# Supplementary material to "Evaluation of Local Importance of Functional Regression via Interval-Based Masking Approach"

Kiyotaka Sugiyama[1,2] and Hidetoshi Matsui[2]

[1] The Japan Research Institute, Limited, Japan
[2] Shiga University, Japan

## Appendix

This appendix presents the theoretical proof that the Shapley value function estimated by MSSIS converges uniformly to the true Shapley value function. We first state the assumptions and notation, then establish bias and variance bounds through three key lemmas, and finally prove the main convergence result (Theorem 1).

We begin by introducing the basic notation and assumptions used in the theoretical analysis. Let $\varphi : I \to \mathbb{R}$ denote the Shapley value function defined on the interval $I = [0, 1]$. At round $r$, let the interval partition be $\Pi^{(r)} = \{I_k^{(r)}\}_{k=1}^{K_r}$, and define the maximum interval width as $\delta_{\max}^{(r)} := \max_k |I_k^{(r)}|$. For each interval $I_k \in \Pi^{(r)}$, the Shapley value is defined by

$$\phi_k^{(r)} := \int_{I_k^{(r)}} \varphi(u)\, du,$$

and the corresponding piecewise constant approximation of $\varphi(t)$ over $\Pi^{(r)}$ is given by

$$\varphi_{\Pi^{(r)}}(t) := \frac{\phi_k^{(r)}}{|I_k^{(r)}|}, \quad \text{for } t \in I_k^{(r)}.$$

This approximation represents the average value of $\varphi(t)$ within each interval of the partition.

**Lemma 1** *Let $\Pi = \{I_1, \ldots, I_n\}$ be a partition of the interval $I = [0, 1]$, and let $\delta_{\max}$ denote the maximum interval width in the partition, i.e.,*

$$\delta_{\max} := \max_{1 \le k \le n} |I_k|.$$

*For any representative point $t_k \in I_k$, the following inequality holds:*

$$\max_{1 \le k \le n} \left| \frac{\phi_{\nu_\Pi}(I_k)}{|I_k|} - \varphi(t_k) \right| \le \frac{C_\varphi}{2^\alpha(\alpha + 1)} \delta_{\max}^\alpha \quad \xrightarrow[\delta_{\max} \to 0]{} 0.$$

*Proof.* From the definition of the Shapley value, the value assigned to each interval $I_k$ under the finite game $\nu_\Pi$ is given by the permutation average:

$$\phi_{\nu_\Pi}(I_k) = \frac{1}{n!} \sum_{\pi \in \mathcal{P}_n} \left[ \nu_\Pi(P^{(\pi)}(I_k) \cup \{I_k\}) - \nu_\Pi(P^{(\pi)}(I_k)) \right],$$

where $\mathcal{P}_n$ denotes all permutations of the $n$ intervals in partition $\Pi$, and $P_\pi(I_k)$ is the set of intervals preceding $I_k$ in permutation $\pi$.

By the additivity of the set function $\nu_\Pi$, we have

$$\nu_\Pi(S) = \int_{\cup I \in S} \varphi(t)\, dt,$$

and thus the marginal contribution simplifies to

$$\nu_\Pi(P^{(\pi)}(I_k) \cup \{I_k\}) - \nu_\Pi(P^{(\pi)}(I_k)) = \int_{I_k} \varphi(t)\, dt,$$

so that

$$\phi_{\nu_\Pi}(I_k) = \int_{I_k} \varphi(t)\, dt.$$

Let $\delta_k := |I_k|$ denote the width of interval $I_k$, and $t_k$ its center. Applying the change of variable $u = t_k + v$ and the $\alpha$–Hölder continuity of $\varphi$, we obtain

$$|\varphi(u) - \varphi(t_k)| \le C_\varphi |v|^\alpha.$$

Therefore, the difference between the average value of $\varphi$ over $I_k$ and its center value is bounded by

$$\left| \frac{\phi_{\nu_\Pi}(I_k)}{\delta_k} - \varphi(t_k) \right| = \left| \frac{1}{\delta_k} \int_{I_k} \varphi(u)\, du - \varphi(t_k) \right| \le \frac{1}{\delta_k} \int_{I_k} |\varphi(u) - \varphi(t_k)|\, du.$$

Since $I_k$ is centered at $t_k$ and has length $\delta_k$, this integral becomes

$$\left| \frac{\phi_{\nu_\Pi}(I_k)}{\delta_k} - \varphi(t_k) \right| \le \frac{C_\varphi}{\delta_k} \int_{-\delta_k/2}^{\delta_k/2} |v|^\alpha\, dv = \frac{C_\varphi}{2^\alpha(\alpha+1)} \delta_k^\alpha.$$

Applying the global bound $\delta_k \le \delta_{\max}$ for all $k$ completes the proof of the lemma.

Consider the interval partition $\Pi^{(r)} = \{I_k^{(r)}\}_{k=1}^{K_r}$ at round $r$, where $K_r$ denotes the number of intervals and each $I_k^{(r)} \in \Pi^{(r)}$. For each interval $I_k$, let $\sigma_1, \ldots, \sigma_M$ be $M$ independently sampled permutations of the interval indices. Define the marginal contribution for the $m$-th permutation as:

$$Z_{m,k} := \nu_{\Pi^{(r)}} \left( P^{(\sigma_m)}(I_k) \cup \{I_k\} \right) - \nu_{\Pi^{(r)}} \left( P^{(\sigma_m)}(I_k) \right),$$

where $P_{\sigma_m}(I_k)$ is the set of intervals preceding $I_k$ in permutation $\sigma_m$. The Monte Carlo estimate of the Shapley value for interval $I_k$ at round $r$ is given by:

$$\hat{\phi}_k^{(r,M)} := \frac{1}{M} \sum_{m=1}^{M} Z_{m,k}.$$

**Lemma 2** *Assume that for each $k$, the random variables $Z_{m,k}$ are bounded within $[a_k, b_k]$, and let $\Delta_k := b_k - a_k$ denote their range. Define the maximal range at round $r$ as $\Delta_r := \max_k \Delta_k$. Then, the following concentration inequality holds:*

$$\Pr\left[\max_{k \le K_r} \left|\hat{\phi}_k^{(r,M)} - \phi_k^{(r)}\right| > \varepsilon\right] \le 2K_r \exp\left(-\frac{2M\varepsilon^2}{\Delta_r^2}\right), \quad \mathrm{MSE}(\hat{\phi}_k^{(r,M)}) = \mathcal{O}(1/M).$$

*Proof.* By applying Hoeffding's inequality to each interval index $k \in \{1, \dots, K_r\}$ and taking a union bound over all $K_r$ intervals, we obtain the following tail bound for the Monte Carlo estimator:

$$\Pr\left(\max_{1 \le k \le K_r} \left|\hat{\phi}_k^{(r,M)} - \phi_k^{(r)}\right| > \varepsilon\right) \le 2K_r \exp\left(-\frac{2M\varepsilon^2}{(b-a)^2}\right),$$

where $[a, b]$ denotes the known range of each summand $Z_{m,k}$.

This concentration inequality shows that the maximum estimation error across all intervals is controlled by the sample size $M$ with high probability. Moreover, since $\hat{\phi}_k^{(r,M)}$ is the average of $M$ independent variables, its variance—and hence its mean squared error—is bounded by

$$\mathrm{MSE}\left(\hat{\phi}_k^{(r,M)}\right) = \mathbb{E}\left[\left(\hat{\phi}_k^{(r,M)} - \phi_k^{(r)}\right)^2\right] = \mathcal{O}\left(\frac{1}{M}\right).$$

*Remark 1 (Choice of Sample Size).* To ensure that the tail probability above is bounded by $\varepsilon$, it suffices to choose the number of Monte Carlo samples $M$ such that

$$M \ge \frac{\Delta_r^2}{2\varepsilon^2} \log(2K_r),$$

where $\Delta_r$ bounds the range of the summands $Z_{m,k}$ at round $r$. As shown in Lemma 3, $\Delta_r$ decreases exponentially with $r$, implying that the required sample size $M$ also decreases in later rounds—thus reducing computational cost over time.

We analyze the convergence rate of the maximum interval width during the refinement steps of MSSIS. To account for Monte Carlo estimation error, we introduce the following event at round $r$:

$$\mathcal{E}_r := \left\{\max_{k \le K_r} \left|\hat{\phi}_k^{(r)} - \phi_k^{(r)}\right| \le \varepsilon_r\right\}.$$

As long as $\mathcal{E}_r$ holds, the estimation error is sufficiently controlled. Even if a non-optimal interval is selected due to sampling noise, the resulting error in interval width remains bounded by $O(\varepsilon_r)$ (see the inequality below).

**Lemma 3** *Under the event $\mathcal{E}_r$, the maximum interval width satisfies the recursive bound*

$$\delta_{\max}^{(r+1)} \le (0.75 + C^* \varepsilon_r)\, \delta_{\max}^{(r)},$$

*where $C^* = 2/m_r$ and $m_r := \min_{k \le K_r} |\phi_k^{(r)}|/|I_k^{(r)}|$ is assumed to be strictly positive. This captures the contraction of interval width at each step, combining deterministic shrinkage and estimation error.*

*Proof.* Under the event $\mathcal{E}_r$, the estimation error is bounded by $\varepsilon_r$ for all $k \leq K_r$, i.e.,

$$\left|\hat{\phi}_k^{(r)} - \phi_k^{(r)}\right| \leq \varepsilon_r.$$

Let $k^\star$ be the index of the true maximum contributor, and let $\hat{k}$ be the interval selected for refinement. Then,

$$\phi_{k^\star}^{(r)} - \phi_{\hat{k}}^{(r)} \leq 2\varepsilon_r.$$

Using the Hölder condition, each true contribution satisfies

$$\phi_k^{(r)} \geq m_r \cdot |I_k^{(r)}|, \quad m_r := \min_{k \leq K_r} \frac{|\phi_k^{(r)}|}{|I_k^{(r)}|}.$$

Thus, the difference in widths is bounded by

$$\left| |I_{k^\star}^{(r)}| - |I_{\hat{k}}^{(r)}| \right| \leq \frac{2\varepsilon_r}{m_r} = C^*\varepsilon_r.$$

Let $I_{\hat{k},1}^{(r+1)}$ and $I_{\hat{k},2}^{(r+1)}$ be the subintervals after splitting. Each satisfies

$$|I_{\hat{k},i}^{(r+1)}| \leq \frac{1}{2}\left(\delta_{\max}^{(r)} + C^*\varepsilon_r\right).$$

Since merging can increase interval width by at most a factor of 1.5, the next-round maximum width is bounded by

$$\delta_{\max}^{(r+1)} \leq 1.5 \cdot \frac{\delta_{\max}^{(r)} + C^*\varepsilon_r}{2}.$$

This completes the proof of Lemma 3.

If the estimation error is set as $\varepsilon_r := \varepsilon_0(3/4)^{\alpha r}$ for some $\varepsilon_0 > 0$, then the update coefficient becomes $0.75 + \eta$ with arbitrarily small $\eta > 0$, implying that $\delta_{\max}^{(r)}$ converges to zero exponentially.

**Theorem 1** *Given any target error $\varepsilon > 0$, if the number of rounds $R$ and the Monte Carlo sample size $M$ are chosen to satisfy the required conditions, then the Shapley density estimator converges uniformly to the true density function with probability at least $1 - \varepsilon$:*

$$\|\hat{\varphi}_{R,M} - \varphi\|_\infty \leq \varepsilon.$$

*Proof.* We set the Monte Carlo estimation error at round $r$ as

$$\varepsilon_r := \varepsilon_0(3/4)^{\alpha r}, \quad \varepsilon_0 := \frac{\varepsilon}{4C^*},$$

where $C^*$ is the constant defined in Lemma 3. Since the maximum interval width decays as $(3/4)^r \delta_{\max}^{(0)}$, we apply Lemma 1 to ensure the bias remains below $\varepsilon/2$ by choosing

$$R \geq \left\lceil \frac{\log\left(\frac{2^\alpha(\alpha+1)}{C_\varphi} \cdot \frac{\varepsilon}{\delta_{\max}^{(0)\alpha}}\right)}{\alpha \log(4/3)} \right\rceil .$$

To keep the variance-related error probability below $\varepsilon/2$, we choose the number of Monte Carlo samples as

$$M \geq \frac{\Delta_0^2}{2\varepsilon_0^2} \cdot \left(\frac{4}{3}\right)^{2\alpha R} \cdot \log\left(2\sum_{r=1}^{R} K_r\right),$$

where $\Delta_0$ bounds the range of marginal contributions.

By Lemmas 1 and 3, the bias is bounded by $\varepsilon/2$. Applying Lemma 2 with the chosen $M$ ensures that the Monte Carlo error exceeds $\varepsilon/2$ with probability at most $\varepsilon/2$. Using the triangle inequality, we obtain

$$\|\hat{\varphi}_{R,M} - \varphi\|_\infty \leq \varepsilon \quad \text{with probability at least } 1-\varepsilon.$$

Given a target error tolerance $\varepsilon > 0$, the number of refinement rounds $R$ and Monte Carlo samples $M$ can be selected as

$$R = \left\lceil \frac{\log\left(\frac{2^\alpha(\alpha+1)}{C_\varphi} \cdot \frac{\varepsilon}{\delta_{\max}^{(0)\alpha}}\right)}{\alpha \log(4/3)} \right\rceil, \quad M = \frac{\Delta_0^2}{2\varepsilon_0^2} \cdot \left(\frac{4}{3}\right)^{2\alpha R} \cdot \log\left(2\sum_{r=1}^{R} K_r\right).$$

In conclusion, these expressions characterize the sufficient conditions under which MSSIS achieves uniform consistency with respect to the true continuous Shapley density.