}, \ldots, X_{i_k}; \exists 1, \nexists 2) \\
&= -\left(\binom{n-1}{k-1}^{-1} - \binom{n}{k}^{-1}\right) \binom{n-2}{k-1} \left[\sum_{j=1}^{k-1} \binom{k-1}{j} \binom{n-2}{j}^{-1} \sum s^j(X_{i_1}, \ldots, X_{i_j}; \nexists 1, \nexists 2) + \right. \\
&\left. \sum_{j=1}^{k} \binom{k-1}{j-1} \binom{n-2}{j-1}^{-1} \sum s^j(X_{i_1}, \ldots, X_{i_j}; \exists 1, \nexists 2)\right]
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{III} &= -\left(\binom{n-1}{k-1}^{-1} - \binom{n}{k}^{-1}\right)\sum s(X_{i_1},\dots,X_{i_k};\nexists 1,\exists 2)\\
&= -\left(\binom{n-1}{k-1}^{-1} - \binom{n}{k}^{-1}\right)\binom{n-2}{k-1}\binom{n-2}{k-1}^{-1}\sum s(X_{i_1},\dots,X_{i_k};\nexists 1,\exists 2)\\
&= -\left(\binom{n-1}{k-1}^{-1} - \binom{n}{k}^{-1}\right)\binom{n-2}{k-1}\left[\sum_{j=1}^{k-1}\binom{k-1}{j}\binom{n-2}{j}^{-1}\sum s^j(X_{i_1},\dots,X_{i_j};\nexists 1,\nexists 2) + \right.\\
&\quad \left.\sum_{j=1}^{k}\binom{k-1}{j-1}\binom{n-2}{j-1}^{-1}\sum s^j(X_{i_1},\dots,X_{i_j};\nexists 1,\exists 2)\right]
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{IV} &= \left(\binom{n-2}{k-2}^{-1} - 2\binom{n-1}{k-1}^{-1} + \binom{n}{k}^{-1}\right)\sum s(X_{i_1},\dots,X_{i_k};\exists 1,\exists 2)\\
&= \left(\binom{n-2}{k-2}^{-1} - 2\binom{n-1}{k-1}^{-1} + \binom{n}{k}^{-1}\right)\binom{n-2}{k-2}\left[\sum_{j=1}^{k-2}\binom{k-2}{j}\binom{n-2}{j}^{-1}\times\right.\\
&\qquad \sum s^j(X_{i_1},\dots,X_{i_j};\nexists 1,\nexists 2)+\\
&\qquad \sum_{j=1}^{k-1}\binom{k-2}{j-1}\binom{n-2}{j-1}^{-1}\times\sum s^j(X_{i_1},\dots,X_{i_j};\exists 1,\nexists 2) +\\
&\qquad \sum_{j=1}^{k-1}\binom{k-2}{j-1}\binom{n-2}{j-1}^{-1}\times\sum s^j(X_{i_1},\dots,X_{i_j};\nexists 1,\exists 2) +\\
&\qquad \left.\sum_{j=2}^{k}\binom{k-2}{j-2}\binom{n-2}{j-2}^{-1}\times\sum s^j(X_{i_1},\dots,X_{i_j};\exists 1,\exists 2)\right]
\end{aligned}
\tag{D.62}
$$

In conclusion, we have $\mathrm{I}+\mathrm{II}+\mathrm{III}+\mathrm{IV} = A+B+C+D$, where A,B,C,D are uncorrelated and given by

$$
\begin{aligned}
A = &\left(\binom{n-2}{k-2}^{-1} - 2\binom{n-1}{k-1}^{-1} + \binom{n}{k}^{-1}\right)\binom{n-2}{k-2}\times\\
&\sum_{j=2}^{k}\binom{k-2}{j-2}\binom{n-2}{j-2}^{-1}\sum s^j(X_{i_1},\dots,X_{i_j};\exists 1,\exists 2)
\end{aligned}
$$

$$B = -\left(\binom{n-1}{k-1} - \binom{n}{k}^{-1}\right)\binom{n-2}{k-1}\left[\sum_{j=1}^{k}\binom{k-1}{j-1}\binom{n-2}{j-1}^{-1}\sum s^j(X_{i_1},\ldots,X_{i_j};\exists 1,\nexists 2)\right]+$$
$$\left(\binom{n-2}{k-2}^{-1} - 2\binom{n-1}{k-1}^{-1} + \binom{n}{k}^{-1}\right)\binom{n-2}{k-2}\times$$
$$\sum_{j=1}^{k-1}\binom{k-2}{j-1}\binom{n-2}{j-1}^{-1}\sum s^j(X_{i_1},\ldots,X_{i_j};\exists 1,\nexists 2)$$

$$C = -\left(\binom{n-1}{k-1} - \binom{n}{k}^{-1}\right)\binom{n-2}{k-1}\left[\sum_{j=1}^{k}\binom{k-1}{j-1}\binom{n-2}{j-1}^{-1}\sum s^j(X_{i_1},\ldots,X_{i_j};\nexists 1,\exists 2)\right]+$$
$$\left(\binom{n-2}{k-2}^{-1} - 2\binom{n-1}{k-1}^{-1} + \binom{n}{k}^{-1}\right)\binom{n-2}{k-2}\times$$
$$\sum_{j=1}^{k-1}\binom{k-2}{j-1}\binom{n-2}{j-1}^{-1}\sum s^j(X_{i_1},\ldots,X_{i_j};\nexists 1,\exists 2)$$

$$D = \binom{n}{k}^{-1}\binom{n-2}{k}\sum_{j=1}^{k}\binom{k}{j}\binom{n-2}{j}^{-1}\sum s^j(X_{i_1},\ldots,X_{i_j};\nexists 1,\nexists 2)-$$
$$\left(\binom{n-1}{k-1} - \binom{n}{k}^{-1}\right)\binom{n-2}{k-1}\left[\sum_{j=1}^{k-1}\binom{k-1}{j}\binom{n-2}{j}^{-1}\sum s(X_{i_1},\ldots,X_{i_j};\nexists 1,\nexists 2)\right]-$$
$$\left(\binom{n-1}{k-1} - \binom{n}{k}^{-1}\right)\binom{n-2}{k-1}\left[\sum_{j=1}^{k-1}\binom{k-1}{j}\binom{n-2}{j}^{-1}\sum s^j(X_{i_1},\ldots,X_{i_j};\nexists 1,\nexists 2)\right]+$$
$$\left(\binom{n-2}{k-2}^{-1} - 2\binom{n-1}{k-1}^{-1} + \binom{n}{k}^{-1}\right)\binom{n-2}{k-2}\times$$
$$\sum_{j=1}^{k-2}\binom{k-2}{j}\binom{n-2}{j}^{-1}\sum s^j(X_{i_1},\ldots,X_{i_j};\nexists 1,\nexists 2).$$

Let $C_2 = \left(\binom{n-2}{k-2}^{-1} - 2\binom{n-1}{k-1}^{-1} + \binom{n}{k}^{-1}\right)\binom{n-2}{k-2}$, $C_1 = \left(\binom{n-1}{k-1}^{-1} - \binom{n}{k}^{-1}\right)\binom{n-2}{k-1}$ and $C_0 = \binom{n}{k}^{-1}\binom{n-2}{k}$. Then we have

$$\begin{aligned}\mathrm{Var}(A) &= \sum_{j=2}^{k}\binom{n-2}{j-2}^{-1}\left(C_2\binom{k-2}{j-2}\right)^2 V_j\\ &= \sum_{j=2}^{k}\binom{n-2}{j-2}^{-1}\left(C_2\binom{k-2}{j-2}\right)^2 V_j\end{aligned} \tag{D.63}$$

$$\begin{aligned}\mathrm{Var}(B) = &\sum_{j=1}^{k-1} \binom{n-2}{j-1} \left(-C_1 \binom{k-1}{j-1} \binom{n-2}{j-1}^{-1} + C_2 \binom{k-2}{j-1} \binom{n-2}{j-1}^{-1}\right)^2 V_j + \\ &\binom{n-2}{k-1}^{-1} \left(-C_1 \binom{k-1}{k-1}\right)^2 V_k \\ = &\sum_{j=1}^{k} \binom{n-2}{j-1}^{-1} \left(-C_1 \binom{k-1}{j-1} + C_2 \binom{k-2}{j-1}\right)^2 V_j\end{aligned} \tag{D.64}$$

$$\mathrm{Var}(B) = \mathrm{Var}(C) \tag{D.65}$$

$$\begin{aligned}\mathrm{Var}(D) = &\sum_{j=1}^{k-2} \binom{n-2}{j}^{-1} \left(C_0 \binom{k}{j} - 2C_1 \binom{k-1}{j} + C_2 \binom{k-2}{j}\right) + \\ &\binom{n-2}{k-1}^{-1} \left(C_0 \binom{k}{k-1} - 2C_1 \binom{k-1}{k-1}\right) + \\ &\binom{n-2}{k} C_0 \binom{k}{k} V_k \\ = &\sum_{j=1}^{k} \binom{n-2}{j}^{-1} \left(C_0 \binom{k}{j} - 2C_1 \binom{k-1}{j} + C_2 \binom{k-2}{j}\right) V_j.\end{aligned} \tag{D.66}$$

Therefore $\mathbb{E}[\text{ps-IJ}_\text{U}(2)] = \binom{k}{2}^2 / \binom{n}{2} \sum_{j=1}^{k} \lambda_j(2) V_j$, where

$$\begin{aligned}\lambda_j(2) = &\binom{n-2}{j-2}^{-1} \left(C_2 \binom{k-2}{j-2}\right)^2 + \\ &2\binom{n-2}{j-1}^{-1} \left(-C_1 \binom{k-1}{j-1} + C_2 \binom{k-2}{j-1}\right)^2 + \\ &\binom{n-2}{j}^{-1} \left(C_0 \binom{k}{j} - 2C_1 \binom{k-1}{j} + C_2 \binom{k-2}{j}\right)^2.\end{aligned} \tag{D.67}$$

□

As a simple example, we can take $n = 20$ and $k = 10$ and plot the curve of $r_j(d)$ to get some insight into how it behaves. In the interest of consistency, we want for r_j/r_1 to be close

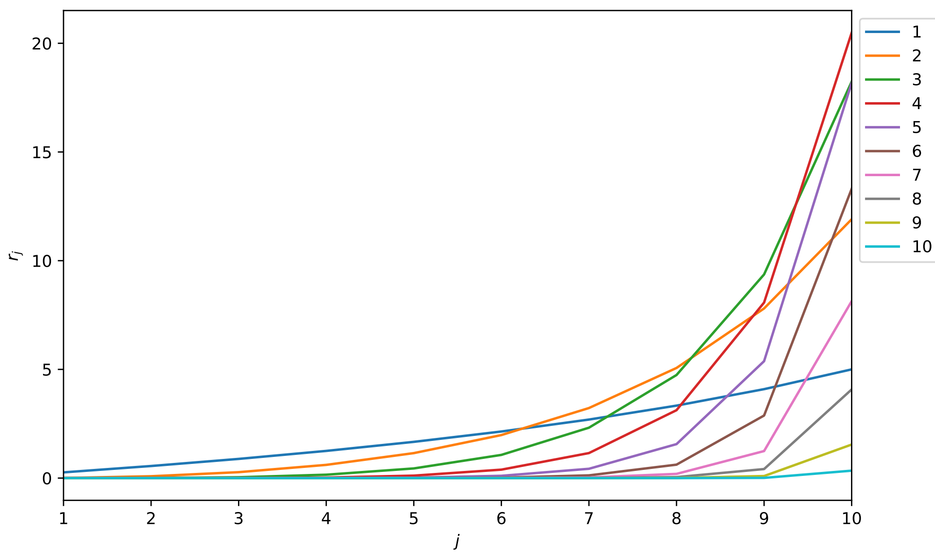

Figure 7: A plot of $\{r_j(d)\}_{j=1}^{k}$, where n=20 and k=10

to 1, at least for small j, because Var(U) should be dominated by the first several terms. From Figure 7 , it appears that ps-$\text{IJ}_{\text{U}}(1)$ still perform better than the other higher-order estimates. It is possible that combining ps-$\text{IJ}_{\text{U}}(d)$ for $d = 1, \ldots, k$ in some way could yield an estimator that outperforms ps-$\text{IJ}_{\text{U}}(1)$; this is a potentially interesting topic for future research.

E. Simulations Comparing the Variance Estimates

We now present a very brief initial simulation study to compare the estimates obtained from the three different methods laid out above as well as the Jackknife-After-Bootstrap (JAB). We consider estimating the variance of a random forest modeling the regression function

$$f(x) = \sin(\pi x_1 x_2) + 0.2(x_3 - 0.5)^2 + 0.5x_4 + 0.1x_5.$$

The covariates are sampled from a multivariate normal distribution with $\mu = (0.2, 0.3, 0.2, 0.7, 0.4)$, $\Sigma_{(i,i)} = 1$ and $\Sigma_{(i,j)} = 0.2$ for $1 \leq i, j \leq 5$. The responses are assigned as $y_i = f(x_i) + \epsilon_i$, for $i = 1, \ldots, 50$, with ϵ_i sampled from $\mathcal{N}(0, 0.1)$. The random forest consists of B fully grown decision trees, built on bootstrap samples of size $n = 50$ with the *mtry* parameter set to 5 so that all variables are available as split candidates at each node. Under this setup, we let s be the prediction of a decision tree at x_0, i.e., $\text{Tree}(X_1, \ldots, X_n; x_0)$, where $x_0 = (0.2, 0.3, 0.2, 0.7, 0.4)$ and the goal is to estimate the variance of the forest's prediction at x_0.

Like e_i, $t(\mathcal{D}_n[i])$ must be approximated via Monte Carlo. We approximate $t(\mathcal{D}_n[i])$ by $t\widehat{(\mathcal{D}_n[i]})$, the average of all s_b^* where X_i is not included in the collection of $X_{b1}^*, ..., X_{bn}^*$, which can be easily obtained by rearranging the original bootstrap replications with no further computation required for resampling. We repeat the process 400 times to obtain the boxplots, where the dashed line denotes the sample variance of the 400 forests. It is clear from Figure 8 that the three methods proposed above appear to perform quite similarly, while the JAB appears a bit more likely to overestimate.

F. Additional Simulations

We now provide a brief simulation using the same data and regression setup as laid out in Appendix E but with the random forests constructed a bit differently. Here we consider a

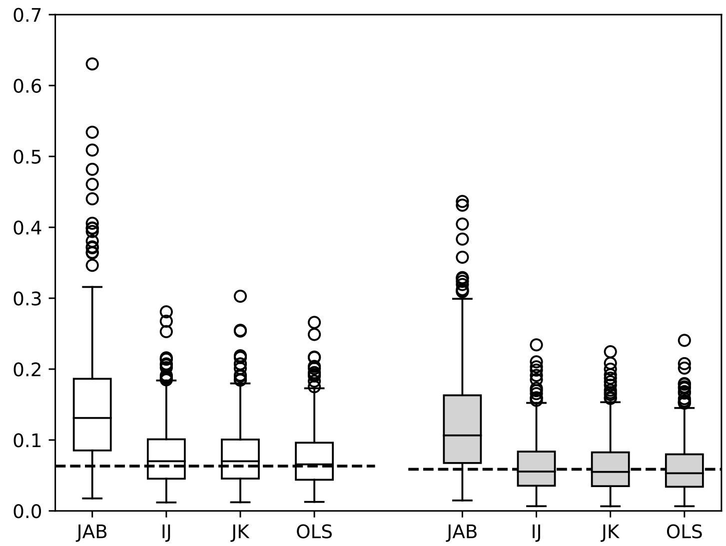

Figure 8: Performance of JAB, $\hat{\sigma}^2_{\text{IJ}}$,$\hat{\sigma}^2_{\text{JK}}$, and $\hat{\sigma}^2_{\text{OLS}}$ in estimating the variance of a random forest that consists of B decision trees with $B = 500$ (White) or $B = 1000$ (Grey). The dotted line indicates the sample variance of the forest.

sample size of $n = 50$, subsample size $k = 20$, and ensemble sizes N of 500, 1000 or 2000. (Note that N here denotes the ensemble size, or number of trees, which was denoted as B in earlier simulations.) We first generate a random variable $\hat{N} \sim \text{Binomial}(\binom{n}{k}, N/\binom{n}{k})$, then build $\hat{N}$ fully-grown decision trees with mtry $= 2$. As before, let s denote the prediction of a decision tree at $x_0 = (0.2, 0.3, 0.2, 0.7, 0.4)$; our goal is to estimate the variance of the random forest's prediction at x_0. The entire process is repeated 400 times to obtain the boxplots and sample

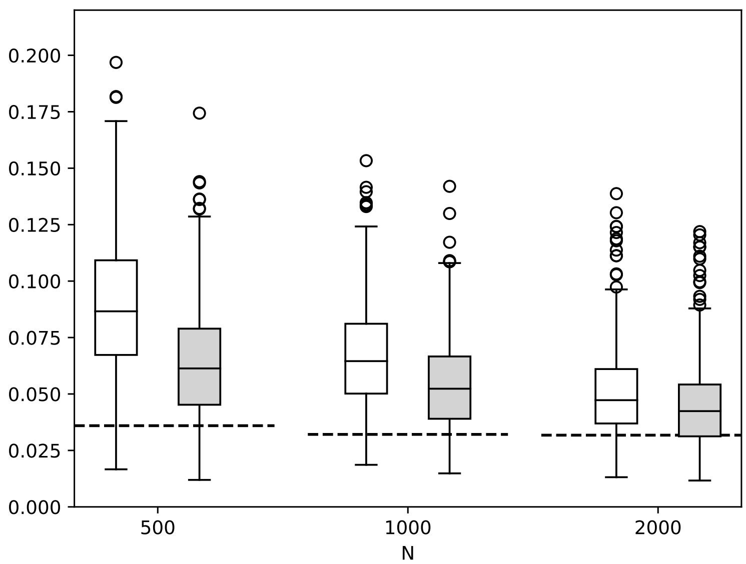

Figure 9: Performance of $\widehat{\text{ps-IJ}_{\text{U}}^{\omega}}$ (White) and $\sum_i \widehat{\text{Cov}}^2(s^*, w_i^*)$ (Grey) as a function of N, in estimating the variance of a random forest that consists of $\hat{N}$ decision trees with $\hat{N} \sim \text{Binomial}(\binom{n}{k}, N/\binom{n}{k})$, where the dotted line indicates the sample variance of the random forest.

variance of the random forest. Since $\frac{n}{n-k} = 1.67$ in our setting, which is not negligible, we apply the adjustment $\frac{n^2}{(n-k)^2}$ and compare $\frac{n^2}{(n-k)^2}\widehat{\text{ps-IJ}_{\text{U}}^{\omega}}$ and $\frac{n^2}{(n-k)^2}\sum_i \widehat{\text{Cov}}(s^*, \omega^*)$. As can be seen from the plots, the two estimates are fairly close to each other and both tend to overestimate the variance as implied by Theorem 2, due to the fact that k is not small enough relative to n so that the effect of the overestimation rates $r_2, \ldots, r_k$ can not be neglected. The estimate $\frac{n^2}{(n-k)^2}\sum_i \widehat{\text{Cov}}(s^*, \omega^*)$ also appears to be slightly more stable.