Online Supplement to Efficient VaR and CVaR Measurement via Stochastic Kriging

Xi Chen[1] and Kyoung-Kuk Kim[2]

Appendix A: Approximation of the Mean Squared Error

Derivation of (13)*:* In the SK predictor, we implicitly assume that $\mathbf{\Sigma}_{\mathsf{M}}+\mathbf{\Sigma}_{\varepsilon}$ and $\mathbf{\Sigma}_{\mathsf{M}}+\widehat{\mathbf{\Sigma}}_{\varepsilon}$ are invertible. Then by applying the results given by (Henderson and Searle 1981) it is straightforward to check that

$$\begin{aligned}\left(\mathbf{\Sigma}_{\mathsf{M}}+\widehat{\mathbf{\Sigma}}_{\varepsilon}\right)^{-1} &= (\mathbf{\Sigma}+\mathbf{\Delta}_{\mathrm{var}})^{-1}\\ &= \mathbf{\Sigma}^{-1}-\mathbf{\Sigma}^{-1}\mathbf{\Delta}_{\mathrm{var}}\mathbf{\Sigma}^{-1}+\mathbf{\Sigma}^{-1}T^2(\mathbf{I}-T)^{-1},\end{aligned}$$

which in turn yields

$$\left(\mathbf{\Sigma}_{\mathsf{M}}+\widehat{\mathbf{\Sigma}}_{\varepsilon}\right)^{-1}\widetilde{\varepsilon} = \mathbf{\Sigma}^{-1}\boldsymbol{\mu}_{\varepsilon}+\mathbf{\Sigma}^{-1}\left(\mathbf{\Delta}_{\mathrm{mean}}-\mathbf{\Delta}_{\mathrm{var}}\mathbf{\Sigma}^{-1}\widetilde{\varepsilon}\right)+\mathbf{\Sigma}^{-1}T^2(\mathbf{I}-T)^{-1}\widetilde{\varepsilon}.$$

Here we set $T=-\mathbf{\Delta}_{\mathrm{var}}\mathbf{\Sigma}^{-1}$. Next, it can be obtained from the expansion above that

$$\begin{aligned}&\left\{\mathsf{E}\left[\mathbf{\Sigma}_{\mathsf{M}}(\mathbf{x}_0,\cdot)^{\top}\left(\mathbf{\Sigma}_{\mathsf{M}}+\widehat{\mathbf{\Sigma}}_{\varepsilon}\right)^{-1}\widetilde{\varepsilon}\right]\right\}^2\\ &= \left\{\mathsf{E}\left[\mathbf{\Sigma}_{\mathsf{M}}(\mathbf{x}_0,\cdot)^{\top}\mathbf{\Sigma}^{-1}\left(\boldsymbol{\mu}_{\varepsilon}-\mathbf{\Delta}_{\mathrm{var}}\mathbf{\Sigma}^{-1}\widetilde{\varepsilon}+T^2(\mathbf{I}-T)^{-1}\widetilde{\varepsilon}\right)\right]\right\}^2\\ &= \left\{\mathsf{E}\left[\mathbf{\Sigma}_{\mathsf{M}}(\mathbf{x}_0,\cdot)^{\top}\mathbf{\Sigma}^{-1}\left(\boldsymbol{\mu}_{\varepsilon}-\mathbf{\Delta}_{\mathrm{var}}\mathbf{\Sigma}^{-1}\widetilde{\varepsilon}\right)\right]\right\}^2+\mathcal{O}_1\left(|\mathbf{\Delta}_{\mathrm{var}}|^2|\widetilde{\varepsilon}|\right)\mathcal{O}_1\left(|\mathbf{\Delta}_{\mathrm{var}}||\widetilde{\varepsilon}|+|\boldsymbol{\mu}_{\varepsilon}|\right)\\ &= \left(\mathbf{\Sigma}_{\mathsf{M}}(\mathbf{x}_0,\cdot)^{\top}\mathbf{\Sigma}^{-1}\boldsymbol{\mu}_{\varepsilon}\right)^2-2\mathbf{\Sigma}_{\mathsf{M}}(\mathbf{x}_0,\cdot)^{\top}\mathbf{\Sigma}^{-1}\mathsf{E}\left[\boldsymbol{\mu}_{\varepsilon}\widetilde{\varepsilon}^{\top}\mathbf{\Sigma}^{-1}\mathbf{\Delta}_{\mathrm{var}}\right]\mathbf{\Sigma}^{-1}\mathbf{\Sigma}_{\mathsf{M}}(\mathbf{x}_0,\cdot)+\mathcal{O}_1\left(|\mathbf{\Delta}_{\mathrm{var}}|^2|\widetilde{\varepsilon}|^2\right)\\ &\quad+\mathcal{O}_1\left(|\mathbf{\Delta}_{\mathrm{var}}|^2|\widetilde{\varepsilon}|\right)\mathcal{O}_1\left(|\mathbf{\Delta}_{\mathrm{var}}||\widetilde{\varepsilon}|+|\boldsymbol{\mu}_{\varepsilon}|\right),\end{aligned}$$

and

$$\begin{aligned}&\mathsf{Var}\left(\mathbf{\Sigma}_{\mathsf{M}}(\mathbf{x}_0,\cdot)^{\top}\left(\mathbf{\Sigma}_{\mathsf{M}}+\widehat{\mathbf{\Sigma}}_{\varepsilon}\right)^{-1}\widetilde{\varepsilon}\right)\\ &= \mathsf{Var}\left(\mathbf{\Sigma}_{\mathsf{M}}(\mathbf{x}_0,\cdot)^{\top}\mathbf{\Sigma}^{-1}\left(\mathbf{\Delta}_{\mathrm{mean}}-\mathbf{\Delta}_{\mathrm{var}}\mathbf{\Sigma}^{-1}\widetilde{\varepsilon}+T^2(\mathbf{I}-T)^{-1}\widetilde{\varepsilon}\right)\right)\\ &= \mathsf{Var}\left(\mathbf{\Sigma}_{\mathsf{M}}(\mathbf{x}_0,\cdot)^{\top}\mathbf{\Sigma}^{-1}\left(\mathbf{\Delta}_{\mathrm{mean}}-\mathbf{\Delta}_{\mathrm{var}}\mathbf{\Sigma}^{-1}\widetilde{\varepsilon}\right)\right)+\mathcal{O}_1\left(|\mathbf{\Delta}_{\mathrm{var}}|^2|\widetilde{\varepsilon}|\right)\mathcal{O}_1\left(|\mathbf{\Delta}_{\mathrm{var}}||\widetilde{\varepsilon}|\right)\\ &\quad+\mathcal{O}_1\left(|\mathbf{\Delta}_{\mathrm{var}}|^2|\widetilde{\varepsilon}||\mathbf{\Delta}_{\mathrm{mean}}|\right)+\mathcal{O}_1\left(|\mathbf{\Delta}_{\mathrm{var}}|^3|\widetilde{\varepsilon}|^2\right)\\ &= \mathbf{\Sigma}_{\mathsf{M}}(\mathbf{x}_0,\cdot)^{\top}\mathbf{\Sigma}^{-1}\mathbf{\Sigma}_{\varepsilon}\mathbf{\Sigma}^{-1}\mathbf{\Sigma}_{\mathsf{M}}(\mathbf{x}_0,\cdot)-2\mathbf{\Sigma}_{\mathsf{M}}(\mathbf{x}_0,\cdot)^{\top}\mathbf{\Sigma}^{-1}\mathsf{E}\left[\mathbf{\Delta}_{\mathrm{mean}}\widetilde{\varepsilon}^{\top}\mathbf{\Sigma}^{-1}\mathbf{\Delta}_{\mathrm{var}}\right]\mathbf{\Sigma}^{-1}\mathbf{\Sigma}_{\mathsf{M}}(\mathbf{x}_0,\cdot)\\ &\quad+\mathcal{O}_1\left(|\mathbf{\Delta}_{\mathrm{var}}|^2|\widetilde{\varepsilon}|^2\right)+\mathcal{O}_1\left(|\mathbf{\Delta}_{\mathrm{var}}|^2|\widetilde{\varepsilon}|\right)\mathcal{O}_1\left(|\mathbf{\Delta}_{\mathrm{var}}||\widetilde{\varepsilon}|\right)+\mathcal{O}_1\left(|\mathbf{\Delta}_{\mathrm{var}}|^2|\widetilde{\varepsilon}||\mathbf{\Delta}_{\mathrm{mean}}|\right).\end{aligned}$$

Note that $\mathcal{O}_1(\cdot)$ stands for L^1 boundedness in the underlying probability space.

On the other hand, we note that

$$\widehat{\mathbf{\Xi}} = \mathbf{\Xi}+\mathbf{\Sigma}^{-1}\mathbf{\Delta}_{\mathrm{var}}\mathbf{\Sigma}^{-1}-\mathbf{\Sigma}^{-1}T^2(\mathbf{I}-T)^{-1}.$$

[1] Grado Department of Industrial and Systems Engineering, Virginia Tech, USA, E-mail: xchen6@vt.edu
[2] Industrial and Systems Engineering, KAIST, Republic of Korea, E-mail: kkim3128@kaist.ac.kr

Since

$$\boldsymbol{\Xi} = \boldsymbol{\Sigma}_{\mathsf{M}}^{-1} - \boldsymbol{\Sigma}^{-1} = \left(\boldsymbol{\Sigma}_{\mathsf{M}}^{-1}\boldsymbol{\Sigma} - \mathbf{I}\right)\boldsymbol{\Sigma}^{-1} = \boldsymbol{\Sigma}_{\mathsf{M}}^{-1}\boldsymbol{\Sigma}_{\varepsilon}\boldsymbol{\Sigma}^{-1} = \boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{\varepsilon}\boldsymbol{\Sigma}_{\mathsf{M}}^{-1},$$

straightforward calculations yield

$$\begin{aligned}\widehat{\boldsymbol{\Xi}}\boldsymbol{\Sigma}_{\mathsf{M}}\widehat{\boldsymbol{\Xi}} &= \boldsymbol{\Xi}\boldsymbol{\Sigma}_{\mathsf{M}}\boldsymbol{\Xi} + \boldsymbol{\Sigma}^{-1}\boldsymbol{\Delta}_{\mathrm{var}}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{\varepsilon}\boldsymbol{\Sigma}^{-1} + \boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{\varepsilon}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Delta}_{\mathrm{var}}\boldsymbol{\Sigma}^{-1} \\ &\quad + \boldsymbol{\Sigma}^{-1}\boldsymbol{\Delta}_{\mathrm{var}}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{\mathsf{M}}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Delta}_{\mathrm{var}}\boldsymbol{\Sigma}^{-1} + \mathcal{O}\left(|\boldsymbol{\Delta}_{\mathrm{var}}|^2|\boldsymbol{\Sigma}_{\varepsilon}|\right) + \mathcal{O}\left(|\boldsymbol{\Delta}_{\mathrm{var}}|^3\right).\end{aligned}$$

Lastly, the last two terms in Equation (12) can be approximated as follows: ignoring higher-order terms,

$$\begin{aligned}&\boldsymbol{\Sigma}_{\mathsf{M}}(\mathbf{x}_0,\cdot)^{\top}\mathsf{E}\left[\widehat{\boldsymbol{\Xi}}\boldsymbol{\Sigma}_{\mathsf{M}}\widehat{\boldsymbol{\Xi}}\right]\boldsymbol{\Sigma}_{\mathsf{M}}(\mathbf{x}_0,\cdot) + \mathsf{E}\left(\left[\boldsymbol{\Sigma}_{\mathsf{M}}(\mathbf{x}_0,\cdot)^{\top}\left(\boldsymbol{\Sigma}_{\mathsf{M}} + \widehat{\boldsymbol{\Sigma}}_{\varepsilon}\right)^{-1}\widetilde{\boldsymbol{\varepsilon}}\right]^2\right) \\ \approx\ &\boldsymbol{\Sigma}_{\mathsf{M}}(\mathbf{x}_0,\cdot)^{\top}\left[\boldsymbol{\Xi}\boldsymbol{\Sigma}_{\mathsf{M}}\boldsymbol{\Xi}\right]\boldsymbol{\Sigma}_{\mathsf{M}}(\mathbf{x}_0,\cdot) + 2\boldsymbol{\Sigma}_{\mathsf{M}}(\mathbf{x}_0,\cdot)^{\top}\boldsymbol{\Sigma}^{-1}\mathsf{E}\left[\boldsymbol{\Delta}_{\mathrm{var}}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{\varepsilon}\right]\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{\mathsf{M}}(\mathbf{x}_0,\cdot) \\ &+\mathsf{E}\left[\left(\boldsymbol{\Sigma}_{\mathsf{M}}(\mathbf{x}_0,\cdot)^{\top}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Delta}_{\mathrm{var}}\boldsymbol{\Sigma}^{-1}\sqrt{\boldsymbol{\Sigma}_{\mathsf{M}}}\right)^2\right] + \left(\boldsymbol{\Sigma}_{\mathsf{M}}(\mathbf{x}_0,\cdot)^{\top}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_{\varepsilon}\right)^2 \\ &+\boldsymbol{\Sigma}_{\mathsf{M}}(\mathbf{x}_0,\cdot)^{\top}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{\varepsilon}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{\mathsf{M}}(\mathbf{x}_0,\cdot) - 2\boldsymbol{\Sigma}_{\mathsf{M}}(\mathbf{x}_0,\cdot)^{\top}\boldsymbol{\Sigma}^{-1}\mathsf{E}\left[\widetilde{\boldsymbol{\varepsilon}}\widetilde{\boldsymbol{\varepsilon}}^{\top}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Delta}_{\mathrm{var}}\right]\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{\mathsf{M}}(\mathbf{x}_0,\cdot).\end{aligned}$$

The desired expression is obtained by rearranging terms and by utilizing the identity

$$\boldsymbol{\Xi}\boldsymbol{\Sigma}_{\mathsf{M}}\boldsymbol{\Xi} + \boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{\varepsilon}\boldsymbol{\Sigma}^{-1} = \boldsymbol{\Xi}.$$

□

Appendix B: Additional Proofs for Section 5.1

Proof of Theorem 1: By the definition of the operator Φ, the jth batch $\mathbf{L}^{(j)}$ yields the jth quantile estimator $\Phi\left(\mathbf{L}^{(j)}\right)$ for $j = 1, 2, \ldots, n_b$. The first two statements are obvious from the definitions of $\widehat{v}_{\alpha}^{\mathrm{batch}} = n_b^{-1}\sum_{j=1}^{n_b}\Phi\left(\mathbf{L}^{(j)}\right)$ and $\widehat{v}_{\alpha}^{\mathrm{sect}} = \Phi(\mathbf{L})$, etc.

For the last statement, recall that $\Psi^{(j)} = n_b\Phi(\mathbf{L}) - (n_b - 1)\Phi\left(\widetilde{\mathbf{L}}^{(j)}\right)$, where $\widetilde{\mathbf{L}}^{(j)}$ is the jth jackknifed sample $\mathbf{L}\backslash\mathbf{L}^{(j)}$, $j = 1, 2, \ldots, n_b$. Then, we have

$$\begin{aligned}\mathsf{E}\left[\widehat{v}_{\alpha}^{\text{jack-b}}\right] &= \mathsf{E}\left[\frac{1}{n_b}\sum_{j=1}^{n_b}\Psi^{(j)}\right] \\ &= \mathsf{E}\left[n_b\Phi(\mathbf{L}) - (n_b-1)\Phi\left(\widetilde{\mathbf{L}}^{(1)}\right)\right] \\ &= n_b\left\{v_{\alpha} + \left(-\frac{\lceil n\alpha\rceil}{n(n+1)} + \frac{\varepsilon_{\alpha}^{n}}{n}\right)\frac{1}{f(v_{\alpha})} - \frac{\alpha(1-\alpha)}{2(n+2)}\frac{f'(v_{\alpha})}{f(v_{\alpha})^3} + \mathcal{O}\left(n^{-2}\right)\right\} \\ &\quad -(n_b-1)\left\{v_{\alpha} + \left(-\frac{\lceil (n-n_s)\alpha\rceil}{(n-n_s)(n-n_s+1)} + \frac{\varepsilon_{\alpha}^{n-n_s}}{n-n_s}\right)\frac{1}{f(v_{\alpha})}\right. \\ &\quad \left. -\frac{\alpha(1-\alpha)}{2(n-n_s+2)}\frac{f'(v_{\alpha})}{f(v_{\alpha})^3} + \mathcal{O}\left((n-n_s)^{-2}\right)\right\} \\ &= v_{\alpha} + \frac{\varepsilon_{\alpha}^{n} - \varepsilon_{\alpha}^{n-n_s}}{n_s f(v_{\alpha})} + \mathcal{O}\left((nn_s)^{-1}\right).\end{aligned}$$

Note that this asymptotic derivation is valid because $n - n_s$ tends to infinity as long as there are multiple batches. □

Proof of Proposition 1: Let us fist consider the jackknife variance estimator and then show that $\widehat{\sigma}^2_{\text{jack}}$ and $\widehat{\sigma}^2_{\text{jack-b}}$ are asymptotically equivalent. By definition, $\widehat{v}_\alpha^{\text{jack}} = \Phi(\mathbf{L})$ and we have

$$\begin{aligned}\widehat{\sigma}^2_{\text{jack}} &= \frac{n_b-1}{n_b}\sum_{j=1}^{n_b}\left(\Phi\left(\widetilde{\mathbf{L}}^{(j)}\right)-\Phi(\mathbf{L})\right)^2\\ &= \frac{n_b-1}{n_b}\sum_{j=1}^{n_b}\left(\widetilde{\varphi}^{(j)}-\varphi_n+\widetilde{R}^{(j)}-R_n\right)^2\end{aligned}$$

where $\varphi_n = n^{-1}\sum_{i=1}^n \phi_F(L_i)$ is used for notational simplicity and $\widetilde{\varphi}^{(j)}$ is similarly defined for the jth jackknifed sample $\widetilde{\mathbf{L}}^{(j)} = \mathbf{L}\backslash\mathbf{L}^{(j)}$. The term $\widetilde{R}^{(j)}$ is the corresponding remainder term in the Bahadur representation of $\Phi\left(\widetilde{\mathbf{L}}^{(j)}\right)$.

Straightforward computations yield

$$\mathsf{E}\left[\frac{n(n_b-1)}{n_b}\sum_{j=1}^{n_b}\left(\widetilde{\varphi}^{(j)}-\varphi_n\right)^2\right] = \frac{\alpha(1-\alpha)}{f(v_\alpha)^2} =: \sigma^2. \quad (1)$$

To see this, first note that with $\varphi^{(j)} = n_s^{-1}\sum_{i\in j\text{th batch}}\phi_F(L_i)$ we have

$$\bar{\varphi} := \frac{1}{n_b}\sum_{j=1}^{n_b}\varphi^{(j)} = \frac{1}{n}\sum_{i=1}^{n}\phi_F(L_i) = \varphi_n$$

and thus

$$\begin{aligned}\widetilde{\varphi}^{(j)} &= \frac{1}{n-n_s}\sum_{i\notin j\text{th batch}}\phi_F(L_i)\\ &= \frac{1}{n-n_s}\left(\sum_{i=1}^{n}\phi_F(L_i)-n_s\varphi^{(j)}\right)\\ &= \frac{n_b}{n_b-1}\bar{\varphi}-\frac{1}{n_b-1}\varphi^{(j)}.\end{aligned}$$

This implies that

$$(n_b-1)\sum_{j=1}^{n_b}\left(\widetilde{\varphi}^{(j)}-\varphi_n\right)^2 = \frac{1}{n_b-1}\sum_{j=1}^{n_b}\left(\varphi^{(j)}-\bar{\varphi}\right)^2,$$

which is nothing but the sample variance of i.i.d. observations $\{\varphi^{(j)}\}_{j=1}^{n_b}$ and it is an unbiased estimator of

$$\mathsf{Var}\left(\varphi^{(1)}\right) = \frac{1}{n_s}\mathsf{Var}\left(\phi_F(L_1)\right) = \frac{\sigma^2}{n_s}.$$

Hence, the equality (1) is obtained.

We further note that the expression inside the expectation of the left side of (1) is equal to

$$\frac{n_s}{n_b-1}\sum_{j=1}^{n_b}\left(\varphi^{(j)}-\varphi_n\right)^2 = \frac{1}{n_b-1}\sum_{j=1}^{n_b}\left(\sqrt{n_s}\varphi^{(j)}\right)^2 - \frac{1}{n_b-1}\left(\sqrt{n}\varphi_n\right)^2.$$

The second term converges to zero if $n_b\to\infty$ thanks to the Central Limit Theorem. For the first term, notice that

$$\mathsf{P}\left(\left|\frac{1}{n_b}\sum_{j=1}^{n_b}\left(\sqrt{n_s}\varphi^{(j)}\right)^2-\sigma^2\right|>\epsilon\right) \le \frac{1}{\epsilon^2 n_b}\mathsf{E}\left[n_s^2\left(\varphi^{(j)}\right)^4\right]. \quad (2)$$

Since $\mathsf{E}\left[\phi_F(L_i)\right]=0$ and $\mathsf{E}\left[\phi_F(L_i)^4\right]<\infty$, Lemma 2.2.2B of Serfling (1980) implies

$$\mathsf{E}\left[\left(\sum_{i=1}^{n_s}\phi_F(L_i)\right)^4\right] = \mathcal{O}\left(n_s^2\right) \Rightarrow \mathsf{E}\left[\left(\varphi^{(j)}\right)^4\right] = \mathcal{O}\left(n_s^{-2}\right).$$

Therefore, the right hand side of (2) converges to zero as n_b increases.

On the other hand, Dutt (1973) proves $n^{3/2}\mathsf{E}\left[R_n^2\right]=\mathcal{O}(1)$. Hence, we have

$$\frac{n(n_b-1)}{n_b}\mathsf{E}\left[\sum_{j=1}^{n_b}\left(R_n-\widetilde{R}^{(j)}\right)^2\right]=n(n_b-1)\mathcal{O}\left((n-n_s)^{-3/2}\right)=\mathcal{O}\left(\sqrt{\frac{n_b}{n_s}}\right).$$

Lastly, one can easily see that the Cauchy-Schwarz inequality takes care of the cross product term. Hence, the consistency and asymptotic unbiasedness of $\widehat{\sigma}^2_{\text{jack}}$ follow from the stated conditions.

Now we discuss the asymptotic equivalence of $\widehat{\sigma}^2_{\text{jack}}$ and $\widehat{\sigma}^2_{\text{jack-b}}$. From the definition of $\widehat{\sigma}^2_{\text{jack-b}}$, we have

$$\begin{aligned}\widehat{\sigma}^2_{\text{jack-b}} &= \frac{1}{n_b(n_b-1)}\sum_{j=1}^{n_b}\left(\Psi^{(j)}-\widehat{v}_\alpha^{\text{jack-b}}\right)^2\\ &= \frac{n_b-1}{n_b}\sum_{j=1}^{n_b}\left(\Phi\left(\widetilde{\mathbf{L}}^{(j)}\right)-\frac{1}{n_b}\sum_{k=1}^{n_b}\Phi\left(\widetilde{\mathbf{L}}^{(k)}\right)\right)^2\\ &= \widehat{\sigma}^2_{\text{jack}}-(n_b-1)\left(\frac{1}{n_b}\sum_{j=1}^{n_b}\Phi\left(\widetilde{\mathbf{L}}^{(j)}\right)-\Phi(\mathbf{L})\right)^2\\ &= \widehat{\sigma}^2_{\text{jack}}-\frac{n_b-1}{n_b^2}\left\{\sum_{j=1}^{n_b}\left(\widetilde{R}^{(j)}-R_n\right)\right\}^2.\end{aligned}$$

Since $(\sum_j x_j)^2\le n_b\sum_j x_j^2$, we see the asymptotic equivalence in mean-square of the two estimators by noting that

$$\mathsf{E}\left[\frac{n_b-1}{n_b^2}\left\{\sum_{j=1}^{n_b}\left(\widetilde{R}^{(j)}-R_n\right)\right\}^2\right]\le\mathsf{E}\left[\sum_{j=1}^{n_b}\left(\widetilde{R}^{(j)}-R_n\right)^2\right]=\mathcal{O}\left(n_b^{-1/2}n_s^{-3/2}\right).$$

□

Proof of Proposition 2: The proof is quite similar to that of the previous proposition. With $\mathbf{L}^{(j)}$ representing the jth batch and $\widehat{v}_\alpha^{\text{batch}}$ as the average of $\Phi\left(\mathbf{L}^{(j)}\right)$'s,

$$\begin{aligned}\widehat{\sigma}^2_{\text{batch}} &= \frac{1}{n_b(n_b-1)}\sum_{j=1}^{n_b}\left(\Phi\left(\mathbf{L}^{(j)}\right)-\frac{1}{n_b}\sum_{k=1}^{n_b}\Phi\left(\mathbf{L}^{(k)}\right)\right)^2\\ &= \frac{1}{n_b(n_b-1)}\sum_{j=1}^{n_b}\left(\varphi^{(j)}-\bar{\varphi}+R^{(j)}-\bar{R}\right)^2,\end{aligned}$$

where $\varphi^{(j)}$ and $R^{(j)}$ are the terms in the Bahadur's representation of $\Phi\left(\mathbf{L}^{(j)}\right)$; and $\bar{\varphi}$ and $\bar{R}$ denote their averages, respectively. One can directly obtain

$$\mathsf{E}\left[\frac{n}{n_b(n_b-1)}\sum_{j=1}^{n_b}\left(\varphi^{(j)}-\bar{\varphi}\right)^2\right]=\frac{\alpha(1-\alpha)}{f(v_\alpha)^2}=:\sigma^2.$$

We can also show that the quantity inside the expectation converges to the right hand side in probability. For this, note that

$$\frac{n}{n_b(n_b-1)}\sum_{j=1}^{n_b}\left(\varphi^{(j)}-\bar{\varphi}\right)^2=\frac{n}{n_b(n_b-1)}\left\{\sum_{j=1}^{n_b}\left(\varphi^{(j)}\right)^2-n_b\bar{\varphi}^2\right\}.$$

The second term is asymptotically equivalent to $\left(\sqrt{n_s}\bar{\varphi}\right)^2$ for which we compute

$$\mathsf{P}\left(|\sqrt{n_s}\bar{\varphi}|>\epsilon\right)\le\frac{n_s\mathsf{E}\left[\bar{\varphi}^2\right]}{\epsilon^2}=\frac{n_s\mathsf{Var}\left(\varphi^{(1)}\right)}{n_b\epsilon^2}=\frac{\sigma^2}{n_b\epsilon^2}\to 0,$$

where $\epsilon > 0$ is arbitrarily given. Hence, by the continuous mapping theorem, the second term converges to zero in probability. Next, note that we can apply the same arguments as given in the proof of Proposition 1 to see that the first term converges to σ^2 in probability.

On the other hand, from Kiefer (1967) we get the exact order of $R^{(j)}$. Using this result, it can be seen that

$$\frac{n}{n_b(n_b-1)}\sum_{j=1}^{n_b}\left(R^{(j)}-\bar{R}\right)^2=\mathcal{O}\left(n_s^{-1/2}\left(\log\log n_s\right)^{3/2}\right)$$

with probability 1. Furthermore, Dutt (1973) implies that the expected value of the summation above satisfies

$$\frac{n}{n_b(n_b-1)}\mathsf{E}\left[\sum_{j=1}^{n_b}\left(R^{(j)}\right)^2-n_b\bar{R}^2\right]=n_s\mathcal{O}\left(n_s^{-2/3}\right)=\mathcal{O}\left(n_s^{-1/2}\right).$$

As shown in the proof of Proposition 1, the remaining terms can be handled by the Cauchy-Schwarz inequality.

The consistency and asymptotic unbiasedness of $\widehat{\sigma}^2_{\text{sect}}$ are derived based on the asymptotic equivalence of two estimators. Specifically,

$$\begin{aligned}
\widehat{\sigma}^2_{\text{sect}} &= \frac{1}{n_b(n_b-1)}\sum_{j=1}^{n_b}\left(\Phi\left(\mathbf{L}^{(j)}\right)-\Phi(\mathbf{L})\right)^2\\
&= \frac{1}{n_b(n_b-1)}\sum_{j=1}^{n_b}\left(\Phi\left(\mathbf{L}^{(j)}\right)-\frac{1}{n_b}\sum_{k=1}^{n_b}\Phi\left(\mathbf{L}^{(k)}\right)+\frac{1}{n_b}\sum_{k=1}^{n_b}\Phi\left(\mathbf{L}^{(k)}\right)-\Phi(\mathbf{L})\right)^2\\
&= \widehat{\sigma}^2_{\text{batch}}+\frac{1}{n_b-1}\left(\frac{1}{n_b}\sum_{k=1}^{n_b}\Phi\left(\mathbf{L}^{(k)}\right)-\Phi(\mathbf{L})\right)^2\\
&= \widehat{\sigma}^2_{\text{batch}}+\frac{1}{n_b^2(n_b-1)}\left\{\sum_{j=1}^{n_b}\left(\varphi^{(j)}-\varphi_n+R^{(j)}-R_n\right)\right\}^2\\
&= \widehat{\sigma}^2_{\text{batch}}+\frac{1}{n_b^2(n_b-1)}\left\{\sum_{j=1}^{n_b}\left(R^{(j)}-R_n\right)\right\}^2
\end{aligned}$$

where φ_n, R_n are the second and third terms in the Bahadur's representation based on a sample $\mathbf{L}$. Note that from $(\sum_j x_j)^2 \le n_b\sum_j x_j^2$, we have

$$\frac{1}{n_b^2(n_b-1)}\mathsf{E}\left[\left\{\sum_{j=1}^{n_b}\left(R^{(j)}-R_n\right)\right\}^2\right]\le\frac{1}{n_b(n_b-1)}\mathcal{O}\left(n_s^{-3/2}\right).$$

Then the result for $\widehat{\sigma}^2_{\text{sect}}$ follows easily. □

Proof of Theorem 2: <u>*Case 1*</u>. The underlying idea is similar to that given in Theorem 1 of Shao (1989). By Proposition 1, the squared bias term has the order $o(n^{-2})$. Hence, we are only concerned with the variance of $\widehat{\sigma}^2_{\text{jack}}$ and $\widehat{\sigma}^2_{\text{jack-b}}$. Recall that

$$\widehat{\sigma}^2_{\text{jack}}=\frac{n_b-1}{n_b}\left\{\sum_{j=1}^{n_b}\left(\widetilde{\varphi}^{(j)}-\varphi_n\right)^2+\sum_{j=1}^{n_b}\left(\widetilde{R}^{(j)}-R_n\right)^2+2\sum_{j=1}^{n_b}\left(\widetilde{\varphi}^{(j)}-\varphi_n\right)\left(\widetilde{R}^{(j)}-R_n\right)\right\}.$$

As shown in Proposition 1, we have

$$\frac{n_b-1}{n_b}\sum_{j=1}^{n_b}\left(\widetilde{\varphi}^{(j)}-\varphi_n\right)^2=\frac{1}{n_b(n_b-1)}\sum_{j=1}^{n_b}\left(\varphi^{(j)}-\bar{\varphi}\right)^2.$$

From the standard result on the variance of the sample variance, we obtain

$$\begin{aligned}\mathsf{Var}\left(\frac{1}{n_b(n_b-1)}\sum_{j=1}^{n_b}\left(\varphi^{(j)}-\bar{\varphi}\right)^2\right) &= \mathcal{O}\left(\frac{1}{n_b^3}\right)\mathsf{E}\left[\left(\varphi^{(1)}\right)^4\right]\\ &= \mathcal{O}\left(\frac{1}{n_b^3 n_s^2}\right).\end{aligned}$$

For the second term, we have

$$\begin{aligned}\mathsf{Var}\left(\frac{n_b-1}{n_b}\sum_{j=1}^{n_b}\left(\widetilde{R}^{(j)}-R_n\right)^2\right) &\le n_b\mathsf{E}\left[\sum_{j=1}^{n_b}\left(\widetilde{R}^{(j)}-R_n\right)^4\right]\\ &= o\left(\frac{n_b^2}{(n-n_s)^2}\right) = o\left(n_s^{-2}\right),\end{aligned}$$

where Lemma 5 is applied in the equality. Lastly, the Cauchy-Schwarz inequality yields

$$\begin{aligned}\mathsf{Var}\left(\sum_{j=1}^{n_b}\left(\widetilde{\varphi}^{(j)}-\varphi_n\right)\left(\widetilde{R}^{(j)}-R_n\right)\right) &\le \mathsf{E}\left[\left(\sum_{j=1}^{n_b}\left(\widetilde{\varphi}^{(j)}-\varphi_n\right)\left(\widetilde{R}^{(j)}-R_n\right)\right)^2\right]\\ &\le \mathsf{E}\left[\sum_{j=1}^{n_b}\left(\widetilde{\varphi}^{(j)}-\varphi_n\right)^2\sum_{j=1}^{n_b}\left(\widetilde{R}^{(j)}-R_n\right)^2\right]\\ &\le \sqrt{\mathsf{Var}\left(\sum_{j=1}^{n_b}\left(\widetilde{\varphi}^{(j)}-\varphi_n\right)^2\right)\mathsf{Var}\left(\sum_{j=1}^{n_b}\left(\widetilde{R}^{(j)}-R_n\right)^2\right)}\\ &\quad +\mathsf{E}\left[\sum_{j=1}^{n_b}\left(\widetilde{\varphi}^{(j)}-\varphi_n\right)^2\right]\mathsf{E}\left[\sum_{j=1}^{n_b}\left(\widetilde{R}^{(j)}-R_n\right)^2\right].\end{aligned}$$

Then previous calculations lead to the conclusion that the right hand side is of order $o\left((nn_s)^{-1}\right)$.

The difference between $\widehat{\sigma}^2_{\text{jack}}$ and $\widehat{\sigma}^2_{\text{jack-b}}$ is given by

$$\frac{n_b-1}{n_b^2}\left\{\sum_{j=1}^{n_b}\left(\widetilde{R}^{(j)}-R_n\right)\right\}^2.$$

Similarly as shown above, the variance of this term can be shown to be of order $o\left(n_s^{-2}\right)$. Hence, the MSE of $\widehat{\sigma}^2_{\text{jack-b}}$ has the same order as $\widehat{\sigma}^2_{\text{jack}}$.

Case 2. By Proposition 2, we again consider the variances of $\widehat{\sigma}^2_{\text{sect}}$ and $\widehat{\sigma}^2_{\text{batch}}$ only. Let us recall that

$$\widehat{\sigma}^2_{\text{batch}} = \frac{1}{n_b(n_b-1)}\left\{\sum_{j=1}^{n_b}\left(\varphi^{(j)}-\bar{\varphi}\right)^2+\sum_{j=1}^{n_b}\left(R^{(j)}-\bar{R}\right)^2+2\sum_{j=1}^{n_b}\left(\varphi^{(j)}-\bar{\varphi}\right)\left(R^{(j)}-\bar{R}\right)\right\}.$$

The first term has been tackled above. For the second term, we have

$$\begin{aligned}\mathsf{Var}\left(\frac{1}{n_b(n_b-1)}\sum_{j=1}^{n_b}\left(R^{(j)}-\bar{R}\right)^2\right) &\le \frac{1}{n_b(n_b-1)^2}\mathsf{E}\left[\sum_{j=1}^{n_b}\left(R^{(j)}-\bar{R}\right)^4\right]\\ &= o\left(\frac{1}{n_b^2 n_s^2}\right),\end{aligned}$$

where in the second equality we applied Lemma 5. The last term can be similarly handled together with the Cauchy-Schwarz inequality, yielding an order of $o\left(n^{-2}\right)$.

As for $\widehat{\sigma}^2_{\text{sect}}$, notice that from the last part of the proof of Proposition 2, the difference between $\widehat{\sigma}^2_{\text{batch}}$ and $\widehat{\sigma}^2_{\text{sect}}$ is given by

$$\frac{1}{n_b^2(n_b-1)}\left\{\sum_{j=1}^{n_b}\left(R^{(j)}-R_n\right)\right\}^2.$$

The assertion follows because

$$\begin{aligned}\mathsf{Var}\left(\frac{1}{n_b^2(n_b-1)}\left\{\sum_{j=1}^{n_b}\left(R^{(j)}-R_n\right)\right\}^2\right) &\le \mathsf{E}\left[\frac{1}{n_b^4(n_b-1)^2}\left\{\sum_{j=1}^{n_b}\left(R^{(j)}-R_n\right)\right\}^4\right]\\ &\le \frac{n_b^3}{n_b^4(n_b-1)^2}\mathsf{E}\left[\sum_{j=1}^{n_b}\left(R^{(j)}-R_n\right)^4\right]\\ &= o\left(\frac{1}{(n_b-1)^2}\cdot\frac{1}{n_s^2}\right),\end{aligned}$$

which is again obtained by applying Lemma 5. This completes the proof. □

Appendix C: Additional Proofs for Section 5.2

Proof of Lemma 2: We utilize the conditioning argument presented in Gribkova and Helmers (2006). Recall that $F^{-1}(U)$ with uniform distribution U on the unit interval $[0,1]$ gives the loss distribution. Hence, $L_{(i)}$ can be written as $F^{-1}(U_{(i)})$ where $U_{(i)}$ is the ith order statistics of n i.i.d. uniform random variables. With $k=\lceil n\alpha\rceil$, observe that

$$\widehat{c}_\alpha^n = \frac{k-n\alpha}{n(1-\alpha)}L_{(k)} + \frac{1}{n(1-\alpha)}\sum_{i=k+1}^{n} L_{(i)},$$

and

$$\begin{aligned}\mathsf{E}\left[\sum_{i=k+1}^{n} L_{(i)}\right] &= \mathsf{E}\left[\sum_{i=k+1}^{n} F^{-1}\left(U_{(i)}\right)\right]\\ &= \mathsf{E}\left[\mathsf{E}\left[\sum_{i=k+1}^{n} F^{-1}\left(U_{(i)}\right)|U_{(k)}\right]\right]\\ &= \mathsf{E}\left[\frac{n-k}{1-U_{(k)}}\int_{U_{(k)}}^{1} F^{-1}(t)dt\right],\end{aligned}$$

where we used that conditional on $U_{(k)}$, $(U_{(k+1)},\dots,U_{(n)})$ is the order statistics of uniform random variables on $(U_{(k)},1)$. If we define $I(x)=(1-x)^{-1}\int_x^1 F^{-1}(t)dt$, then $I(\alpha)=c_\alpha$ and

$$I'(\alpha) = \frac{1}{(1-\alpha)^2}\int_\alpha^1 F^{-1}(t)dt - \frac{1}{1-\alpha}F^{-1}(\alpha) = \frac{c_\alpha - v_\alpha}{1-\alpha}.$$

Furthermore,

$$\begin{aligned}I''(\alpha) &= \frac{2}{(1-\alpha)^3}\int_\alpha^1 F^{-1}(t)dt - \frac{2F^{-1}(\alpha)}{(1-\alpha)^2} - \frac{1}{(1-\alpha)f(v_\alpha)}\\ &= -\frac{2}{1-\alpha}\left\{\frac{1}{2f(v_\alpha)} - \frac{c_\alpha - v_\alpha}{1-\alpha}\right\}.\end{aligned}$$

For notational simplicity, we write $U_{(k)}$ as X_n. As a next step, consider a function

$$H(x) = I(x) - I(\alpha) - I'(\alpha)(x-\alpha) - \frac{I''(\alpha)}{2}(x-\alpha)^2.$$

It is obvious that $H(\alpha)=H'(\alpha)=H''(\alpha)=0$. The following relationship also becomes useful:

$$I(x) - I(\alpha) = \frac{x-\alpha}{1-\alpha}\frac{1}{1-x}\int_x^1 F^{-1}(t)dt + \frac{1}{1-\alpha}\int_x^\alpha F^{-1}(t)dt.$$

Then, using the above equation, we observe that $H(X_n)$ can be explicitly written as

$$
\begin{aligned}
H(X_n) &= \frac{X_n-\alpha}{1-\alpha}\frac{1}{1-X_n}\int_{X_n}^{1} F^{-1}(t)dt + \frac{1}{1-\alpha}\int_{X_n}^{\alpha} F^{-1}(t)dt \\
&\quad -\left\{\frac{1}{(1-\alpha)^2}\int_{\alpha}^{1} F^{-1}(t)dt - \frac{F^{-1}(\alpha)}{1-\alpha}\right\}(X_n-\alpha) \\
&\quad -\left\{\frac{1}{(1-\alpha)^3}\int_{\alpha}^{1} F^{-1}(t)dt - \frac{F^{-1}(\alpha)}{(1-\alpha)^2} - \frac{1}{2(1-\alpha)f(v_\alpha)}\right\}(X_n-\alpha)^2 \\
&= \frac{X_n-\alpha}{1-\alpha}\{I(X_n)-I(\alpha)\} + \frac{1}{1-\alpha}\left\{\int_{X_n}^{\alpha} F^{-1}(t)dt - \int_{X_n}^{\alpha} F^{-1}(\alpha)dt\right\} \\
&\quad -\left\{\frac{1}{(1-\alpha)^3}\int_{\alpha}^{1} F^{-1}(t)dt - \frac{F^{-1}(\alpha)}{(1-\alpha)^2} - \frac{1}{2(1-\alpha)f(v_\alpha)}\right\}(X_n-\alpha)^2 \\
&= \frac{X_n-\alpha}{1-\alpha}\left\{\frac{X_n-\alpha}{1-\alpha}\frac{1}{1-X_n}\int_{X_n}^{1} F^{-1}(t)dt + \frac{1}{1-\alpha}\int_{X_n}^{\alpha} F^{-1}(t)dt\right\} \\
&\quad + \frac{1}{1-\alpha}\int_{X_n}^{\alpha}\left(F^{-1}(t)-F^{-1}(\alpha)\right)dt \\
&\quad -\left\{\frac{1}{(1-\alpha)^3}\int_{\alpha}^{1} F^{-1}(t)dt - \frac{F^{-1}(\alpha)}{(1-\alpha)^2} - \frac{1}{2(1-\alpha)f(v_\alpha)}\right\}(X_n-\alpha)^2 \\
&= \left(\frac{X_n-\alpha}{1-\alpha}\right)^2\{I(X_n)-I(\alpha)\} + \frac{X_n-\alpha}{(1-\alpha)^2}\left\{\int_{X_n}^{\alpha}\left(F^{-1}(t)-F^{-1}(\alpha)\right)dt\right\} \\
&\quad + \frac{1}{1-\alpha}\left\{\int_{X_n}^{\alpha}\left(F^{-1}(t)-F^{-1}(\alpha)-\frac{t-\alpha}{f(v_\alpha)}\right)dt\right\}.
\end{aligned}
$$

Let us denote each term by A_n, B_n, and C_n, respectively. We claim that $n^{3/2}|A_n|$ (and similarly for the other two) is bounded in expectation. To see this, note that $F^{-1}(U) \overset{d}{=} L$, $F^{-1}(X_n) = \widehat{v}_\alpha^n$. Hence,

$$
\begin{aligned}
|A_n| &= \left(\frac{X_n-\alpha}{1-\alpha}\right)^2\left|\frac{X_n-\alpha}{1-\alpha}\frac{1}{1-X_n}\int_{X_n}^{1} F^{-1}(t)dt + \frac{1}{1-\alpha}\int_{X_n}^{\alpha} F^{-1}(t)dt\right| \\
&\le \frac{|X_n-\alpha|^3}{(1-\alpha)^3}\frac{\mathsf{E}[|L|]}{1-X_n} + \frac{|X_n-\alpha|^3}{(1-\alpha)^3}\left(|v_\alpha| + |\widehat{v}_\alpha^n|\right).
\end{aligned}
$$

For all suitably large n, we have $\mathsf{E}\left[\left(n^{3/2}|X_n-\alpha|^3\right)^2\right] < \infty$, $\mathsf{E}\left[(1-X_n)^{-2}\right] < \infty$, and the finite variance of the quantile estimator. Subsequently, $\mathsf{E}\left[n^{3/2}|A_n|\right] = \mathcal{O}(1)$ follows from the Cauchy-Schwarz inequality. All the relevant results on sample quantiles can be found in David and Nagaraja (2003) and Serfling (1980).

Regarding $n^{3/2}B_n$, we see that

$$
|B_n| \le \frac{(X_n-\alpha)^2}{(1-\alpha)^2}\left|v_\alpha - \widehat{v}_\alpha^n\right|,
$$

since F^{-1} is non-decreasing. The fact that $\mathsf{E}\left[n^{3/2}|B_n|\right] = \mathcal{O}(1)$ also easily follows using similar arguments as given above. In particular, we used the asymptotic normality of the quantile estimator.

For the last term C_n, take a small $\epsilon > 0$ such that the density f is positive and differentiable on the region $(F^{-1}(\alpha-\epsilon), F^{-1}(\alpha+\epsilon))$. On $\{|X_n-\alpha| < \epsilon\}$, C_n can be expressed as

$$
\begin{aligned}
C_n &= \frac{1}{1-\alpha}\int_{X_n}^{\alpha}\int_{t}^{\alpha}\left\{\frac{1}{f(v_\alpha)} - \frac{1}{f(F^{-1}(z))}\right\}dzdt \\
&= \frac{1}{1-\alpha}\int_{X_n}^{\alpha}\left\{\frac{1}{f(v_\alpha)} - \frac{1}{f(F^{-1}(z))}\right\}(z-X_n)dz.
\end{aligned}
$$

By the mean value theorem,

$$
\left|\frac{1}{f(v_\alpha)} - \frac{1}{f(F^{-1}(z))}\right| = \left|\frac{f'\left(F^{-1}(z^*)\right)}{f^3\left(F^{-1}(z^*)\right)}\right| \cdot \left|v_\alpha - F^{-1}(z)\right| \le C_\epsilon\left|v_\alpha - F^{-1}(z)\right|
$$

for some $z^* \in (\alpha-\epsilon, \alpha+\epsilon)$ and C_ϵ is some constant that depends on ϵ. Consequently, on $\{|X_n-\alpha|<\epsilon\}$,

$$|C_n| \le \frac{C_\epsilon}{2(1-\alpha)}(X_n-\alpha)^2\,|\widehat{v}_\alpha^n - v_\alpha| =: C_n^{(1)}.$$

On the other hand, when $|X_n-\alpha|\ge\epsilon$, we get

$$|C_n| \le \left\{|\widehat{v}_\alpha^n - v_\alpha| + \frac{|X_n-\alpha|}{f(v_\alpha)}\right\}|X_n-\alpha| \le |\widehat{v}_\alpha^n - v_\alpha| + \frac{1}{f(v_\alpha)} =: C_n^{(2)}.$$

Therefore,

$$\begin{aligned}
\mathsf{E}\left[n^{3/2}|C_n|\right] &\le \mathsf{E}\left[n^{3/2}C_n^{(1)}\mathbf{1}_{|X_n-\alpha|<\epsilon}\right] + \mathsf{E}\left[n^{3/2}C_n^{(2)}\mathbf{1}_{|X_n-\alpha|\ge\epsilon}\right] \\
&\le \mathsf{E}\left[n^{3/2}C_n^{(1)}\right] + \left\{\mathsf{E}\left[\left(C_n^{(2)}\right)^2\right] n^3\mathsf{P}\left(|X_n-\alpha|\ge\epsilon\right)\right\}^{1/2}.
\end{aligned}$$

The finiteness of the first term on the right-hand side follows from the Cauchy-Schwarz inequality as used for proving results on A_n and B_n. For the second term, we note that $\mathsf{P}\left(|X_n-\alpha|\ge\epsilon\right) \le 2e^{-2n\delta_\epsilon^2}$ for some constant δ_ϵ. See Theorem 2.3.2 in Serfling (1980). This makes the second term finite as n increases. These two observations ensure $\mathsf{E}\left[n^{3/2}|C_n|\right] = \mathcal{O}(1)$.

The previous result implies that we can write

$$\frac{1}{n-k}\mathsf{E}\left[\sum_{i=k+1}^{n} L_{(i)}\right] = I(\alpha) + I'(\alpha)\mathsf{E}\left[U_{(k)}-\alpha\right] + \frac{I''(\alpha)}{2}\mathsf{E}\left[\left(U_{(k)}-\alpha\right)^2\right] + \mathcal{O}\left(n^{-3/2}\right).$$

The standard results given by David and Nagaraja (2003) read $\mathsf{E}[U_{(k)}] = k/(n+1) =: \alpha_k$ and $\mathsf{Var}(U_{(k)}) = \alpha_k(1-\alpha_k)/(n+2)$. With $\varepsilon = \alpha - k/(n+1) = \mathcal{O}(n^{-1})$, straightforward calculations yield

$$\begin{aligned}
\frac{1}{n-k}\mathsf{E}\left[\sum_{i=k+1}^{n} L_{(i)}\right] &= I(\alpha) - I'(\alpha)\varepsilon + \frac{I''(\alpha)}{2}\left(\frac{\alpha_k(1-\alpha_k)}{(n+2)} + \varepsilon^2\right) + \mathcal{O}\left(n^{-3/2}\right) \\
&= I(\alpha) - I'(\alpha)\varepsilon + I''(\alpha)\frac{\alpha(1-\alpha)}{2n} + \mathcal{O}\left(n^{-3/2}\right) \\
&= c_\alpha + \left(\frac{k}{n+1}-\alpha\right)\frac{c_\alpha - v_\alpha}{1-\alpha} - \frac{\alpha}{n}\left(\frac{1}{2f(v_\alpha)} - \frac{c_\alpha - v_\alpha}{1-\alpha}\right) + \mathcal{O}\left(n^{-3/2}\right) \\
&= c_\alpha - \frac{\alpha}{2nf(v_\alpha)} + \left(\frac{k}{n+1} - \alpha + \frac{\alpha}{n}\right)\frac{c_\alpha - v_\alpha}{1-\alpha} + \mathcal{O}\left(n^{-3/2}\right).
\end{aligned}$$

Let us denote the last expression by Υ for notational simplicity.

Recall the following result for the quantiles (David and Nagaraja 2003, Ressler and Lewis 1990): With $\alpha_k = k/(n+1)$,

$$\mathsf{E}\left[\widehat{v}_\alpha^n\right] = v_\alpha + \frac{\alpha_k - \alpha}{f(v_\alpha)} - \frac{\alpha(1-\alpha)}{2(n+2)}\frac{f'(v_\alpha)}{f(v_\alpha)^3} + \mathcal{O}\left(n^{-2}\right).$$

Consequently, we have

$$\begin{aligned}
\mathsf{E}\left[\widehat{c}_\alpha^n\right] &= \frac{k-n\alpha}{n(1-\alpha)}\mathsf{E}\left[\widehat{v}_\alpha^n\right] + \frac{1}{n(1-\alpha)}\mathsf{E}\left[\sum_{i=k+1}^{n} L_{(i)}\right] \\
&= \frac{k-n\alpha}{n(1-\alpha)}v_\alpha + \frac{n-k}{n(1-\alpha)}\Upsilon + \mathcal{O}\left(n^{-3/2}\right) \\
&= \Upsilon + \frac{n\alpha-k}{n(1-\alpha)}\left(-v_\alpha + \Upsilon\right) + \mathcal{O}\left(n^{-3/2}\right) \\
&= \Upsilon + \left(\alpha - \frac{k}{n}\right)\frac{c_\alpha - v_\alpha}{1-\alpha} + \mathcal{O}\left(n^{-3/2}\right) \\
&= c_\alpha - \frac{\alpha}{2nf(v_\alpha)} + \mathcal{O}\left(n^{-3/2}\right).
\end{aligned}$$

□

Proof of Theorem 3: The first two statements follow easily by observing that $\widehat{c}_\alpha^{\text{batch}}$ is the average of i.i.d. estimates of c_α, each of which is calculated based on a sample of size n_s; while the other three estimators in the second statement are obtained from a sample of size n.

For the last statement, recall the definition of pseudovalues and that of jackknife bias-corrected estimator given in Equation (2), it follows that

$$\begin{aligned}
\mathsf{E}\left[\widehat{c}_\alpha^{\text{jack-b}}\right] &= \mathsf{E}\left[\frac{1}{n_b}\sum_{i=1}^{n_b}\Psi^{(j)}\right] = \mathsf{E}\left[\frac{1}{n_b}\sum_{j=1}^{n_b}\left\{n_b\Phi(\mathbf{L}) - (n_b-1)\Phi\left(\widetilde{\mathbf{L}}^{(j)}\right)\right\}\right] \\
&= \mathsf{E}\left[n_b\widehat{c}_\alpha^{n} - (n_b-1)\widehat{c}_\alpha^{(n_b-1)n_s}\right] \\
&= n_b\left(c_\alpha - \frac{\alpha}{2nf(v_\alpha)} + \mathcal{O}\left(n^{-3/2}\right)\right) \\
&\quad -(n_b-1)\left(c_\alpha - \frac{\alpha}{2(n_b-1)n_s f(v_\alpha)} + \mathcal{O}\left((n_b-1)^{-3/2}n_s^{-3/2}\right)\right) \\
&= c_\alpha + \mathcal{O}\left(n_s^{-3/2}n_b^{-1/2}\right).
\end{aligned}$$

Here $\Psi^{(j)}$ is the jth pseudovalue, $n_b\Phi(\mathbf{L})-(n_b-1)\Phi\left(\widetilde{\mathbf{L}}^{(j)}\right)$; Φ represents the operator that estimates c_α based on a given sample, here either on the entire sample $\mathbf{L}$ or the jth jackknifed sample $\mathbf{L}^{(j)}$ for $j = 1, 2, \ldots, n_b$.
□

LEMMA 3. *Suppose that Assumption 1 holds. For a given $\epsilon > 0$ and almost all i.i.d. sequences $\omega = \{L_i\}_{i=1}^{\infty}$, there exists an integer $n_\epsilon(\omega)$ such that if $n, m > n_\epsilon(\omega)$, then*

$$\left|\widehat{c}_\alpha^{n} - \widehat{c}_\alpha^{m} - \int \phi_F(x)d(F_n - F_m)(x)\right| \le \epsilon \|F_n - F_m\|_\infty$$

where $\phi_F(x) = -\int\{\mathbf{1}_{\{y\ge x\}} - F(y)\}J(F(y))dy$, F_n and F_m depend on any (overlapping) subsets of ω, and $\|\cdot\|_\infty$ is the usual supremum norm.

Proof: *Step 1.* Let us temporarily call $\int_0^1 F^{-1}(t)J(t)dt$ by $T(F)$. We introduce a functional $T_a(\cdot)$ and a function $\psi_a(\cdot)$ for a fixed real value $a > v_\alpha$:

$$T_a(G) = -\int_{-\infty}^{a}\int_0^{G(x)} J(u)dudx, \quad \psi_a(x) = -\int_{-\infty}^{a}\left\{\mathbf{1}_{\{y\ge x\}} - F(y)\right\}J(F(y))dy,$$

where G denotes an arbitrary distribution. Let us fix two arbitrary distributions G, H with finite means. Note first that $\int|\psi_a(x)|d(G+H)(x) < \infty$ because

$$\int|\psi_a(x)|d(G+H)(x) \le \int\int_{-\infty}^{a} 2J(F(y))dyd(G+H)(x) \le \frac{4(a-v_\alpha)}{1-\alpha} < \infty.$$

Hence, by Fubini's theorem, we obtain

$$\begin{aligned}
&T_a(H) - T_a(G) - \int\psi_a(x)d(H-G)(x) \\
&= T_a(H) - T_a(G) + \int_{-\infty}^{a}\left(\mathsf{E}^H\left[J(F(y))\left(\mathbf{1}_{\{y\ge X\}} - F(y)\right)\right] - \mathsf{E}^G\left[J(F(y))\left(\mathbf{1}_{\{y\ge X\}} - F(y)\right)\right]\right)dy \\
&= \int_{-\infty}^{a}\left\{\int_0^{G(x)}J(u)du - \int_0^{H(x)}J(u)du - \left(G(x)-H(x)\right)J(F(x))\right\}dx \\
&= \int_{-\infty}^{a}\left\{\int_{F(x)}^{G(x)}\left(J(u)-J(F(x))\right)du - \int_{F(x)}^{H(x)}\left(J(u)-J(F(x))\right)du\right\}dx
\end{aligned}$$

where E^H and E^G denote taking expectations when X follows distribution H and G, respectively.

Next, we set $\Upsilon_{G,F}(x) := \int_{F(x)}^{G(x)} \Big(J(u) - J(F(x))\Big) du$. If $G(x) > F(x)$, then

$$\begin{aligned}(1-\alpha)\Upsilon_{G,F}(x) &= \int_{F(x)}^{G(x)} \Big(\mathbf{1}_{\{u>\alpha\}} - \mathbf{1}_{\{F(x)>\alpha\}}\Big) du = \int_{F(x)}^{G(x)} \mathbf{1}_{\{F(x)\le\alpha<u\}} du \\ &= \mathbf{1}_{\{F(x)\le\alpha\}} \int_{\alpha}^{G(x)} \mathbf{1}_{\{u>\alpha\}} du = \mathbf{1}_{\{F(x)\le\alpha\}} \times \mathbf{1}_{\{G(x)>\alpha\}} (G(x)-\alpha).\end{aligned}$$

Similarly if $G(x) < F(x)$, then we get

$$\begin{aligned}(1-\alpha)\Upsilon_{G,F}(x) &= \int_{G(x)}^{F(x)} \Big(\mathbf{1}_{\{F(x)>\alpha\}} - \mathbf{1}_{\{u>\alpha\}}\Big) du = \int_{G(x)}^{F(x)} \mathbf{1}_{\{F(x)>\alpha\ge u\}} du \\ &= \mathbf{1}_{\{F(x)>\alpha\}} \int_{G(x)}^{\alpha} \mathbf{1}_{\{u\le\alpha\}} du = \mathbf{1}_{\{F(x)>\alpha\}} \times \mathbf{1}_{\{G(x)\le\alpha\}} (\alpha - G(x)).\end{aligned}$$

These two cases can be rewritten succinctly as follows:

$$(1-\alpha)\Upsilon_{G,F}(x) = \mathbf{1}_{\{F(x)\le\alpha\}} (G(x)-\alpha)^+ + \mathbf{1}_{\{F(x)>\alpha\}} (\alpha - G(x))^+.$$

Utilizing this equality, we compute

$$\begin{aligned}&(1-\alpha)\int_{-\infty}^{a} \Big|\Upsilon_{G,F}(x) - \Upsilon_{H,F}(x)\Big| dx \\ =& \int_{-\infty}^{a} \Big|\mathbf{1}_{\{F(x)\le\alpha\}} \big((G(x)-\alpha)^+ - (H(x)-\alpha)^+\big) + \mathbf{1}_{\{F(x)>\alpha\}} \big((\alpha - G(x))^+ - (\alpha - H(x))^+\big)\Big| dx \\ \le& \|G-H\|_\infty \int_{-\infty}^{a} \Big(\mathbf{1}_{\{F(x)\le\alpha\le G(x)\}} + \mathbf{1}_{\{F(x)\le\alpha\le H(x)\}} + \mathbf{1}_{\{F(x)\le\alpha\le G(x)\}}\mathbf{1}_{\{F(x)\le\alpha\le H(x)\}} \\ &+ \mathbf{1}_{\{G(x)<\alpha<F(x)\}} + \mathbf{1}_{\{H(x)<\alpha<F(x)\}} + \mathbf{1}_{\{G(x)<\alpha<F(x)\}}\mathbf{1}_{\{H(x)<\alpha<F(x)\}}\Big) dx.\end{aligned}$$

For the preceding inequality, we used the following simple inequalities:

$$\begin{aligned}\big|(x-\alpha)^+ - (y-\alpha)^+\big| &\le |x-y|\Big(\mathbf{1}_{\{x<\alpha\le y\}} + \mathbf{1}_{\{y<\alpha\le x\}} + \mathbf{1}_{\{x\ge\alpha\}}\mathbf{1}_{\{y\ge\alpha\}}\Big), \\ \big|(\alpha-x)^+ - (\alpha-y)^+\big| &\le |x-y|\Big(\mathbf{1}_{\{x<\alpha\le y\}} + \mathbf{1}_{\{y<\alpha\le x\}} + \mathbf{1}_{\{x<\alpha\}}\mathbf{1}_{\{y<\alpha\}}\Big).\end{aligned}$$

Since $G(x) \ge \alpha$ and $G(x) < \alpha$ imply $G^{-1}(\alpha) \le x$ and $x \le G^{-1}(\alpha)$, respectively, we see that together with Assumption 1

$$\mathbf{1}_{\{F(x)\le\alpha\le G(x)\}} + \mathbf{1}_{\{G(x)<\alpha<F(x)\}} \le \mathbf{1}_{\{G^{-1}(\alpha)\le x\le F^{-1}(\alpha)\}} + \mathbf{1}_{\{F^{-1}(\alpha)\le x\le G^{-1}(\alpha)\}}.$$

A similar inequality for H holds as well. Consequently, we have $(1-\alpha)\int_{-\infty}^{a} |\Upsilon_{G,F}(x) - \Upsilon_{H,F}(x)| dx$ bounded above by

$$\Psi_{G,H} := 3\|G-H\|_\infty \Big(\big|G^{-1}(\alpha) - F^{-1}(\alpha)\big| + \big|H^{-1}(\alpha) - F^{-1}(\alpha)\big|\Big).$$

In terms of T_a and ψ_a, this result says

$$\left|T_a(H) - T_a(G) - \int \psi_a(x) d(H-G)(x)\right| \le \frac{\Psi_{G,H}}{1-\alpha}.$$

Step 2. Next, let us fix two distribution functions G, H with finite means and take a to be large enough to ensure $a > \max\{G^{-1}(\alpha), H^{-1}(\alpha)\}$ in addition to $a > v_\alpha$. We start by computing $(1-\alpha)T(G)$ (similarly for

$T(H)$ as well). When U is a uniform random variable on $(0,1)$ and $X = G^{-1}(U)$, it is well known that X is distributed as G. Suppose that $G(t-) \le \alpha < G(t)$ for some t. Then,

$$\begin{aligned}(1-\alpha)T(G) &= \int_\alpha^1 G^{-1}(u)du = \mathsf{E}\left[G^{-1}(U)\mathbf{1}_{\{U\ge\alpha\}}\right] \\ &= \mathsf{E}\left[G^{-1}(U)\left(\mathbf{1}_{\{\alpha\le U<G(t)\}} + \mathbf{1}_{\{U\ge G(t)\}}\right)\right] \\ &= t(G(t)-\alpha) + \mathsf{E}^G\left[X\mathbf{1}_{\{X>t\}}\right] \\ &= t(G(t)-\alpha) + \int y\mathbf{1}_{\{y>t\}}dG(y).\end{aligned}$$

The third equality follows because, for given t, $\{u|u > G(t)\} \subset \{u|G^{-1}(u) > t\} \subset \{u|u \ge G(t)\}$ and $\mathsf{P}(U > G(t)) = \mathsf{P}(U \ge G(t))$.

On the other hand, as for $T_a(G)$, we have

$$\begin{aligned}(1-\alpha)T_a(G) &= -\int_{-\infty}^a \int_0^{G(x)} \mathbf{1}_{\{u>\alpha\}}dudx = -\int_t^a \int_0^{G(x)} \mathbf{1}_{\{u>\alpha\}}dudx \\ &= -\int_t^a G(x)dx + \alpha(a-t) = -\int\int_t^a \mathbf{1}_{\{y\le x\}}dxdG(y) + \alpha(a-t) \\ &= -\int_{\{y\le t\}}(a-t)dG(y) - \int_{\{t<y\le a\}}(a-y)dG(y) + \alpha(a-t) \\ &= tG(t) - aG(a) + \alpha(a-t) + \int y\mathbf{1}_{\{t<y\le a\}}dG(y).\end{aligned}$$

Combining the two observations, we get

$$T(G) - T_a(G) = \frac{aG(a) - \alpha a}{1-\alpha} + \int y\frac{\mathbf{1}_{\{y>a\}}}{1-\alpha}dG(y).$$

This equality does not change when $G(t-) = \alpha = G(t)$. Since a similar result holds for H, consequently we arrive at

$$\begin{aligned}&T(H) - T(G) - \int \phi_F(x)d(H-G)(x) \\ &= T_a(H) - T_a(G) - \int \psi_a(x)d(H-G)(x) + \frac{a(H(a)-G(a))}{1-\alpha} + \int y\frac{\mathbf{1}_{\{y>a\}}}{1-\alpha}d(H-G)(y) \\ &\quad + \int\int_a^\infty \left\{\mathbf{1}_{\{y\ge x\}} - F(y)\right\} J(F(y))dyd(H-G)(x).\end{aligned}$$

With respect to the last term, we see it is bounded above by

$$\int\int_a^\infty \left|\bar{F}(y) - \mathbf{1}_{\{y<x\}}\right| J(F(y))dyd(H+G)(x) \le \frac{2\mathsf{E}[|L|]}{1-\alpha} + \int \frac{(x-a)^+}{1-\alpha}d(H+G)(x) < \infty.$$

Hence, Fubini's theorem implies that the term can be written as

$$\int_a^\infty \int \left\{\mathbf{1}_{\{y\ge x\}} - F(y)\right\} d(H-G)(x)J(F(y))dy = \int_a^\infty (H(y)-G(y))\, J(F(y))dy.$$

Finally, combining the inequality above and the result of Step 1 leads to the following inequality:

$$\begin{aligned}&\left|T(H) - T(G) - \int \phi_F(x)d(H-G)(x)\right| \\ &\le \frac{\Psi_{G,H}}{1-\alpha} + \frac{a|H(a)-G(a)|}{1-\alpha} + \left|\int y\frac{\mathbf{1}_{\{y>a\}}}{1-\alpha}d(H-G)(y)\right| + \left|\int_a^\infty (H(y)-G(y))\, J(F(y))dy\right|.\end{aligned}$$

Note that the inequality above holds as long as G and H have finite means and for any value a greater than $\max\{v_\alpha, G^{-1}(\alpha), H^{-1}(\alpha)\}$.

Step 3. Suppose that we are given a positive real number ϵ and i.i.d. realizations $\omega := \{L_i\}_{i=1}^{\infty}$. We can think of ω as an element in Ω. Then, Theorem 2.3.1 in Serfling (1980) says $\lim_n F_n^{-1}(\alpha) = v_\alpha$ for almost all realizations ω. In other words, for such an ω we can find an integer $n_\epsilon(\omega)$ such that $n > n_\epsilon(\omega)$ implies $|F_n^{-1}(\alpha) - F^{-1}(\alpha)| < \epsilon(1-\alpha)/6$.

Now let us take G and H as any two empirical distribution functions that depend on such ω, say F_n and F_m with $n, m > n_\epsilon(\omega)$. Since the last inequality of Step 2 holds for any sufficiently large a as long as G and H are fixed, we can take a greater than the largest value among the samples that underlie F_n and F_m. This simplifies the inequality so that we obtain

$$\left| T(F_n) - T(F_m) - \int \phi_F(x) d(F_n - F_m)(x) \right| \le \frac{\Psi_{F_n,F_m}}{1-\alpha} < \epsilon \|F_n - F_m\|_\infty.$$

By definition of $T(\cdot)$, $T(F_n) = \widehat{c}_\alpha^n$ and $T(F_m) = \hat{c}_\alpha^m$. This completes the proof. $\square$

Lemma 4 below is useful in deriving results that follow and it is a detailed description of a claim given on page 1193 of Shao and Wu (1989).

Lemma 4. *For an empirical distribution F_n, the following representation holds:*

$$\widehat{c}_\alpha^n - c_\alpha = \frac{1}{n}\sum_{i=1}^{n} \phi_F(L_i) + R_n$$

where $\phi_F(x)$ is the same as defined in Lemma 3 and

$$R_n = \int_{-\infty}^{\infty} \left\{ \int_0^{F(y)} J(u)du - \int_0^{F_n(y)} J(u)du - J(F(y))(F(y) - F_n(y)) \right\} dy.$$

Proof: For the right hand side, it is clear that

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^{n} \phi_F(L_i) + R_n &= \int (F(y) - F_n(y)) J(F(y)) dy + R_n \\ &= \int \int_0^1 J(u) \Big\{ \mathbf{1}_{\{u \le F(y)\}} - \mathbf{1}_{\{u \le F_n(y)\}} \Big\} du dy. \end{aligned} \tag{3}$$

We observe that the last expression is bounded by

$$\begin{aligned} \int \int_0^1 J(u) \Big\{ \mathbf{1}_{\{F_n(y) < u \le F(y)\}} + \mathbf{1}_{\{F(y) < u \le F_n(y)\}} \Big\} du dy &\le \int |F(y) - F_n(y)|\, dy \\ &= \int \left|\bar{F}(y) - \bar{F}_n(y)\right| dy < \infty. \end{aligned}$$

Hence by Funibi's theorem,

$$\begin{aligned} (3) &= \int_0^1 \int J(u) \Big\{ \mathbf{1}_{\{u \le F(y)\}} - \mathbf{1}_{\{u \le F_n(y)\}} \Big\} dy du \\ &= \int_0^1 J(u) \int \Big\{ \mathbf{1}_{\{F_n(y) < u \le F(y)\}} - \mathbf{1}_{\{F(y) < u \le F_n(y)\}} \Big\} dy du \\ &= \int_0^1 J(u) \int \Big\{ \mathbf{1}_{\{F^{-1}(u) \le y < F_n^{-1}(u)\}} - \mathbf{1}_{\{F_n^{-1}(u) \le y < F^{-1}(u)\}} \Big\} dy du \\ &= \int_0^1 J(u) (F_n^{-1}(u) - F^{-1}(u)) du, \end{aligned}$$

which is equal to $\widehat{c}_\alpha^n - c_\alpha$. $\square$

LEMMA 5. *Suppose that Assumption 1 holds. Then, for an integer $k > 0$, we have*

1. $\left\{n^k \left(\widehat{v}_\alpha^n - v_\alpha\right)^{2k}\right\}$ *is uniformly integrable;*
2. $\lim_n n^k \mathsf{E}\left[(R_n^v)^{2k}\right] = 0$*;*
3. $n^k \mathsf{E}\left[|R_n|^k\right] = \mathcal{O}(1)$

where R_n^v is the nonlinear term of the Bahadur representation of $\widehat{v}_\alpha^n$.

Proof: The proof given here is inspired by Lemma 1 of Shao (1989). We have the usual representation of the quantile estimator

$$\widehat{v}_\alpha^n = v_\alpha + \frac{1}{n}\sum_{i=1}^n Z_i + R_n^v, \quad Z_i = \frac{\alpha - \mathbf{1}_{\{L_i \le v_\alpha\}}}{f(v_\alpha)}.$$

We use different notation to avoid confusion. From $\lim_n n^{3/2}\mathsf{E}\left[(R_n^v)^2\right] = \mathcal{O}(1)$ (Dutt 1973), $\sqrt{n}R_n^v$ converges to zero in probability and thus

$$\sqrt{n}\left(\widehat{v}_\alpha^n - v_\alpha\right) \Rightarrow \mathsf{N}(0,\sigma^2), \quad \sigma_Z^2 = \mathsf{E}\left[Z_i^2\right] = \frac{\alpha(1-\alpha)}{f(v_\alpha)^2}.$$

Under the given assumption, in Chapter 6 of Reiss (1989) it is shown that

$$n^k \mathsf{E}\left[\left(\widehat{v}_\alpha^n - v_\alpha\right)^{2k}\right] = (2k-1)!!\sigma_Z^{2k} + \mathcal{O}\left(n^{-1/2}\right),$$

where $(2k-1)!! = (2k-1)(2k-3)\cdots 1$ is the $2k$th moment of a standard normal distribution. By page 15 of Serfling (1980), we then conclude that $\left\{n^k\left(\widehat{v}_\alpha^n - v_\alpha\right)^{2k}\right\}$ is uniformly integrable.

On the other hand, one can show that

$$\mathsf{E}\left[\left(\frac{1}{n}\sum_{i=1}^n Z_i\right)^{2k}\right] = (2k-1)!!\sigma_Z^{2k} n^{-k} + \mathcal{O}\left(n^{-k-1}\right).$$

Indeed, since $\mathsf{E}[Z_i] = 0$, we can directly expand the left hand-side and observe that

$$\begin{aligned}
\mathsf{E}\left[\left(\sum_{i=1}^n Z_i\right)^{2k}\right] &= \mathsf{E}\left[\sum_{(i_1,\dots,i_k)} \left(Z_{i_1}\right)^2 \cdots \left(Z_{i_k}\right)^2\right] \\
&\quad + \mathsf{E}\left[\sum_{(i_1,\dots,i_{k-1})} \left(Z_{i_1}\right)^3 \left(Z_{i_2}\right)^3 \left(Z_{i_3}\right)^2 \cdots \left(Z_{i_{k-1}}\right)^2\right] + \cdots + \mathsf{E}\left[\sum_i \left(Z_i\right)^{2k}\right] \\
&= \binom{n}{k}\binom{2k}{2}\binom{2k-2}{2}\cdots\binom{2}{2}\sigma_Z^{2k} + \mathcal{O}\left(n^{k-1}\right) = n^k(2k-1)!!\sigma_Z^{2k} + \mathcal{O}\left(n^{k-1}\right),
\end{aligned}$$

where the summations in the first equality are taken over distinct indices.

Finally, notice that

$$n^k (R_n^v)^{2k} = n^k\left\{\left(\widehat{v}_\alpha^n - v_\alpha\right) - \frac{1}{n}\sum_{i=1}^n Z_i\right\}^{2k} \le 2^{2k-1} n^k \left\{\left(\widehat{v}_\alpha^n - v_\alpha\right)^{2k} + \left(\frac{1}{n}\sum_{i=1}^n Z_i\right)^{2k}\right\},$$

and thus $\{n^k (R_n^v)^{2k}\}$ is uniformly integrable. Therefore, $\sqrt{n}R_n^v \to 0$ in probability implies the second statement.

From Lemma 4, we have $R_n = \int\int_{F_n(y)}^{F(y)} \left(J(u) - J(F(y))\right) du dy$. Using the definition of $J(\cdot)$,

$$\begin{aligned}
-(1-\alpha)R_n &= -\int\int_{F_n(y)}^{F(y)} \left(\mathbf{1}_{\{u>\alpha\ge F(y)\}} - \mathbf{1}_{\{F(y)>\alpha\ge u\}}\right) du dy \\
&= \int\int_0^1 \left(\mathbf{1}_{\{F(y)\le\alpha<u<F_n(y)\}} + \mathbf{1}_{\{F_n(y)<u\le\alpha<F(y)\}}\right) du dy \\
&= \int \left\{(F_n(y) - \alpha)\mathbf{1}_{\{F(y)\le\alpha<F_n(y)\}} + (\alpha - F_n(y))\mathbf{1}_{\{F_n(y)<\alpha<F(y)\}}\right\} dy.
\end{aligned}$$

This implies that $(1-\alpha)^k|R_n|^k \le |F_n(v_\alpha)-\alpha|^k|v_\alpha - \widehat{v}_\alpha^n|^k$. The Cauchy-Schwarz inequality leads us to

$$n^k \mathsf{E}\left[|R_n|^k\right] \le \frac{1}{(1-\alpha)^k}\sqrt{\mathsf{E}\left[n^k\left(F_n(v_\alpha)-\alpha\right)^{2k}\right]\mathsf{E}\left[n^k\left(\widehat{v}_\alpha^n - v_\alpha\right)^{2k}\right]}.$$

The first expectation is easily shown to be bounded in n because $F_n(\cdot)$ is the sum of i.i.d. random variables and $\alpha = \mathsf{E}[F_n(v_\alpha)]$ with finite higher moments. The second expectation is also $\mathcal{O}(1)$ due to the previous result. □

Proof of Proposition 3: We first prove the results for $\widehat{\sigma}^2_{\text{jack}}$, and the results corresponding to $\widehat{\sigma}^2_{\text{jack-b}}$ follow from the asymptotic equivalence between the two variance estimators as given in the proof of Proposition 1.

Recall that $\widehat{c}_\alpha^{\text{jack}} = \widehat{c}_\alpha^n = \Phi(\mathbf{L})$ and, by definition, we have

$$\begin{aligned}\widehat{\sigma}^2_{\text{jack}} &= \frac{n_b-1}{n_b}\sum_{j=1}^{n_b}\left(\Phi\left(\widetilde{\mathbf{L}}^{(j)}\right)-\Phi(\mathbf{L})\right)^2 \\ &= \frac{n_b-1}{n_b}\sum_{j=1}^{n_b}\left(\widetilde{\varphi}^{(j)}-\varphi_n+\widetilde{R}^{(j)}-R_n\right)^2,\end{aligned}$$

where $\widetilde{\varphi}^{(j)}$ and $\widetilde{R}^{(j)}$ denote the two terms on the right-hand side from Lemma 4 with respect to the jth jackknifed sample $\widetilde{\mathbf{L}}^{(j)}$. Similarly, φ_n and R_n are the corresponding terms for the entire sample $\mathbf{L}$.

One can prove that

$$\mathsf{E}\left[\frac{n(n_b-1)}{n_b}\sum_{j=1}^{n_b}\left(\widetilde{\varphi}^{(j)}-\varphi_n\right)^2\right]=\sigma^2,$$

by following the same steps as given for proving Proposition 1. The only difference is that the variance $\sigma^2 = \mathsf{Var}(\phi_F(L))$ is now given by

$$\begin{aligned}\sigma^2 &= \mathsf{Var}\left(\int\left\{\mathbf{1}_{\{y\ge L\}}-F(y)\right\}J(F(y))dy\right) \\ &= \frac{1}{(1-\alpha)^2}\mathsf{Var}\left(\int \mathbf{1}_{\{y<L\}}\mathbf{1}_{\{F(y)>\alpha\}}dy\right) \\ &= \frac{1}{(1-\alpha)^2}\mathsf{Var}\left((L-v_\alpha)\mathbf{1}_{\{L\ge v_\alpha\}}\right).\end{aligned}$$

The proof for convergence in probability can also be given in a similar fashion provided that $\mathsf{E}[L^4]<\infty$ and $n_b\to\infty$.

On the other hand, the Cauchy-Schwarz inequality implies that

$$\begin{aligned}n\mathsf{E}\left[\left|\sum_{j=1}^{n_b}\left(\widetilde{\varphi}^{(j)}-\varphi_n\right)\left(\widetilde{R}^{(j)}-R_n\right)\right|\right] &\le n\sum_{j=1}^{n_b}\sqrt{\mathsf{E}\left[\left(\widetilde{\varphi}^{(j)}-\varphi_n\right)^2\right]\mathsf{E}\left[\left(\widetilde{R}^{(j)}-R_n\right)^2\right]} \\ &= n\sum_{j=1}^{n_b}\sqrt{\frac{\sigma^2}{n(n_b-1)}\mathsf{E}\left[\left(\widetilde{R}^{(j)}-R_n\right)^2\right]} \\ &\le n\sqrt{n_b\sum_{j=1}^{n_b}\frac{\sigma^2}{n(n_b-1)}\mathsf{E}\left[\left(\widetilde{R}^{(j)}-R_n\right)^2\right]} \\ &= \sqrt{\frac{n_b\sigma^2}{n_b-1}n\sum_{j=1}^{n_b}\mathsf{E}\left[\left(\widetilde{R}^{(j)}-R_n\right)^2\right]}.\end{aligned}$$

Here, we used the symmetry of $\widetilde{\varphi}_n^{(j)}$ in the first equality and $\left(\sum_{j=1}^{n_b} x_i\right)^2 \le n_b \sum_{j=1}^{n_b} x_i^2$ in the second inequality. Therefore, $\widehat{\sigma}^2_{\text{jack}}$ is consistent and asymptotically unbiased as long as

$$\lim_{n\to\infty} \frac{n(n_b-1)}{n_b} \sum_{j=1}^{n_b} \mathsf{E}\left[\left(\widetilde{R}^{(j)} - R_n\right)^2\right] = 0 \Leftrightarrow \lim_{n\to\infty} n(n_b-1)\mathsf{E}\left[\left(\widetilde{R}^{(1)} - R_n\right)^2\right] = 0.$$

Let us first assume that n_s is bounded. From Lemma 3, we know that for a given $\epsilon > 0$ it holds with probability 1 that

$$|\widetilde{R}^{(j)} - R_n| \le \epsilon \|\widetilde{F}^{(j)} - F_n\|_\infty$$

for all sufficiently large n. Here $\widetilde{F}^{(j)}$ denotes the empirical distribution based on $\widetilde{\mathbf{L}}^{(j)}$. We further observe that $\|\widetilde{F}^{(j)} - F_n\|_\infty \le n_s/n$. Since $\epsilon > 0$ is arbitrary, we conclude that

$$\lim_{n\to\infty} n^2 \max_{j=1,\dots,n_b} \left(\widetilde{R}^{(j)} - R_n\right)^2 = 0$$

almost surely. Therefore, $n\sum_{j=1}^{n_b} \left(\widetilde{R}^{(j)} - R_n\right)^2 \to 0$ almost surely. On the other hand, we note that Lemma 5 gives us $n^2\mathsf{E}\left[|\widetilde{R}^{(j)} R_n|\right] = \mathcal{O}(1)$ as well as $n^2\mathsf{E}\left[R_n^2\right] = \mathcal{O}(1)$. Hence, the dominated convergence theorem guarantees the asymptotic unbiasedness.

Lastly, assume that $n_s \to \infty$. Then, from Lemma 5, we have

$$\frac{n(n_b-1)}{n_b} \sum_{j=1}^{n_b} \mathsf{E}\left[\left(\widetilde{R}^{(j)} - R_n\right)^2\right] = nn_b\mathcal{O}\left((n-n_s)^{-2}\right) + nn_b\mathcal{O}\left(n^{-2}\right) = \mathcal{O}\left(n_s^{-1}\right).$$

Thus it converges to zero in L^1. □

References

H. A. David and H. N. Nagaraja. *Order Statistics.* Wiley, New Jersey, 3rd edition, 2003.

D. L. Duttweiler. The mean-square error of Bahabur's order-statistic approximation. *The Annals of Statistics*, 1:446–453, 1973.

N. Gribkova and R. Helmers. The empirical Edgeworth expansion for a studentized trimmed mean. *Mathematical Methods of Statistics*, 15:61–87, 2006.

H. V. Henderson and S. R. Searle. On deriving the inverse of a sum of matrices. *SIAM Review*, 23:53–60, 1981.

J. Kiefer. On Bahadur's representation of sample quantiles. *Annals of Mathematical Statistics*, 38:1323–1342, 1967.

R.-D. Reiss. *Approximate Distributions of Order Statistics.* Springer, New York, 1989.

R. L. Ressler and P. A. W. Lewis. Variance reducton for quantile estimates in simulations via nonlinear controls. *Communications in Statistics - Simluation and Computation*, 19:1045–1077, 1990.

R. J. Serfling. *Approximation Theorems of Mathematical Statistics.* Wiley, New York, 1980.

J. Shao. The efficiency and consistency of approximations to the jackknife variance estimators. *Journal of the American Statistical Association*, 84:114–119, 1989.

J. Shao and C. F. J. Wu. A general theory for jackknife variance estimation. *The Annals of Statistics*, 17:1176–1197, 1989.