ASSS : Adaptive Stratified Sampling for Shapley-like Values

JUNYUAN PANG, The State Key Laboratory of Blockchain and Data Security, Zhejiang University, Shenzhen Loop Area Institute
JIAYAO ZHANG, The State Key Laboratory of Blockchain and Data Security, Zhejiang University
JIAZHENG SONG, The State Key Laboratory of Blockchain and Data Security, Zhejiang University
JINFEI LIU*, Zhejiang University, Hangzhou High-Tech Zone (Binjiang) Institute of Blockchain and Data Security
KUI REN, The State Key Laboratory of Blockchain and Data Security, Zhejiang University

The Shapley-like values, rooted in cooperative game theory, have gained significant recognition in data valuation and query answering. Due to the exponential complexity of their exact computation, many sampling-based approaches have been developed for Shapley-like value estimation. One widely used approach is stratified sampling, where samples are stratified into n strata and n is the number of players. However, while stratified sampling offers variance reduction, its unbiasedness is contingent upon full stratum coverage, necessitating an onerous $\Omega(n^2 \log n)$ samples. This dependency renders static n-stratified sampling vulnerable under limited budgets, resulting in biased or unstable estimates. To tackle the problem, we introduce novel adaptive stratification strategies for Shapley-like values to retain unbiasedness for stratified sampling with fewer samples. Based on the intuition "the fewer samples, the fewer strata", we develop ASSS (**A**daptive **S**tratified **S**ampling for **S**hapley-like values). Two types of approaches ASSS-A (stratify **A**fter sample) and ASSS-B (stratify **B**efore sample) are proposed to offer flexible trade-offs between sample efficiency and memory usage. Experimental results on real and synthetic datasets exhibit the effectiveness and efficiency of ASSS-A and ASSS-B.

CCS Concepts: • **Information systems → Data management systems**.

Additional Key Words and Phrases: Shapley-like value; Stratifeid sampling

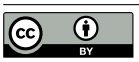
ACM Reference Format:
Junyuan Pang, Jiayao Zhang, Jiazheng Song, Jinfei Liu, and Kui Ren. 2026. ASSS : Adaptive Stratified Sampling for Shapley-like Values. *Proc. ACM Manag. Data* 4, 3 (SIGMOD), Article 139 (June 2026), 28 pages. https://doi.org/10.1145/3802016

1 Introduction

The Shapley value [55], originated from cooperative game theory, offers a moral framework for fairly attributing the collective payoff of a coalition to individual players based on their contributions. In essence, the Shapley value of a player is the average marginal contribution they make

*Corresponding author.

Authors' Contact Information: Junyuan Pang, The State Key Laboratory of Blockchain and Data Security, Zhejiang University, Shenzhen Loop Area Institute, junyuanpang@zju.edu.cn; Jiayao Zhang, The State Key Laboratory of Blockchain and Data Security, Zhejiang University, jiayaozhang@zju.edu.cn; Jiazheng Song, The State Key Laboratory of Blockchain and Data Security, Zhejiang University, jiazhengsong@zju.edu.cn; Jinfei Liu, Zhejiang University, Hangzhou High-Tech Zone (Binjiang) Institute of Blockchain and Data Security, jinfeiliu@zju.edu.cn; Kui Ren, The State Key Laboratory of Blockchain and Data Security, Zhejiang University, kuiren@zju.edu.cn.



ACM 2836-6573/2026/6-ART139
https://doi.org/10.1145/3802016

across all permutations. Their capacity to fairly allocate value or contribution among participants in cooperative scenarios makes them indispensable in numerous domains. In data management [15, 54, 60, 62, 67], the Shapley value has gained prominence for its ability to evaluate the importance of individual data points of a machine learning model or query answering. Particularly in data valuation [4, 10, 18, 19, 21, 22, 41, 42, 45–47, 53, 62], researchers leverage the fairness and interpretability of the Shapley value, making it a valuable tool for addressing challenges in modern data-driven environments.

In recent years, various derivatives of the Shapley value, such as the Banzhaf value [59] and the Beta Shapley value [32], have been introduced for improved data valuation. These values assess player contributions in cooperative game scenarios based on their marginal contributions in different ways of Shapley values and have found applications in data valuation due to their unique properties. We collectively refer to these values as "Shapley-like values."

Calculating the exact Shapley-like values involves evaluating the marginal contribution of each player across all possible coalitions, which grows exponentially with the number of players. Despite the well-acknowledged properties of Shapley-like values, this significant complexity limits their applications on large-scale datasets. To address the challenge, many approximation methods for the Shapley value and its derivatives have been developed, among which Monte Carlo sampling-based approaches have been extensively studied [9, 30, 36, 50, 64]. To reduce sample variance, stratified sampling techniques are introduced in [64]. Specifically, stratified sampling divides the coalitions into n strata based on the coalition size. Samples are drawn from these strata individually and the estimator is computed as the weighted average of the means in each stratum. Stratified sampling leads to more reliable results than simple random sampling owing to the lower stratum variance.

However, conventional stratified sampling is fundamentally limited by the onerous sample requirement for unbiasedness. Specifically, theoretical unbiasedness requires full stratum coverage, necessitating $\Omega(n^2 \log n)$ utility samples. Under limited budgets, the inevitable presence of empty strata compromises the unbiasedness guarantee. Moreover, static stratification fails to adapt to highly non-uniform weight distributions (e.g., the Banzhaf value), where variance-optimal sample allocation needs equally onerous pilot sampling for stratum-variance calculation.

Can we design a stratification strategy for Shapley-like value estimation that provides unbiased yet effective estimators with insufficient samples? The key challenge in existing methods is that the excessive number of strata calls for a large sample size to fill. A compelling idea is to reduce the number of strata, enabling faster filling and thus requiring fewer samples for unbiased estimation. Furthermore, significant variance reduction in stratified sampling can be achieved without relying on many strata. We illustrate the case with an example as follows.

Example 1.1. For exposition, assume X is uniform on $[0, 2]$ (i.e., $X \sim U[0, 2]$). In unstratified sampling, we estimate the expectation $\mathbb{E}[X]$ by sampling m independent values $X_1, X_2, \ldots, X_m$ and computing the sample mean $\hat{\mu}_{\text{unstrat}} = \frac{1}{m}\sum_{i=1}^{m} X_i$. The variance of $X \sim U[0, 2]$ $\text{Var}(X) = \frac{(2-0)^2}{12} = \frac{1}{3}$. Thus, the variance of the sample mean $\text{Var}(\hat{\mu}_{\text{unstrat}}) = \frac{1}{m}\text{Var}(X) = \frac{1}{3m}$.

In stratified sampling, we divide the interval $[0, 2]$ into two subintervals $[0, 1]$ and $[1, 2]$. We sample $m/2$ values from each subinterval as shown in Figure 1. Let $X_1^{(1)}, X_2^{(1)}, \ldots, X_{m/2}^{(1)}$ be the samples from $U[0, 1]$, and $X_1^{(2)}, X_2^{(2)}, \ldots, X_{m/2}^{(2)}$ be the samples from $U[1, 2]$. The stratified estimator for the expectation is the average of the sample means from each subinterval $\hat{\mu}_{\text{strat}} = \frac{1}{2}(\hat{\mu}_1 + \hat{\mu}_2)$, where $\hat{\mu}_1 = \frac{2}{m}\sum_{i=1}^{m/2} X_i^{(1)}$ and $\hat{\mu}_2 = \frac{2}{m}\sum_{i=1}^{m/2} X_i^{(2)}$. The variance of X in each subinterval $\text{Var}(X^{(1)}) = \text{Var}(X^{(2)}) = \frac{(1-0)^2}{12} = \frac{1}{12}$. Thus, the variance of the sample mean from each subinterval $\text{Var}(\hat{\mu}_1) = \text{Var}(\hat{\mu}_2) = \frac{1}{6m}$. Since the subintervals are independent, the variance of the stratified estimator

$\mathrm{Var}(\hat{\mu}_{\mathrm{strat}}) = \frac{1}{4}\left(\frac{1}{6m} + \frac{1}{6m}\right) = \frac{1}{12m}$. With only two strata, the variance is reduced by a factor of four. In addition, only two samples are required to ensure unbiased estimation in this stratification.

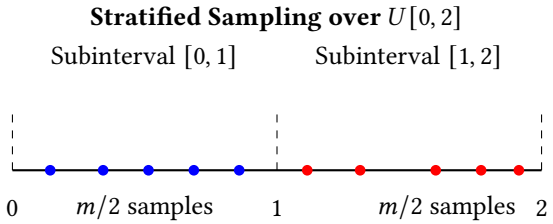


Fig. 1. Stratified sampling with two strata in $[0,2]$.

Table 1. Comparison of sampling methods. *Required samples* refers to the minimum number of samples required to cover all strata. *Lower variance* indicates whether the method can reduce variance compared to unstratified sampling. The unbiasedness of ASSS-B holds with high probability, as detailed in Theorem 5.10.

	unstratified sampling	stratified sampling	ASSS (type A)	ASSS (type B)
number of strata	1	n	$\leq n$ (adaptive)	L (adaptive)
required samples	$\Omega(1)$	$\Omega(n^2 \log n)$	$\Omega(1)$	$\Omega(L^2 \log L)$
space cost	$O(n)$	$O(n^2)$	$O(n^2)$	$O(Ln)$
lower variance	✗	✓	✓	✓
unbiased guarantee	✓	✗	✓	✓

We can see that a relatively small number of strata can still effectively reduce the estimator variance while requiring much fewer samples to provide unbiased estimators. Based on the simple yet effective intuition, we propose ASSS (**A**daptive **S**tratified **S**ampling for **S**hapley-like values), an innovative sampling approach that flexibly adjusts the stratification structure and properly sets probability distribution for sampling according to the number of samples and sampling results. We introduce two kinds of approaches ASSS-A (stratify **A**fter sample) and ASSS-B (stratify **B**efore sample). The advantages of our methods are shown in Table 1. Our approaches maintain the variance reduction benefits of stratified sampling while providing unbiased estimators with fewer samples and lower space costs. The superiority of our approaches is confirmed through both mathematical analysis in Section 5 and experimental validation in Section 6. Our main contributions are summarized as follows.

- We propose **ASSS (Adaptive Stratified Sampling for Shapley-like values)**, a novel framework that fundamentally addresses the limitation of conventional static stratified sampling. By adaptively decoupling the stratification structure from the number of players n, ASSS enables stratified sampling that retains unbiasedness and achieves superior efficiency compared to existing methods in limited budget scenarios. We introduced two variations, ASSS-A and ASSS-B, to address different computational constraints, offering flexibility in trade-offs between sample efficiency and memory usage.
- We conduct **mathematical analysis** to establish the theoretical superiority of the proposed methods. We identify the "empty stratum" bottleneck in existing methods, and prove that ASSS secures unbiasedness and establishes strictly lower variance bounds compared to unstratified sampling, effectively bridging the gap between estimation stability and budget limitations.
- We design **efficient algorithms** for both ASSS-A and ASSS-B to offer a flexible trade-off between sample efficiency and memory usage.

Table 2. A summary of frequently used notations.

Notation	Definition
$\mathcal{N}$	the set of players
n	the number of players
i	the i-th player
m	the number of samples
$v(\cdot)$	utility function
$\mathcal{S}$	a coalition of players
ϕ_i	the Shapley-like value of player i
w_k	the weight of the k-th stratum
$\mathcal{L}$	a stratification
$\mathcal{L}_k$	the k-th stratum of $\mathcal{L}$
ω_k	the weight of the stratum $\mathcal{L}_k$

- Extensive **experimental validation** is performed on both real-world and synthetic datasets, verifying the effectiveness, efficiency, scalability, and robustness of the proposed methods. Additionally, we provide practical guidelines to assist practitioners in selecting the appropriate ASSS variant based on specific application constraints.

The remainder of this paper is organized as follows. Section 2 provides a brief review of existing research on the Shapley-like values. Section 3 introduces the concept of Shapley-like values and a state-of-the-art method for Shapley-like value estimation. Section 4 presents the application of existing stratified sampling techniques to the introduced state-of-the-art method and discusses its limitations. In Section 5, we propose an adaptive stratified sampling technique, introducing two stratification strategies along with their corresponding algorithms. Section 6 presents experimental results and key findings. Finally, Section 7 concludes the paper by summarizing the main contributions and insights. Frequently used notations are summarized in Table 2, and all proofs are provided in the appendix to maintain brevity in the main body of the paper.

2 Related Work

In this section, we introduce the applications and computation of Shapley-like values, respectively.

Applications. Shapley value [55], rooted in cooperative game theory, has been widely adopted for its fairness properties in evaluating the contribution of each data point. It has been applied to numerous downstream tasks, such as explanations in database queries [13, 28, 29, 43, 54], data pricing [2, 4, 10, 19, 41, 42, 45–47], data/feature selection [18, 21, 37, 38, 53, 60, 62], database tuning [22, 34, 66], contribution evaluation in federated learning [3, 6, 7, 11, 16, 17, 48, 56–58, 63, 67], influence maximization in social networks [68], and model interpretation [23, 35, 44, 51, 53]. The Banzhaf value [5], a variation of the Shapley value with robustness to varying coalition structures, has also gained traction in data valuation [59] and query answering [1, 31]. Further extending these ideas, the Beta Shapley value [32] introduces adjustable weighting to balance different coalition sizes, enabling tailored data valuation for specific tasks and enhancing flexibility in data valuation [27, 33]. Reference [29] explored Shapley-like values in probabilistic settings, extending their applicability to probabilistic databases.

Computation. Although Shapley-like values have widespread applications, their computation requires evaluating the marginal contributions of exponentially many coalitions, which is #P-hard [14]. Numerous approximation techniques have been developed to address this challenge [8, 12, 21, 24, 26, 30, 36, 44, 49, 50, 64, 65]. Reference [49] pioneered the use of Monte Carlo sampling for Shapley value estimation, followed by methods using permutation sampling to estimate Shapley values as expectations of marginal contributions [9, 50]. These approaches have been further enhanced by Quasi-Monte Carlo techniques [9] and stratified sampling algorithms [24], as refined in [8]. Reference [44] proposed computing the Shapley value through a weighted optimization based on sampled utilities, later improved in [12] using paired sampling. Reference [64] significantly improved the sampling efficiency of Shapley value estimation by introducing complementary contributions instead of marginal contributions, which enhanced the reusability of samples.

Reference [26] designed algorithms with polynomial complexity specifically for KNN classifiers given the computational difficulty of calculating Shapley-like values in general cases. Reference [21] improved efficiency by omitting marginal contributions below a certain threshold or approximating utilities via gradients within permutation sampling, albeit at the cost of unbiasedness. Reference [65] investigated efficient approximation methods for updating Shapley values in dynamic datasets, addressing challenges posed by data changes over time.

The most relevant works to our study are [36] and [64]. Reference [36] introduced an unstratified utility-sampling approach to estimate Shapley-like values, while [64] showed that static stratification can reduce variance when strata are sufficiently covered. However, a key challenge for existing stratified estimators arises when the sample budget is limited——the "empty stratum" problem. Static stratification schemes often fail to cover all strata, which breaks the unbiasedness guarantee and leads to biased or undefined estimates. This challenge is further amplified for Shapley-like values whose coalition-size weights $p(k)$ can be highly non-uniform, making a static stratification design poorly matched across different targets. Moreover, variance-optimal allocation rules typically rely on pilot sampling to estimate within-stratum variances, which is also onerous under limited budget scenarios. To address these limitations, this paper proposes an adaptive stratified sampling approach (ASSS) that adaptively decouples the stratification structure from the number of players n.

3 Preliminaries

In this section, we introduce the definitions of Shapley-like values in Section 3.1 and the sampling algorithms for approximating Shapley-like values in Section 3.2.

3.1 Shapley-like Values

The Shapley-like value is a class of values that evaluates the contribution of a player by calculating the expected marginal contribution of the player to all possible coalitions. To further understand the concept, we first explore the fundamental idea of the Shapley value.

3.1.1 Shapley Value. The Shapley value [55] is a concept in cooperative game theory, named in honor of Nobel laureate Lloyd Shapley, who introduced it in 1953. It provides a fair division of payoffs among cooperative players based on their contributions to various coalitions.

Consider a set of players $\mathcal{N} = \{1, 2, \ldots, n\}$ in a cooperative game. A coalition is a subset $\mathcal{S}$ within $\mathcal{N}$. $\mathcal{N}$ itself is called the **grand coalition**. The utility function $v : 2^{\mathcal{N}} \to \mathbb{R}$ assigns a value to evaluate the utility of each coalition $\mathcal{S} \subseteq \mathcal{N}$. The Shapley value $\phi_i(v)$ for player i is defined as

$$\phi_i(v) = \frac{1}{n!} \sum_{\pi \in \Pi(\mathcal{N})} \left[v(\mathcal{S}_i^\pi \cup \{i\}) - v(\mathcal{S}_i^\pi) \right] \tag{1}$$

$$= \frac{1}{n} \sum_{\mathcal{S} \subseteq \mathcal{N} \setminus \{i\}} \frac{1}{\binom{n-1}{|\mathcal{S}|}} \left(v(\mathcal{S} \cup \{i\}) - v(\mathcal{S}) \right) \tag{2}$$

$$= \mathrm{E}[v(\mathcal{S} \cup \{i\}) - v(\mathcal{S}) \mid \mathcal{S} \subseteq \mathcal{N} \setminus \{i\}, \Pr(\mathcal{S}) = \frac{p_k^{\text{Shap}}}{\binom{n-1}{k-1}}], \tag{3}$$

where $\mathcal{S}_i^\pi$ represents the coalition containing all the players before player i in permutation π, $k = |\mathcal{S}| + 1$, and $p_k^{\text{Shap}} = \frac{1}{n} (1 \le k \le n)$. $v(\mathcal{S}_i^\pi \cup \{i\}) - v(\mathcal{S}_i^\pi)$ represents the **marginal contribution** of player i in permutation π, and $v(\mathcal{S} \cup \{i\}) - v(\mathcal{S})$ represents the marginal contribution of player i to coalition $\mathcal{S}$. Equation 1 expresses the Shapley value as the average marginal contribution of each player across all permutations; Equation 2 expresses the Shapley value as the weighted average marginal contributions of each player to all possible coalitions; Equation 3 expresses the Shapley value as the expectation of the marginal contributions of each player to all possible coalitions, where coalition $\mathcal{S}$ is drawn from $\mathcal{N} \setminus \{i\}$ with probability $p_{|\mathcal{S}|+1}^{\text{Shap}}$. The probability $p_{|\mathcal{S}|+1}^{\text{Shap}}$ corresponds to the weight of the marginal contribution of each player to a coalition with the same size of $\mathcal{S}$ in the calculation of its Shapley value.

The Shapley value is the unique measure that satisfies four significant properties of fairness.

- **Efficiency**: The sum of the Shapley values equals the utility of the grand coalition, i.e., $\sum_{i=1}^n \phi_i(v) = v(\mathcal{N}) - v(\varnothing)$.
- **Symmetry**: If the marginal contributions of two players are equal to all coalitions, they should receive the same Shapley value.
- **Dummy Player**: If a player does not contribute to any coalition (i.e., zero marginal contribution), its Shapley value should be zero.
- **Additivity**: For two utility functions v and w, $\phi_i(v + w) = \phi_i(v) + \phi_i(w)$.

The fairness properties and the flexibility of the utility function in the applications of the Shapley value have led to its widespread implementation in data valuation [21] and the interpretability of query answering [15]. Below we employ an example to show the Shapley value computation based on Equation 1.

Example 3.1. Consider a game with three players $\mathcal{N} = \{1, 2, 3\}$ and the utility function v defined as follows.

$$v(\varnothing) = 0, \quad v(\{1\}) = 1, \quad v(\{2\}) = 2, \quad v(\{3\}) = 3,$$
$$v(\{1,2\}) = 4, \quad v(\{1,3\}) = 5, \quad v(\{2,3\}) = 6, \quad v(\{1,2,3\}) = 11.$$

The Shapley value computation of each player is presented in Table 3. Equation $(\mathbf{1}, 2, 3) : 1 - 0 = 1$ represents that the marginal contribution of player 1 in permutation $(1, 2, 3)$ is $v(\{1\}) - v(\varnothing) = 1 - 0 = 1$. For player 1, the marginal contributions across all permutations are 1, 1, 2, 2, 5, and 5, respectively. Shapley value ϕ_1 is calculated as $\frac{1}{6}(1 + 1 + 2 + 2 + 5 + 5) = \frac{8}{3}$.

3.1.2 Shapley-derived Values. Capitalizing on the proven benefits of the Shapley value, various Shapley-derived values such as the Banzhaf value [59] and the Beta Shapley value [32] have been employed to improve data valuation for their advanced robustness and flexibility.

Banzhaf value. The Banzhaf value [5] is also a concept in cooperative game theory, named in honor of John F. Banzhaf III, who introduced it in 1964. The Banzhaf value treats coalitions of different sizes equally and measures the average marginal contribution of each player to all possible

Player	Marginal Contribution	$\phi_i(v)$
1	$(\mathbf{1},2,3):1-0=1$ $(\mathbf{1},3,2):1-0=1$ $(2,\mathbf{1},3):4-2=2$ $(3,\mathbf{1},2):5-3=2$ $(2,3,\mathbf{1}):11-6=5$ $(3,2,\mathbf{1}):11-6=5$ average: $\frac{1}{6}(1+1+2+2+5+5)=\frac{8}{3}$	$\frac{8}{3}$
2	$(\mathbf{2},1,3):2-0=2$ $(\mathbf{2},3,1):2-0=2$ $(1,\mathbf{2},3):4-1=3$ $(3,\mathbf{2},1):6-3=3$ $(3,1,\mathbf{2}):11-5=6$ $(1,3,\mathbf{2}):11-5=6$ average: $\frac{1}{6}(2+2+3+3+6+6)=\frac{11}{3}$	$\frac{11}{3}$
3	$(\mathbf{3},2,1):3-0=3$ $(\mathbf{3},1,2):3-0=3$ $(1,\mathbf{3},2):5-1=4$ $(2,\mathbf{3},1):6-2=4$ $(2,1,\mathbf{3}):11-4=7$ $(1,2,\mathbf{3}):11-4=7$ average: $\frac{1}{6}(3+3+4+4+6+6)=4$	$\frac{14}{3}$
sum	$\frac{8}{3}+\frac{11}{3}+\frac{14}{3}=11$	11

Table 3. The computation of the Shapley value.

coalitions. Consider a set of players $\mathcal{N}=\{1,2,\ldots,n\}$ and a utility function $v:2^{\mathcal{N}}\to\mathbb{R}$. The Banzhaf value $\varphi_i(v)$ for player i is defined as

$$\begin{aligned}\varphi_i(v) &= \frac{1}{2^{n-1}}\sum_{\mathcal{S}\subseteq\mathcal{N}\setminus\{i\}}\left(v(\mathcal{S}\cup\{i\})-v(\mathcal{S})\right)\\ &= \mathrm{E}[v(\mathcal{S}\cup\{i\})-v(\mathcal{S})\mid\mathcal{S}\subseteq\mathcal{N}\setminus\{i\},\Pr(\mathcal{S})=\frac{p_k^{\mathrm{Banz}}}{\binom{n-1}{k-1}}],\end{aligned}$$

where $k=|\mathcal{S}|+1$ and $p_k^{\mathrm{Banz}}=\frac{1}{2^{n-1}}(1\le k\le n)$. The Banzhaf value computes the expectation of the marginal contribution of each player with each possible coalition the same probability to be drawn, unlike the Shapley value.

Beta Shapley value. The Beta Shapley value [32] is a generalization of the Shapley value designed to enhance data valuation in machine learning. Unlike the original Shapley value, which gives uniform weights to marginal contributions across different coalition sizes, the Beta Shapley value introduces a weighting function to adjust the importance of contributions from different coalition sizes.

Consider a set of players $\mathcal{N}=\{1,2,\ldots,n\}$ and a utility function $v:2^{\mathcal{N}}\to\mathbb{R}$. The $\mathrm{Beta}(a,b)$-Shapley value $\psi_i(v)$ for player i is defined as

$$\begin{aligned}\psi_i(v) &= \sum_{\mathcal{S}\subseteq\mathcal{N}\setminus\{i\}}\frac{\mathrm{Beta}(b+|\mathcal{S}|,n-|\mathcal{S}|-1+a)}{\mathrm{Beta}(a,b)\binom{n-1}{|\mathcal{S}|}}\left(v(\mathcal{S}\cup\{i\})-v(\mathcal{S})\right),\\ &= \mathrm{E}[v(\mathcal{S}\cup\{i\})-v(\mathcal{S})\mid\mathcal{S}\subseteq\mathcal{N}\setminus\{i\},\Pr(\mathcal{S})=\frac{p_k^{\mathrm{Beta}(a,b)}}{\binom{n-1}{k-1}}],\end{aligned}$$

where $k=|\mathcal{S}|+1$ and $p_k^{\mathrm{Beta}(a,b)}=\frac{\mathrm{Beta}(b+k-1,n-k+a)}{\mathrm{Beta}(a,b)\binom{n-1}{k-1}}(1\le k\le n)$.

The Banzhaf value and the Beta Shapley value share the common properties with the Shapley value including symmetry, the dummy player, and additivity, but do not satisfy the efficiency

property since they are no longer the average marginal contribution of each player across all permutations. The Banzhaf value is used in data valuation owing to its robustness, while the Beta Shapley value can flexibly adjust weights for different tasks in data valuation.

We collectively refer to the Shapley value, Banzhaf value, and Beta Shapley value as Shapley-like values, and represent them uniformly as

$$\phi_i(v, \mathbf{p}) = \mathrm{E}\left[v(\mathcal{S} \cup \{i\}) - v(\mathcal{S}) \mid \mathcal{S} \subseteq \mathcal{N} \setminus \{i\}, \Pr(\mathcal{S}) = \frac{p_k}{\binom{n-1}{k-1}}\right],$$

where $k = |\mathcal{S}| + 1$ and $\mathbf{p} \in \mathbb{R}^n$ can be $\mathbf{p}^{\mathrm{Shap}}$, $\mathbf{p}^{\mathrm{Banz}}$, $\mathbf{p}^{\mathrm{Beta}(a,b)}$ or any other vectors satisfying that $\sum_{k=1}^{n} p_k = 1$. For brevity, we denote Shapley-like values with fixed parameters as ϕ_i.

3.2 The Monte Carlo Method

The exact computation of Shapley-like values requires evaluating the utilities of 2^n coalitions. The exponential computational complexity makes their use impractical in large-scale scenarios. As a result, researchers turn to the study of Shapley-like value estimation. Efficient sampling-based methods for estimating Shapley-like values have been extensively studied [9, 36, 64].

Reference [36] proposed GELS to enhance the approximation efficiency by adding a constant to the Shapley-like value of each player, enabling the reuse of each utility sample in Shapley-like value estimation of all players. The reformulated Shapley-like value is given by

$$\begin{aligned}\phi_i(v, \mathbf{p}) = {} & c(n, \mathbf{p})\mathrm{E}\left[v(\mathcal{S}) \mid i \in \mathcal{S} \subsetneq \mathcal{N} \cup \{0\}, \Pr(\mathcal{S}) = p_{|\mathcal{S}|}\right] \\ & - c(n, \mathbf{p})\mathrm{E}\left[v(\mathcal{S}) \mid 0 \in \mathcal{S} \subsetneq \mathcal{N} \cup \{0\}, \Pr(\mathcal{S}) = p_{|\mathcal{S}|}\right],\end{aligned} \tag{4}$$

where player 0 is an additional dummy player and $c(n, \mathbf{p}) = \sum_{k=1}^{n} \frac{n}{n-k+1} p_k$. Given that the second term in Equation 4 remains constant across each player, it suffices only to estimate the first term in Equation 4. Denoted by $\phi_i^* = \mathrm{E}[v(\mathcal{S}) \mid i \in \mathcal{S} \subsetneq \mathcal{N} \cup \{0\}, \Pr(\mathcal{S}) = p_{|\mathcal{S}|}]$ $(0 \le i \le n)$. The Shapley-like values can then be derived by adjusting these estimates based on the value of the dummy player, i.e., $\phi_i = c(n, \mathbf{p})(\phi_i^* - \phi_0^*)$. Using the Monte Carlo method, we randomly sample some coalitions and compute the average utility of the coalitions containing player i as the estimator $\hat{\phi}_i^*$ of ϕ_i^*. The detailed algorithm is presented in Algorithm 1. We first sample coalition $\mathcal{S}$ from $\mathcal{N} \cup \{0\}$ and calculate utility $v(\mathcal{S} \cap \mathcal{N})$ (Lines 3-6). $v(\mathcal{S} \cap \mathcal{N})$ is then reused in the estimation of ϕ_i^* for all player $i \in \mathcal{S}$, which improves efficiency (Lines 7-9). After m samples, we average sampled utilities and derive the estimated Shapley-like values (Lines 10-11).

4 STATIC STRATIFIED SAMPLING

Building on the Monte Carlo method for Shapley-like value estimation, this section investigates the potential of stratified sampling to enhance estimation accuracy. Stratified sampling is particularly appealing due to its ability to reduce variance by dividing the sample space into distinct strata. However, this section also critically evaluates the limitations of the stratification strategy, especially when the number of strata is large or samples are insufficient to fill all strata effectively.

Stratified sampling is a widely used technique for variance reduction. Following existing work on static stratified sampling for estimating Shapley values [9, 30, 36, 50, 64], we first consider a straightforward extension of Algorithm 1 to the Shapley-like value setting. Specifically, we partition all coalitions in $\mathcal{N} \cup \{0\}$ into n strata, where the k-th stratum contains all coalitions of size k.

Let $\phi_{i,k}^* = \mathrm{E}\left[v(\mathcal{S}) \mid i \in \mathcal{S} \subsetneq \mathcal{N} \cup \{0\}, |\mathcal{S}| = k\right]$ denote the expected utility of coalitions in the k-th stratum $(1 \le k \le n)$. The expectation of utilities among all coalitions can then be expressed as

Algorithm 1: Monte Carlo method for Shapley-like value estimation.

input : $\mathcal{N} = \{1, \ldots, n\}$, $v : 2^{\mathcal{N}} \to \mathbb{R}$, $\mathbf{p} \in \mathbb{R}^n$, and $m > 0$
output: the estimated Shapley-like values $\hat{\phi}_1, \ldots, \hat{\phi}_n$

1 $\hat{\phi}_i^*, m_i \leftarrow 0$ $(0 \le i \le n)$;
2 **for** $t = 1$ *to* m **do**
3 Select k with probability $\propto \frac{n(n+1)}{k(n+1-k)} p_k$ $(1 \le k \le n)$;
4 Let π^t be a random permutation of $\{0, 1, \ldots, n\}$;
5 $\mathcal{S} \leftarrow \{\pi^t(1), \ldots, \pi^t(k)\}$;
6 $u \leftarrow v(\mathcal{S} \cap \mathcal{N})$;
7 **for** $i = 1$ *to* k **do**
8 $\hat{\phi}^*_{\pi^t(i)} = \frac{m_{\pi^t(i)}}{m_{\pi^t(i)}+1} \hat{\phi}^*_{\pi^t(i)} + \frac{1}{m_{\pi^t(i)}+1} u$;
9 $m_{\pi^t(i)} += 1$;
10 **for** $i = 1$ *to* n **do**
11 $\hat{\phi}_i \leftarrow \left(\sum_{k=1}^{n} \frac{n}{n-k+1} p_k\right) \left(\hat{\phi}_i^* - \hat{\phi}_0^*\right)$;
12 **return** $\hat{\phi}_1, \ldots, \hat{\phi}_n$.

a weighted average of the expectations of utilities across different strata:

$$\phi_i^*(v, \mathbf{p}) = \sum_{k=1}^{n} w_k \phi_{i,k}^*, \tag{5}$$

where $w_k = \frac{n}{n-k+1} p_k / c(n, \mathbf{p})$.

To estimate ϕ_i^*, we employ post-stratification [25] based on the above equation. Using the Monte Carlo method, we randomly draw samples from each stratum to form coalitions. For ϕ_i^*, we first compute the average utility of coalitions in each stratum containing player i, yielding the estimator $\hat{\phi}_{i,k}^*$ for $\phi_{i,k}^*$ $(1 \le k \le n)$. Subsequently, a weighted average of these estimates across all strata is computed to obtain the estimator $\hat{\phi}_i^*$.

The detailed algorithm is outlined in Algorithm 2. Initially, coalitions $\mathcal{S}$ are sampled from $\mathcal{N} \cup \{0\}$, and their utilities $v(\mathcal{S} \cap \mathcal{N})$ are calculated (Lines 3-6). These utilities are then reused to estimate $\phi_{i,k}^*$ for all players $i \in \mathcal{S}$ in the k-th stratum (Lines 7-9). After m samples are drawn, we compute the average utility in each stratum and derive the estimated Shapley-like values (Lines 10-11).

The following example illustrates the difference between stratified and unstratified sampling.

Example 4.1. We adopt the same setting as in Example 3.1. Table 4 and Figure 2 show the process of unstratified sampling (Algorithm 1) and stratified sampling (Algorithm 2) for Shapley value estimation. Coalitions $\{0\}$, $\{1\}$, $\{0, 1\}$, $\{1, 2\}$, $\{1, 3\}$, $\{0, 1, 3\}$, and $\{0, 2, 3\}$ are sampled. In unstratified sampling as shown in the left side, $\hat{\phi}_0^*$ is calculated as the average of the utilities of four coalitions containing player 0 ($v\{0\} = 0$, $v\{0, 1\} = 1$, $v\{0, 3\} = 3$, and $v\{0, 2, 3\} = 6$). In stratified sampling as shown in the right side, we first calculate the average utilities of the coalitions with same size k as the estimator $\hat{\phi}_{0,k}^*$ of $\phi_{0,k}^*(1 \le k \le 3)$, i.e., $\hat{\phi}_{0,1}^* = 0$, $\hat{\phi}_{0,2}^* = 1$, and $\hat{\phi}_{0,3}^* = \frac{1}{2}(5 + 6) = \frac{11}{2}$. $\hat{\phi}_0^*$ is calculated as the weighted average of the average utilities in three stratum, i.e., $\hat{\phi}_0^* = \frac{2}{11}\hat{\phi}_{0,1}^* + \frac{3}{11}\hat{\phi}_{0,2}^* + \frac{6}{11}\hat{\phi}_{0,3}^*$.

Limitations. The utilities of coalitions of the same size tend to be more homogeneous than those of different sizes. By dividing the coalitions into n strata based on coalition sizes, stratified sampling

Algorithm 2: Stratified Monte Carlo method for Shapley-like value estimation.

input : $\mathcal{N} = \{1, \ldots, n\}$, $v : 2^{\mathcal{N}} \to \mathbb{R}$, $\mathbf{p} \in \mathbb{R}^n$, and $m > 0$
output: the estimated Shapley-like values $\hat{\phi}_1, \ldots, \hat{\phi}_n$

$\hat{\phi}^*_{i,k}, m_{i,k} \leftarrow 0$ $(0 \le i \le n, 1 \le k \le n)$;
for $t = 1$ *to* m **do**
  Select k with probability $\frac{1}{n}$ $(1 \le k \le n)$;
  Let π^t be a random permutation of $\{0, 1, \ldots, n\}$;
  $\mathcal{S} \leftarrow \{\pi^t(1), \ldots, \pi^t(k)\}$;
  $u \leftarrow v(\mathcal{S} \cap \mathcal{N})$;
  for $i = 1$ *to* k **do**
    $\hat{\phi}^*_{\pi^t(i),k} = \frac{m_{\pi^t(i),k}}{m_{\pi^t(i),k}+1} \hat{\phi}^*_{\pi^t(i),k} + \frac{1}{m_{\pi^t(i),k}+1} u$;
    $m_{\pi^t(i),k} += 1$;
for $i = 1$ *to* n **do**
  $\hat{\phi}_i \leftarrow \sum_{k=1}^{n} \frac{n}{n-k+1} p_k \hat{\phi}^*_{i,k} - \sum_{k=1}^{n} \frac{n}{n-k+1} p_k \hat{\phi}^*_{0,k}$;
return $\hat{\phi}_1, \ldots, \hat{\phi}_n$.

Unstratified Sampling	Stratified Sampling	
$\boxed{\{\mathbf{0}\}}$	$\boxed{\{\mathbf{0}\}}$	0
$\boxed{\{\mathbf{0},1\}}$ $\{\mathbf{0},2\}$ $\{\mathbf{0},3\}$	$\boxed{\{\mathbf{0},1\}}$ $\{\mathbf{0},2\}$ $\{\mathbf{0},3\}$	1
$\{\mathbf{0},1,2\}$ $\boxed{\{\mathbf{0},1,3\}}$ $\boxed{\{\mathbf{0},2,3\}}$	$\{\mathbf{0},1,2\}$ $\boxed{\{\mathbf{0},1,3\}}$ $\boxed{\{\mathbf{0},2,3\}}$	$\frac{11}{2}$
$\hat{\phi}^*_0 = \frac{1}{4}(0+1+5+6) = 3$	$\hat{\phi}^*_0 = \frac{2}{11} \cdot 0 + \frac{3}{11} \cdot 1 + \frac{6}{11} \cdot \frac{11}{2} = \frac{36}{11}$	
$\boxed{\{\mathbf{1}\}}$	$\boxed{\{\mathbf{1}\}}$	1
$\boxed{\{\mathbf{0},1\}}$ $\boxed{\{\mathbf{1},2\}}$ $\boxed{\{\mathbf{1},3\}}$	$\boxed{\{\mathbf{0},1\}}$ $\boxed{\{\mathbf{1},2\}}$ $\boxed{\{\mathbf{1},3\}}$	$\frac{10}{3}$
$\{0,\mathbf{1},2\}$ $\boxed{\{0,\mathbf{1},3\}}$ $\{\mathbf{1},2,3\}$	$\{0,\mathbf{1},2\}$ $\boxed{\{0,\mathbf{1},3\}}$ $\{\mathbf{1},2,3\}$	5
$\hat{\phi}^*_1 = \frac{1}{5}(1+1+4+5+5) = \frac{16}{5}$	$\hat{\phi}^*_1 = \frac{2}{11} \cdot 1 + \frac{3}{11} \cdot \frac{10}{3} + \frac{6}{11} \cdot 5 = \frac{41}{11}$	
$\hat{\phi}_1 = \frac{11}{6}(\hat{\phi}^*_1 - \hat{\phi}^*_0) = \frac{11}{30}$	$\hat{\phi}_1 = \frac{11}{6}(\hat{\phi}^*_1 - \hat{\phi}^*_0) = \frac{5}{6}$	
$\{\mathbf{2}\}$	$\{\mathbf{2}\}$	-
$\{0,\mathbf{2}\}$ $\boxed{\{1,\mathbf{2}\}}$ $\{\mathbf{2},3\}$	$\{0,\mathbf{2}\}$ $\boxed{\{1,\mathbf{2}\}}$ $\{\mathbf{2},3\}$	4
$\{0,1,\mathbf{2}\}$ $\boxed{\{0,\mathbf{2},3\}}$ $\{1,\mathbf{2},3\}$	$\{0,1,\mathbf{2}\}$ $\boxed{\{0,\mathbf{2},3\}}$ $\{1,\mathbf{2},3\}$	6
$\hat{\phi}^*_2 = \frac{1}{2}(4+6) = 5$	$\hat{\phi}^*_2 = \frac{2}{11} \cdot 0 + \frac{3}{11} \cdot 4 + \frac{6}{11} \cdot 6 = \frac{48}{11}$	
$\hat{\phi}_2 = \frac{11}{6}(\hat{\phi}^*_2 - \hat{\phi}^*_0) = \frac{11}{3}$	$\hat{\phi}_2 = \frac{11}{6}(\hat{\phi}^*_2 - \hat{\phi}^*_0) = 2$	
$\{\mathbf{3}\}$	$\{\mathbf{3}\}$	-
$\{0,\mathbf{3}\}$ $\boxed{\{1,\mathbf{3}\}}$ $\{2,\mathbf{3}\}$	$\{0,\mathbf{3}\}$ $\boxed{\{1,\mathbf{3}\}}$ $\{2,\mathbf{3}\}$	5
$\boxed{\{0,1,\mathbf{3}\}}$ $\boxed{\{0,2,\mathbf{3}\}}$ $\{1,2,\mathbf{3}\}$	$\boxed{\{0,1,\mathbf{3}\}}$ $\boxed{\{0,2,\mathbf{3}\}}$ $\{1,2,\mathbf{3}\}$	$\frac{11}{2}$
$\hat{\phi}^*_3 = \frac{1}{3}(5+5+6) = \frac{16}{3}$	$\hat{\phi}^*_3 = \frac{2}{11} \cdot 0 + \frac{3}{11} \cdot 5 + \frac{6}{11} \cdot \frac{11}{2} = \frac{48}{11}$	
$\hat{\phi}_3 = \frac{11}{6}(\hat{\phi}^*_3 - \hat{\phi}^*_0) = \frac{77}{18}$	$\hat{\phi}_3 = \frac{11}{6}(\hat{\phi}^*_3 - \hat{\phi}^*_0) = 2$	

Table 4. Unstratified Sampling and Stratified Sampling for Shapley Value Estimation.

generally achieves lower estimation variance than unstratified sampling. However, despite its

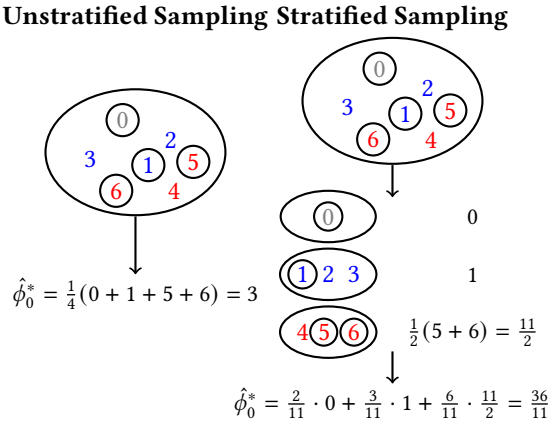


Fig. 2. Unstratified sampling and stratified sampling .

potential for variance reduction, static stratified sampling (e.g., Algorithm 2) is not a one-size-fits-all solution. As implied by Theorem 4.2, static stratified sampling cannot guarantee unbiasedness without sufficient samples to ensure full coverage of all n strata. For example, in Table 4, the stratum corresponding to coalition size 1 is empty due to the lack of samples. Consequently, the term $\phi_{2,1}$ is forced to a default value of 0, and the estimator becomes $\hat{\phi}_2^* = \frac{2}{11}\cdot 0 + \frac{3}{11}\hat{\phi}_{2,2}^* + \frac{6}{11}\hat{\phi}_{2,3}^*$. This resulting estimate is strictly biased, as visualized in Figure 3. Therefore, while static stratification is theoretically superior in variance, it becomes structurally flawed and biased when the sample budget falls below the threshold required for full coverage.

For stratified sampling to yield an unbiased estimator, a fundamental prerequisite is that every stratum must be non-empty (i.e., full coverage). We propose the following theorem to quantify the expected number of samples required to fill all strata for Shapley-like value estimation.

Theorem 4.2. *Let m_n denote the number of utility samples required to fill n strata in stratified sampling. The expected sample size is lower bounded by*

$$\mathrm{E}[m_n] \geq n(n-1)\sum_{k=1}^{n}\frac{1}{k} \approx n^2(\ln(n)+\gamma) = \Theta(n^2\log n),$$

where γ refers to the Euler-Mascheroni constant with an approximate value of 0.57721.

Theorem 4.2 illustrates that static stratified sampling is impractical when the number of players is large, as it requires $\Omega(n^2\log n)$ samples to ensure full coverage of all strata—the necessary condition for unbiasedness. In scenarios with insufficient samples (i.e., limited budgets), the "empty stratum" problem is likely to occur, thereby breaking the unbiasedness guarantee. In contrast, according to [36], GELS only requires $O(\frac{n}{\varepsilon^2}\log\frac{n}{\delta})$ unstratified utility samples to achieve an (ε,δ)-approximation. Consequently, static stratified sampling is often unsuitable for many practical scenarios, particularly when scaling to large datasets or operating under strict budgetary limits.

5 Adaptive Stratified Sampling

Considering the limitations of stratified sampling, one may wonder: can we retain the benefits of stratified sampling while mitigating its drawbacks? The answer is affirmative. The primary reason stratified sampling fails to provide unbiased estimators with insufficient samples is the large number of strata, making them difficult to fill efficiently. To tackle the problem, a practical solution is to reduce the number of strata.

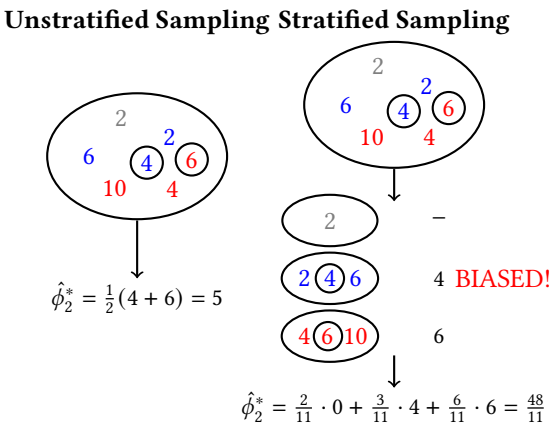


Fig. 3. Limitation of static stratified sampling

In this section, we propose ASSS (**A**daptive **S**tratified **S**ampling for **S**hapley-like estimation) based on this intuition, which adaptively stratifies according to the number of samples. We propose two strategies, ASSS-A (stratify **A**fter sampling) and ASSS-B (stratify **B**efore sampling). Overall, ASSS-A eliminates unsampled strata after sampling to ensure unbiased estimation; ASSS-B determines a smaller number of strata based on the sample size to ensure that all strata are filled.

5.1 Proper Sample Allocation for Stratification

We first introduce how to allocate samples effectively across various strata. Existing methods [30, 64] distribute equal sample sizes across all strata, assuming no prior knowledge. However, stratified sampling guarantees variance reduction compared to non-stratified sampling only when the probability of sampling coalitions of each length aligns with that of non-stratified sampling. Otherwise, stratified sampling may fail to reduce variance.

Let $\sigma_{i,k}^2$ represent the utility variance of coalitions containing player i with size k. For player i, given $m_{i,k}$ samples in the k-th stratum, the variance of the stratified estimator $\hat{\phi}_i^{*strat}$ of ϕ_i^* is

$$\mathrm{Var}(\hat{\phi}_i^{*strat}) = \sum_{k=1}^{n} w_k^2 \frac{\sigma_{i,k}^2}{m_{i,k}}. \tag{6}$$

For player i, given m_i total samples, the variance of the unstratified estimator $\hat{\phi}_i^{*unstrat}$ is

$$\mathrm{Var}(\hat{\phi}_i^{*unstrat}) = \frac{1}{m_i} \sum_{k=1}^{n} w_k \sigma_{i,k}^2 + \frac{1}{m_i} \sum_{k=1}^{n} w_k (\phi_{i,k}^* - \phi_i^*)^2. \tag{7}$$

Equations 6 and 7 define the conditions under which stratified sampling reduces the variance of Shapley-like value estimates.

Theorem 5.1. *When $m_{i,k} = w_k m_i$ for all $1 \le k \le n$, we have $\mathrm{Var}(\hat{\phi}_i^{*strat}) \le \mathrm{Var}(\hat{\phi}_i^{*unstrat})$.*

Corollary 5.2. *The following corollary generalizes Theorem 5.1 under specific sampling conditions. When $m_k \propto \frac{n(n+1)}{k(n+1-k)} p_k$ for all $1 \le k \le n$, we have $\mathrm{Var}(\hat{\phi}_i^{*strat}) \le \mathrm{Var}(\hat{\phi}_i^{*unstrat})$ given the same amount of samples.*

According to Theorem 5.1 and Corollary 5.2, when stratified sampling ratios match the unstratified proportions, stratified sampling is guaranteed to reduce variance. However, the variance of stratified sampling may exceed that of direct sampling for other sample allocations. In other

words, when performing stratified sampling, without prior knowledge, we should set the sampling probabilities to match those used in the unstratified sampling algorithm, Algorithm 1. These results offer a theoretical basis for optimizing stratified sampling in practical scenarios, especially when prior knowledge about coalition utility distributions is limited.

5.2 Stratification after Sampling (ASSS-A)

Based on Section 5.1, setting the same sampling probabilities as in unstratified sampling ensures variance reduction. However, when the number of samples is insufficient, some strata may remain empty. This raises a natural question: what if we simply ignore the unsampled strata? By doing so, we can retain the structure of the sampled strata while discarding the unsampled ones. This approach still yields unbiased estimates and, by preserving the structure of sampled strata, continues to reduce variance.

The main idea of ASSS-A is to perform stratification after drawing samples. First, we collect samples using the original stratified sampling method. During the computation of results, we discard strata that have not been sampled and compute the weighted average of the remaining strata. As shown on the left side of Table 5, we compute the weighted average of $\hat{\phi}^*_{2,2}$ and $\hat{\phi}^*_{2,3}$, i.e., $\hat{\phi}^*_1 = \frac{1}{3}\hat{\phi}^*_{2,2} + \frac{2}{3}\hat{\phi}^*_{2,3}$. We refer to this type of stratified sampling as ASSS-A (stratify **A**fter sampling).

To ensure unbiased estimation, the probability of selecting a coalition of size k must remain consistent with unstratified sampling, i.e., $\propto \frac{n(n+1)}{k(n+1-k)} p_k$. By setting such sampling ratios, we can maintain unbiased estimators while reducing variance. The detailed algorithm is presented in Algorithm 3. First, we sample coalitions $\mathcal{S}$ from $\mathcal{N} \cup \{0\}$ based on the proper probability distribution and calculate the utility $v(\mathcal{S} \cap \mathcal{N})$ (Lines 3-6). The computed utility $v(\mathcal{S} \cap \mathcal{N})$ is then reused to estimate $\phi^*_{i,k}$ for all players $i \in \mathcal{S}$ in the k -th stratum (Lines 7-9). After collecting m samples, we average the utilities in each stratum, remove unsampled strata, and derive the estimated Shapley-like values (Lines 10-17).

We now demonstrate that the resulting estimators under this method are unbiased. This property allows the method to effectively leverage the variance reduction benefits of stratified sampling, resulting in higher estimation accuracy.

Theorem 5.3. *Algorithm 3 gives unbiased estimators.*

5.3 Stratification before Sampling (ASSS-B)

Beyond deciding which strata to retain after sampling, we propose a more resource-efficient method. Specifically, we determine the number of strata based on the sample size, ensuring that each stratum receives a sufficient number of samples. To achieve this, we generalize the concept of stratification beyond the traditional approach of dividing coalitions into n strata based solely on their sizes.

Definition 5.4 (Stratification for Shapley-like values). A stratification $\mathcal{L}$ for the Shapley-like value with parameters $n, v(\cdot), \mathbf{p}$ is a partition of $\{1, \ldots, n\}$ into $\{\mathcal{L}_1, \ldots, \mathcal{L}_L\}$ such that $\bigcup_{k=1}^{L} \mathcal{L}_k = \{1, \ldots, n\}$ and $\mathcal{L}_{k1} \cap \mathcal{L}_{k2} = \varnothing$ $(1 \le k1 < k2 \le L)$. The expectation of utilities in the k-th stratum $\phi^*_i(\mathcal{L}_k)$ is defined as

$$\phi^*_i(\mathcal{L}_k) = \mathrm{E}[v(\mathcal{S}) \mid i \in \mathcal{S} \subsetneq \mathcal{N} \cup \{0\}, |\mathcal{S}| \in \mathcal{L}_k, \Pr(\mathcal{S}) \propto p_{|\mathcal{S}|+1}].$$

With this generalized stratification $\mathcal{L} = \{\mathcal{L}_1, \ldots, \mathcal{L}_L\}$, we redefine ϕ^*_i as follows:

Theorem 5.5.

$$\phi^*_i = \sum_{k=1}^{L} \omega_k \phi^*_i(\mathcal{L}_k),$$

where $\omega_k = \sum_{l \in \mathcal{L}_k} w_l$.

Algorithm 3: Stratification after sampling (ASSS-A).

input : $\mathcal{N} = \{1, \ldots, n\}$, $v : 2^{\mathcal{N}} \to \mathbb{R}$, $\mathbf{p} \in \mathbb{R}^n$, and $m > 0$
output: the estimated Shapley-like values $\hat{\phi}_1, \ldots, \hat{\phi}_n$

$\hat{\phi}^*_{i,k}, m_{i,k} \leftarrow 0$ $(0 \le i \le n, 1 \le k \le n)$;
for $t = 1$ *to* m **do**
  Select k with probability $\propto \frac{n(n+1)}{k(n+1-k)} p_k$ $(1 \le k \le n)$;
  Let π^t be a random permutation of $\{0, 1, \ldots, n\}$;
  $\mathcal{S} \leftarrow \{\pi^t(1), \ldots, \pi^t(k)\}$;
  $u \leftarrow v(\mathcal{S} \cap \mathcal{N})$;
  for $i = 1$ *to* k **do**
    $\hat{\phi}^*_{\pi^t(i),k} = \frac{m_{\pi^t(i),k}}{m_{\pi^t(i),k}+1} \hat{\phi}^*_{\pi^t(i),k} + \frac{1}{m_{\pi^t(i),k}+1} u$;
    $m_{\pi^t(i),k} += 1$;
$c \leftarrow \sum_{k=1}^{n} \frac{n}{n-k+1} p_k$;
for $i = 0$ *to* n **do**
  $\mathcal{L}_i \leftarrow \varnothing$;
  for $k = 1$ *to* n **do**
    if $m_{i,k} \neq 0$ **then**
      $\mathcal{L}_i = \mathcal{L}_i \cup \{k\}$;
  $c_i \leftarrow \sum_{k \in \mathcal{L}_i} \frac{n}{n-k+1} p_k$;
  $\hat{\phi}_i \leftarrow \sum_{k \in \mathcal{L}_i} \frac{c}{c_i} \frac{n}{n-k+1} p_k \hat{\phi}^*_{i,k} - \sum_{k \in \mathcal{L}_i} \frac{c}{c_0} \frac{n}{n-k+1} p_k \hat{\phi}^*_{0,k}$;
return $\hat{\phi}_1, \ldots, \hat{\phi}_n$.

We can proceed with generalized stratified sampling using Definition 5.4 and Theorem 5.5. For a stratum $\mathcal{L}$ and ϕ^*_i, we first estimate the expected value in each stratum $\hat{\phi}^*_i(\mathcal{L}_k)$ using Monte Carlo sampling. Then, we compute $\hat{\phi}^*_i$ as the weighted average of $\hat{\phi}^*_i(\mathcal{L}_k)$ $(1 \le k \le L)$ based on Theorem 5.5.

As shown on the right side of Table 5, for Example 4.1, regarding ϕ^*_2, we first divide all the coalitions into two strata. The first stratum includes coalitions of size 1 and 2, and the second stratum includes coalitions of size 3, i.e., $\mathcal{L}_1 = \{1, 2\}$ and $\mathcal{L}_2 = \{3\}$. We first calculate the mean utility within each stratum, $\hat{\phi}^*_2(\mathcal{L}_1) = 4$ and $\hat{\phi}^*_2(\mathcal{L}_2) = 6$, and compute $\omega_1 = \frac{2}{11} + \frac{3}{11} = \frac{5}{11}$, and $\omega_2 = \frac{6}{11}$. Then, we calculate $\hat{\phi}^*_2 = \omega_1 \hat{\phi}^*_2(\mathcal{L}_1) + \omega_2 \hat{\phi}^*_2(\mathcal{L}_2)$.

Similar to Section 5.1, we can calculate the variance of the estimator $\hat{\phi}^{*strat}_i(\mathcal{L})$ of ϕ^*_i with the stratification $\mathcal{L} = \{\mathcal{L}_1, \ldots, \mathcal{L}_L\}$ as

$$\mathrm{Var}(\hat{\phi}^{*strat}_i(\mathcal{L})) = \sum_{k=1}^{L} \omega_k^2 \frac{\sigma_i^2(\mathcal{L}_k)}{m_{i,k}}, \tag{8}$$

where $\sigma_i^2(\mathcal{L}_k)$ is the variance of the utility of the coalitions containing player i in the stratum $\mathcal{L}_k$ and $m_{i,k}$ is the number of samples in the k-th stratum $\mathcal{L}_k$ for player i.

Given that we now have a generalized definition of stratification, we can extend Theorem 5.1 and Corollary 5.2 to the generalized case. We first compute the variance using the generalized concept of stratification. The variance of the unstratified estimator $\hat{\phi}^{*unstrat}_i$ can be computed as

$$\mathrm{Var}(\hat{\phi}_i^{*unstrat}) = \frac{1}{m_i}\sum_{k=1}^{L}\omega_k\sigma_{i,k}^2(\mathcal{L}) + \frac{1}{m_i}\sum_{k=1}^{n}\omega_k(\phi_i^*(\mathcal{L}_k) - \phi_i^*)^2. \tag{9}$$

Similarly, based on Equations 8 and 9, we can also calculate the sampling probabilities under which stratified sampling can guarantee variance reduction with a stratification $\mathcal{L}$.

Theorem 5.6. *When $m_{i,k} = \omega_k m_i$ for all $1 \le k \le L$, we have $\mathrm{Var}(\hat{\phi}_i^{*strat}(\mathcal{L})) \le \mathrm{Var}(\hat{\phi}_i^{*unstrat})$. Specifically, we have*

$$\mathrm{Var}(\hat{\phi}_i^{*unstrat}) - \mathrm{Var}(\hat{\phi}_i^{*strat}(\mathcal{L})) = \frac{1}{m_i}\sum_{k=1}^{n}\omega_k(\phi_i^*(\mathcal{L}_k) - \phi_i^*)^2.$$

Corollary 5.7. *When $m_k \propto \sum_{l\in\mathcal{L}_i}\frac{n(n+1)}{l(n+1-l)}p_l$ for all $1 \le k \le L$ for all $1 \le k \le L$, we have $\mathrm{Var}(\hat{\phi}_i^{*strat}) \le \mathrm{Var}(\hat{\phi}_i^{*unstrat})$ given the same amount of samples.*

According to Theorem 5.6 and Corollary 5.7, a generalized non-stratified sampling method that adopts the same sampling proportions as unstratified sampling can also reduce variance. We consider the conditions that should be satisfied when performing stratification to minimize variance.

Theorem 5.8. *The optimal stratification $\mathcal{L} = \mathcal{L}_1, \ldots, \mathcal{L}_L$ for the objective function $\min \mathrm{Var}(\hat{\phi}_i^{*strat}(\mathcal{L}))$ satisfies that $\forall k_1 < k_2, \forall l_1 \in \mathcal{L}_{k_1}, l_2 \in \mathcal{L}_{k_2}$, we have $\phi_{i,l_1}^* < \phi_{i,l_2}^*$.*

Theorem 5.8 tells that the optimal stratification for ϕ_i^* should divide the strata consecutively according to the ordering of $\phi_{i,k}^*(1 \le k \le n)$. Suppose $\phi_{i,k}^*$ $(1 \le k \le n)$ is monotonically increasing with k, which holds in general cases. We can conclude that the optimal $\mathcal{L}$ should divide the strata consecutively according to the order $1, 2, \ldots, n$. Leveraging this structural property, we can formulate the search for the theoretically optimal stratification boundaries as a dynamic programming problem.

Proposition 5.9 (Optimal K-Layer Stratification – DP Formulation). *For a given player $i \in \mathcal{N}$ and an integer $K \le n$, consider a K-layer stratification $\mathcal{L} := (\mathcal{L}_1, \ldots, \mathcal{L}_K)$ induced by cut indices $0 = \tau_0 < \tau_1 < \cdots < \tau_K = n$, where $\mathcal{L}_r := \{\tau_{r-1} + 1, \ldots, \tau_r\}$ for $r = 1, \ldots, K$. The optimal K-layer stratification solves*

$$\min_{0=\tau_0<\tau_1<\cdots<\tau_{K-1}<\tau_K=n}\ \sum_{r=1}^{K}\ \sum_{k=\tau_{r-1}+1}^{\tau_r}\left(\phi_{i,k}^* - \phi_i^*(\mathcal{L}_r)\right)^2.$$

$\forall 1 \le p, q \le n$, define the mean operator on the interval $\{p, \ldots, q\}$ by

$$\mathcal{M}_{p:q}(\phi_i^*) := \frac{1}{q-p+1}\sum_{k=p}^{q}\phi_{i,k}^*,$$

and define the within-layer squared error

$$\mathrm{c}(p,q) := \min_{\mu\in\mathbb{R}}\sum_{k=p}^{q}(\phi_{i,k}^* - \mu)^2 = \sum_{k=p}^{q}\left(\phi_{i,k}^* - \mathcal{M}_{p:q}(\phi_i^*)\right)^2.$$

Let $\mathrm{dp}[g][q]$ be the minimum objective value for partitioning $\{1, \ldots, q\}$ into g contiguous layers. Then the optimum value is given by

$$\mathrm{dp}[1][q] = \mathrm{c}(1,q),\ \mathrm{dp}[g][q] = \min_{g-1\le p\le q}\Big(\mathrm{dp}[g-1][p-1] + \mathrm{c}(p,q)\Big),$$

for $g = 2, \ldots, K$ *and* $q = 1, \ldots, n$.

As presented in Proposition 5.9, the optimal stratification strategy requires solving a dynamic programming problem to minimize variance. However, implementing this theoretical optimality relies on the variance of utility distributions. Estimating the variances would consume significant pilot sampling budgets, and solving the DP entails a non-trivial time complexity of $O(Ln^2)$. Consequently, given these constraints, we simply adopt a naïve equal-weight strategy to construct the strata. To balance the need for stratum coverage and the available sample size, we define $L = \frac{m}{2n \log n}$ (bounded by n). As we will demonstrate in Theorem 5.10, this specific choice is sufficient to ensure full stratum coverage with high probability.

The detailed algorithm is presented in Algorithm 4. First, we determine the number of strata L based on the number of samples (Line 1). Then we properly divide all the coalitions into L strata (Lines 4-9). We sample coalitions $\mathcal{S}$ with proper probability distribution from $\mathcal{N} \cup \{0\}$ in each stratum and calculate utility $v(\mathcal{S} \cap \mathcal{N})$ (Lines 12-15). $v(\mathcal{S} \cap \mathcal{N})$ is then reused in the estimation of $\phi^*_{i,k}$ for all players $i \in \mathcal{S}$ (Lines 16-18). After m samples, we average coalition utilities in each stratum and derive the estimated Shapley-like values (Lines 19-20).

Algorithm 4: Stratification before sampling (ASSS-B).

input : $\mathcal{N} = \{1, \ldots, n\}$, $v : 2^{\mathcal{N}} \to \mathbb{R}$, $\mathbf{p} \in \mathbb{R}^n$, and $m > 0$
output: the estimated Shapley-like values $\hat{\phi}_1, \ldots, \hat{\phi}_n$

1 $L \leftarrow \max(\lfloor \frac{m}{2n \log n} \rfloor, n)$;
2 $\hat{\phi}^*_{i,k}, m_{i,k} \leftarrow 0$ $(0 \le i \le n, 1 \le k \le L)$;
3 $s \leftarrow 1$;
4 **for** $k = 1$ *to* L **do**
5 $\mathcal{L}_k \leftarrow \emptyset$;
6 **while** $\sum_{l \in \mathcal{L}_k} p_l < \frac{1}{L}$ *and* $s \le n$ **do**
7 $\mathcal{L}_k = \mathcal{L}_k \cup \{s\}$;
8 $s+ = 1$;
9 $m_k \leftarrow m \sum_{l \in \mathcal{L}_k} p_l$;
10 **for** $k = 1$ *to* L **do**
11 **for** $t = 1$ *to* m_k **do**
12 Select l with probability $\propto \frac{n(n+1)}{l(n+1-l)} p_l$ $(l \in \mathcal{L}_i)$;
13 Let π^t be a random permutation of $\{0, 1, \ldots, n\}$;
14 $\mathcal{S} \leftarrow \{\pi^t(1), \ldots, \pi^t(l)\}$;
15 $u \leftarrow v(\mathcal{S} \cap \mathcal{N})$;
16 **for** $i = 1$ *to* l **do**
17 $\hat{\phi}^*_{\pi^t(i),k} = \frac{m_{\pi^t(i),k}}{m_{\pi^t(i),k}+1} \hat{\phi}^*_{\pi^t(i),k} + \frac{1}{m_{\pi^t(i),k}+1} u$;
18 $m_{\pi^t(i),k}+ = 1$;
19 **for** $i = 1$ *to* n **do**
20 $\hat{\phi}_i \leftarrow \sum_{k=1}^{L} \sum_{l \in \mathcal{L}_k} \frac{n}{n-l+1} p_l (\hat{\phi}^*_{i,k} - \hat{\phi}^*_{0,k})$;
21 **return** $\hat{\phi}_1, \ldots, \hat{\phi}_n$.

We now demonstrate that the resulting estimators achieve unbiasedness with a probability that asymptotically approaches 1 as the number of players n increases.

Theorem 5.10. *Algorithm 4 needs $\Omega(L^2 \log L)$ samples to fill L strata, and gives unbiased estimators with probability greater than $1 - \frac{2}{n}$.*

ASSS (Type A)		ASSS (Type B)	
$\{2\}$	-	$\{2\}$	4
$\boxed{\{0,2\}}$ $\boxed{\{1,2\}}$ $\{2,3\}$	4	$\boxed{\{0,2\}}$ $\boxed{\{1,2\}}$ $\{2,3\}$	
$\{0,1,2\}$ $\boxed{\{0,2,3\}}$ $\{1,2,3\}$	6	$\{0,1,2\}$ $\boxed{\{0,2,3\}}$ $\{1,2,3\}$	6
$\hat{\phi}_2^* = \frac{1}{3} \cdot 4 + \frac{2}{3} \cdot 6 = \frac{16}{3}$		$\hat{\phi}_2^* = \frac{2+3}{11} \cdot 4 + \frac{6}{11} \cdot 5 = \frac{56}{11}$	
$\{3\}$	-	$\{3\}$	5
$\boxed{\{0,3\}}$ $\{1,3\}$ $\{2,3\}$	5	$\boxed{\{0,3\}}$ $\{1,3\}$ $\{2,3\}$	
$\{0,1,3\}$ $\{0,2,3\}$ $\boxed{\{1,2,3\}}$	$\frac{11}{2}$	$\{0,1,3\}$ $\{0,2,3\}$ $\boxed{\{1,2,3\}}$	$\frac{11}{2}$
$\hat{\phi}_3^* = \frac{1}{3} \cdot 5 + \frac{2}{3} \cdot \frac{11}{2} = \frac{16}{3}$		$\hat{\phi}_3^* = \frac{2+3}{11} \cdot 5 + \frac{6}{11} \cdot \frac{11}{2} = \frac{58}{11}$	

Table 5. Adaptive stratified sampling.

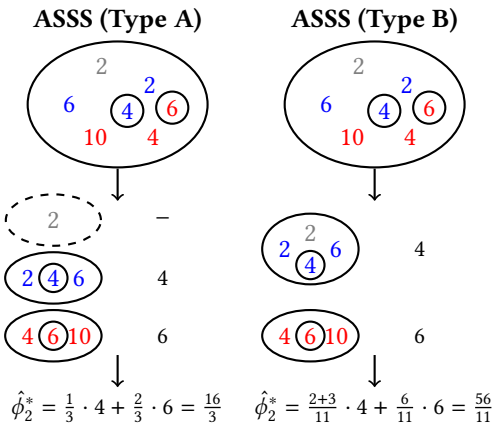


Fig. 4. Adaptive stratified sampling.

5.4 Theoretical Analysis

In this section, we present a comprehensive theoretical analysis of our proposed approaches, including a comparison with the state-of-the-art method GELS [36] and an evaluation of computational complexity.

5.4.1 Comparison. A theoretical comparison was conducted between our proposed methods (ASSS-A and ASSS-B) and GELS in terms of estimation variance. Specifically, we consider the stratification of ASSS-B the setting in Algorithm 4.

Theorem 5.11. *Denote by $\hat{\phi}_i^{*A}$, $\hat{\phi}_i^{*B}$, and $\hat{\phi}_i^{*G}$ the estimators of ϕ_i^* provided by ASSS-A, ASSS-B, and GELS with same samples, we have $\mathrm{Var}(\hat{\phi}_i^{*A}) \leq \mathrm{Var}(\hat{\phi}_i^{*G})$ and $\mathrm{Var}(\hat{\phi}_i^{*B}) \leq \mathrm{Var}(\hat{\phi}_i^{*G})$ for any arbitrary i.*

Method	Sample Complexity	Stratification	Variance Reductions
GELS	$O(\frac{n}{\varepsilon^2}\ln\frac{n}{\delta})$	✗	✗
ASSS-A	$O(\frac{n}{\varepsilon^2}\ln\frac{n}{\delta})$	Adaptive	$\frac{E_i}{E_i+V_i}+O(m^{-2})$
ASSS-B	$O(\frac{n}{\varepsilon^2}\ln\frac{n}{\delta})$	Adaptive	$\frac{E_i}{E_i+V_i}+O(m^{-2})$

Table 6. Comparison of GELS, ASSS-A, and ASSS-B.

Theorem 5.11 establishes that our stratified approaches theoretically achieve lower variance than the unstratified GELS. Furthermore, we derive explicit bounds on the reduction in variance for ASSS-A and ASSS-B based on the distribution of utilities.

Theorem 5.12. *Assume that $v(\mathcal{S}) < 2\phi^*_{i,|\mathcal{S}|}$ for all $i \in \mathcal{S}$, then the estimator ratio satisfies $\frac{\hat{\phi}_i^{*A}}{\hat{\phi}_i^{*G}} \le \frac{E_i}{E_i+V_i}+O(m^{-2})$, where $E_i = \mathrm{E}[(\phi^*_{i,k})^2], V_i = \mathrm{Var}(\phi^*_{i,k}), 1 \le k \le n$, and $\Pr(\phi^*_{i,k}) = w_k$.*

Theorem 5.13. *Given a stratification $\mathcal{L} = \{\mathcal{L}_1, \mathcal{L}_2, \ldots, \mathcal{L}_L\}$, assume that $v(\mathcal{S}) < 2\phi^*_{i,|\mathcal{S}|}$ for all $i \in \mathcal{S}$. Then, the estimator ratio satisfies $\frac{\hat{\phi}_i^{*B}}{\hat{\phi}_i^{*G}} \le \frac{E_i}{E_i+V_i}+O(m^{-2})$, where $E_i = \mathrm{E}[\phi_i^*(\mathcal{L}_k)^2], V_i = \mathrm{Var}(\phi_i^*(\mathcal{L}_k)), 1 \le k \le L$, and $\Pr(\phi_i^*(\mathcal{L}_k)) = \omega_k$.*

Note that the inequalities for ASSS-B holds with full stratum coverage. In practical regimes, our design satisfies the condition with high probability, validating the variance reduction.

5.4.2 Computational Complexity. We determined the computational complexity by calculating the expected sample size. To provide a theoretical benchmark, we first derived the expected sample size required to achieve a specified level of estimation precision.

Theorem 5.14. *Assuming the expected error of ASSS-A and ASSS-B with m utility samples is ε, we have $\varepsilon = O(\sqrt{\frac{n}{m}})$, i.e., $m = O(\frac{n}{\varepsilon^2})$.*

Theorem 5.14 highlights the scalability of our methods, as the sample complexity scales favorably with the number of players n and the desired precision ε. This result demonstrates the efficiency of our approaches in practical scenarios, particularly when computational resources are limited.

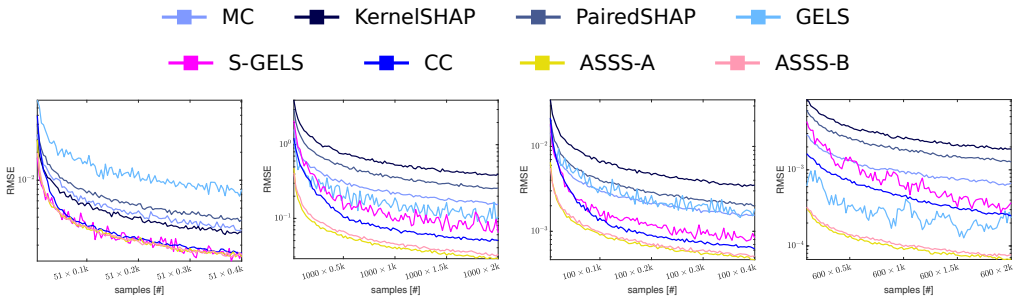
In addition to theoretical expectations, the sample size can also be determined by imposing confidence interval constraints under different distributional assumptions. Specifically, we estimate the minimum number of samples required to ensure that the confidence interval remains below a specified error threshold.

Case 1: Normal Distribution. Under the normality assumption, the required sample size can be determined with a tight bound.

Theorem 5.15. *If $\phi_i^* \sim \mathcal{N}(\mu_i, \sigma_i^2)$ and $\sigma_i \le \sigma$, then with at least $\frac{(2z_{1-\delta/2}\sigma)^2}{\varepsilon^2}$ samples, ASSS-A and ASSS-B can achieve an (ε,δ)-approximation, where $z_{1-\delta/2}$ denotes the standard normal quantile. Specifically, for a 95% confidence level, $z_{0.975} = 1.96$.*

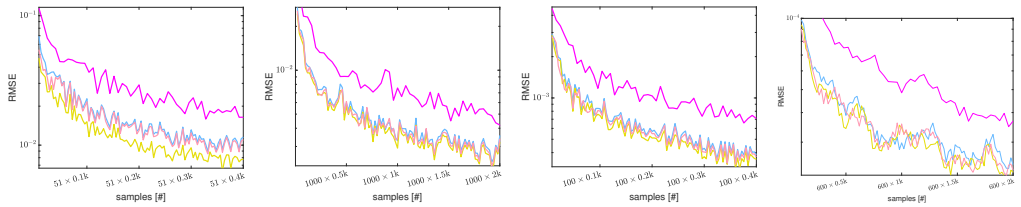
Case 2: Chebyshev Bound. For arbitrary distributions with finite variance, the Chebyshev inequality yields a bound when the variance is bounded.

Theorem 5.16. *If $\mathrm{Var}(\phi_i^*) \le \sigma^2$, then with at least $\frac{4\sigma^2}{\delta\varepsilon^2}$ samples, ASSS-A and ASSS-B can achieve an (ε,δ)-approximation.*



(a) Voting. (b) Airport. (c) Iris. (d) Breast Cancer.

Fig. 5. Shapley value effectiveness (RMSE).



(a) Voting. (b) Airport. (c) Iris. (d) Breast Cancer.

Fig. 6. Banzhaf value effectiveness (RMSE).

In Cases 1 and 2, the variance can be estimated from samples using Bessel's correction; however, it prevents a direct estimate of the required sample size. In the following, we introduce an approach that only requires knowledge of the utility bound.

Case 3: Hoeffding Bound. Hoeffding's bound requires only the known range of the variable, making it the most practical and widely used approach for bounded variables.

THEOREM 5.17. *If* $\mathcal{U}(\mathcal{S}) \in [0, u_0]$, *then with at least* $\frac{2nu_0^2}{\epsilon^2} \ln \frac{4n}{\delta}$ *samples, ASSS-A and ASSS-B can achieve an* (ϵ, δ)*-approximation.*

Theorems 5.15-5.17 establish the expected sample sizes under different conditions and indicate that the proposed methods achieve competitive sample complexity while guaranteeing the prescribed precision and confidence levels. Table 6 summarizes the comparison between GELS and our two adaptive stratified estimators. ASSS-A and ASSS-B match GELS in the leading-order Hoeffding bound sample complexity, while reducing variance through adaptive stratification by a factor $\frac{E_i}{E_i+V_i} + O(m^{-2})$. This highlights that our gain primarily comes from variance reduction rather than altering the asymptotic sample complexity.

6 Experiments

This section presents experiments evaluating the efficiency and effectiveness of our proposed algorithms.

6.1 Experiment Setup

6.1.1 Methods Compared. Our main baseline is the state-of-the-art method GELS for Shapley-like values [36].

- **GELS** [36]: a utility-based Monte Carlo simulation algorithm (Algorithm 1).

- **S-GELS**: an extension of GELS using a straightforward stratification strategy (Algorithm 2).

For Shapley values, we compare additional baselines as follows.

- **MC** [9]: a Monte Carlo simulation algorithm based on marginal contributions.
- **CC** [64]: a Monte Carlo simulation algorithm using complementary contributions.
- **KernelSHAP** [44]: a kernel-based optimization method for utility sampling.
- **PairedSHAP** [12]: an enhanced version of KernelSHAP leveraging pairwise utility sampling.

We also evaluate the two versions of our proposed methods.

- **ASSS-A**: adaptive stratified sampling with stratification applied after sampling (Algorithm 3).
- **ASSS-B**: adaptive stratified sampling with stratification applied before sampling (Algorithm 4).

6.1.2 Evaluation Tasks. To evaluate the performance of our algorithms, we estimate Shapley-like values across four application scenarios: two cooperative games and two data valuation tasks. These tasks represent diverse settings, ranging from structured game-theoretic problems to machine learning-based data valuation.

A voting game [52]. In a non-symmetric voting game simulating a U.S. presidential election, the Shapley value measures the voting impact of individual players. The set of players is $\mathcal{N} = \{1, \ldots, 51\}$. For any coalition $\mathcal{S} \subseteq \mathcal{N}$, the utility function is defined as

$$v(\mathcal{S}) = \begin{cases} 1, & \text{if } \sum_{i \in \mathcal{S}} \nu_i \geq \frac{1}{2} \sum_{j \in \mathcal{N}} \nu_j, \\ 0, & \text{otherwise,} \end{cases}$$

where the weights $\{\nu_1, \ldots, \nu_{51}\}$ represent the voting power of each player.

An airport game [40]. In this game, the Shapley value allocates airstrip costs among planes of varying sizes. The set of players is $\mathcal{N} = \{1, \ldots, 1000\}$. For any coalition $\mathcal{S} \subseteq \mathcal{N}$, the utility function is

$$v(\mathcal{S}) = \max_{i \in \mathcal{S}} \{i\}.$$

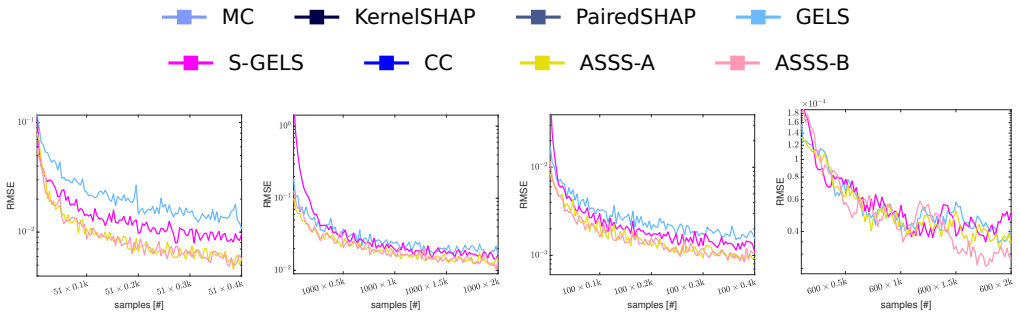
Data valuation tasks on Iris and Breast Cancer datasets. In these tasks, players are data points used to train a predictive model. We employ two benchmark datasets from the UCI Machine Learning Repository [20, 61]—**Iris** and **Breast Cancer**. The **Iris** dataset consists of 150 instances with 4 numerical features and 3 classes. We randomly select $n = 100$ instances as players for the training set, and the remaining 50 are used for validation. The **Breast Cancer** dataset contains 699 instances characterized by 9 features and 2 classes. We selected $n = 600$ instances as players, and the remaining 99 are used for validation. For both tasks, the utility function $v(\mathcal{S})$ is defined as the classification accuracy of a Support Vector Machine (SVM) trained on the subset $\mathcal{S}$ and evaluated on the fixed validation set.

6.1.3 Evaluation Metric. We use RMSE as the evaluation metric to quantify the accuracy of estimated Shapley-like values.

Root Mean Square Error (RMSE). Given benchmark Shapley values ϕ_i and estimated Shapley values $\hat{\phi}_i$ $(i = 1, \ldots, n)$, RMSE is defined as:

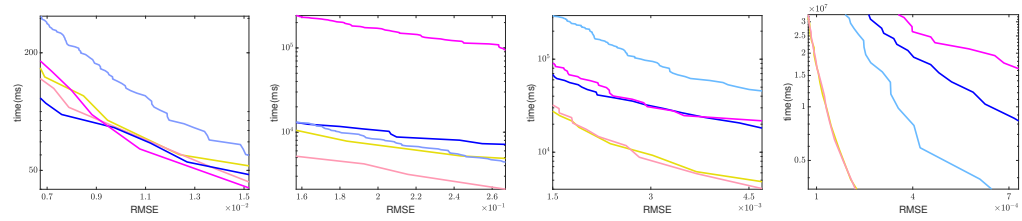
$$\text{RMSE} = \left(\frac{1}{n} \sum_{i=1}^{n} \left| \hat{\phi}_i - \phi_i \right|^2 \right)^{1/2}.$$

Calculating the exact Shapley value ϕ_i is computationally prohibitive due to the exponential growth in complexity with the number of players. Therefore, we use the estimated Shapley-like values computed using GELS [36], with 10000k samples as benchmark Shapley-like values for data



(a) Voting. (b) Airport. (c) Iris. (d) Breast Cancer.

Fig. 7. Beta(2, 2)-Shapley value effectiveness (RMSE).



(a) Voting. (b) Airport. (c) Iris. (d) Breast Cancer.

Fig. 8. Shapley value efficiency.

valuation tasks. For cooperative games, we employ the Shapley values reported in [9] for the voting game and those computed in [39] for the airport game as the benchmark Shapley and Banzhaf values. For other cooperative Shapley-like values that can not be computed in polynomial time, we similarly use the method in [36] with 500k samples as the benchmark Shapley values.

6.2 Effectiveness

We experimentally analyze the performance of ASSS-A and ASSS-B with an equal number of samples across four tasks: Voting, Airport, Iris, and Breast Cancer for Shapley-like value estimation. We set different amounts of samples based on the number of players and examine the performance of each method under conditions where the sample size is relatively insufficient. The empirical results, presented in Figures 5-7, consistently demonstrate the superiority of our approaches over the baselines on the computation of Shapley values, Banzhaf values, and Beta Shapley values.

ASSS-A and ASSS-B perform similarly overall, with ASSS-B slightly underperforming ASSS-A on some tasks. However, ASSS-B requires less space and has more potential for optimization in sample allocation. In Shapley value estimation, it can be observed that as the number of samples increases and that all strata in the CC and S-GELS gradually become filled, the advantage of ASSS-A and ASSS-B gradually diminishes. However, ASSS-A and ASSS-B demonstrate significant improvements over both CC and S-GELS as the number of players increases (Airport and Breast Cancer), where CC and S-GELS suffer due to unfilled strata with relatively insufficient samples. In the estimation of Banzhaf values, S-GELS is less effective due to inappropriate stratification ratios, resulting in its performance lagging behind GELS. Meanwhile, ASSS-A and ASSS-B exhibit consistent performance advantages in all cases. ASSS-A and ASSS-B show little advantage over GELS in the estimation of Banzhaf values, mainly because the Banzhaf value assigns equal weight

Method	Shapley	Banzhaf	Beta(2,2)-Shapley
GELS	$(6.19 \pm 1.41) \times 10^{-4}$	$(3.09 \pm 0.14) \times 10^{-5}$	$(4.48 \pm 0.39) \times 10^{-4}$
S-GELS	$(2.29 \pm 0.25) \times 10^{-3}$	$(8.08 \pm 0.73) \times 10^{-5}$	$(4.34 \pm 0.85) \times 10^{-4}$
ASSS-A	$(2.08 \pm 0.03) \times 10^{-4}$	$(3.02 \pm 0.20) \times 10^{-5}$	$(3.97 \pm 0.47) \times 10^{-4}$
ASSS-B	$(2.24 \pm 0.02) \times 10^{-4}$	$(3.06 \pm 0.17) \times 10^{-5}$	$(4.16 \pm 0.38) \times 10^{-4}$

Table 7. 95% confidence intervals of RMSE.

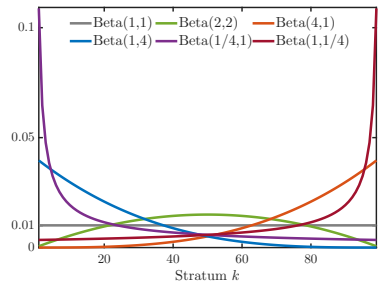


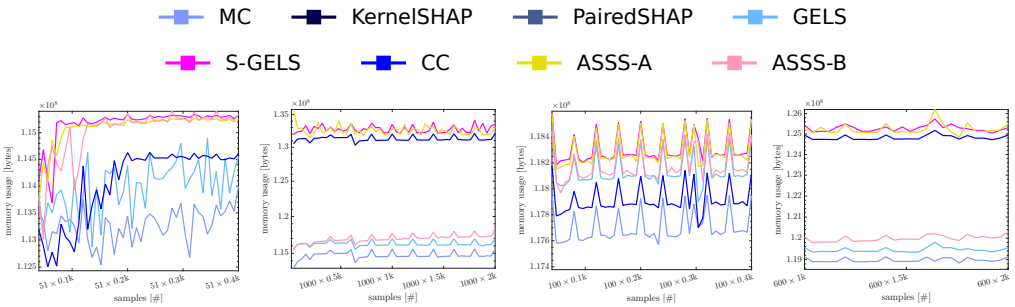
Fig. 9. Beta-Shapley value stratification weights ($n = 100$).

to the marginal contribution of coalitions. This gives coalitions of size around $n/2$ a significant proportion in the calculation. However, since these coalitions are similar in size, their utilities are close to each other, and according to Theorem 5.6, stratified sampling becomes less effective at reducing variance. In contrast, for the voting task, we observe a more noticeable reduction in error with our method, as even when coalition sizes are similar, the utility differences can still be large (between 0 and 1).

Confidence intervals. Table 7 presents the RMSE with 95% confidence intervals (calculated over 20 independent runs) for the Breast Cancer dataset at a fixed sample budget of $m = 600$k. The results confirm that both ASSS-A and ASSS-B significantly outperform GELS and S-GELS, achieving consistently lower errors and narrower confidence intervals. Comparing the two adaptive methods reveals a nuanced trade-off. While ASSS-A achieves a marginally lower mean RMSE (better accuracy), ASSS-B exhibits slightly tighter confidence intervals (better stability). This difference stems from stratification granularity: ASSS-A's fine-grained structure effectively minimizes bias but is more sensitive to stochastic sampling fluctuations across its numerous strata. In contrast, ASSS-B's coarser aggregation mitigates this variability, resulting in statistically more robust estimates.

6.3 Efficiency

We perform an empirical analysis across four tasks on Shapley value, Banzhaf value, and Beta Shapley value estimation, each with varying target RMSE values. Figure 8 illustrates the required time of each algorithm to reach the same target RMSE. For clarity, we omit some of the underperforming baselines in each figure. The experimental results demonstrate that our methods consistently achieve comparable RMSE with less time across all tasks. Notably, although ASSS-B does not always achieve more accurate results than ASSS-A with the time in some tasks, it is more efficient in practice due to its lower resource consumption. In the voting task, ASSS-A slightly underperforms compared to the traditional baseline algorithm MC for Shapley value estimation. This is because utility computation in voting requires minimal time, making simple baseline algorithms like MC highly efficient. However, our algorithms demonstrate a significant advantage in more complex tasks such as data valuation, where utility computation is resource-intensive.



(a) Voting. (b) Airport. (c) Iris. (d) Breast Cancer.

Fig. 10. Shapley value memory usage.

6.4 Scalability

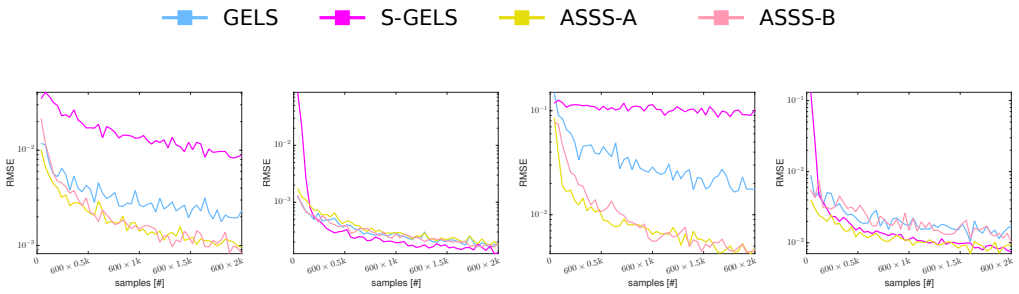
In addition to sample efficacy, we analyze the memory usage of each method across varying problem scales to assess their practical applicability. Figure 10 shows the peak memory usage during Shapley value computation for each method at corresponding sample budgets. The memory usage of most methods is largely insensitive to the sample size, whereas the memory usage of ASSS-B increases slightly as the sample size grows. A distinct pattern emerges when varying the number of players n. When n is small, as shown in Figure 10(a) and 10(c), the memory usage across all methods is comparable. For larger n, as shown in Figure 10(b) and 10(d), a significant gap emerges. CC, ASSS-A, and S-GELS consume more memory than others because they maintain an $O(n^2)$ utility matrix. In contrast, ASSS-B uses less memory due to its more compact $O(Ln)$ structure. Although ASSS-A incurs a higher memory cost, it provides finer-grained stratification and thus lower estimation error. ASSS-B, on the other hand, significantly reduces space complexity, making it the preferable choice for large-scale tasks where memory is the primary bottleneck, especially with n extremely large.

6.5 Robustness

We evaluate ASSS within the Beta-Shapley framework under four extreme parameter settings (a, b): $(4, 1)$, $(1, 4)$, $(1/4, 1)$, and $(1, 1/4)$ to assess the robustness of our proposed algorithms. Their weight distributions are illustrated in Figure 9. Figure 11 presents the experimental results for the Breast Cancer dataset. We observe that the RMSE of GELS and S-GELS is significantly affected by the variation in parameter settings, exhibiting large fluctuations. In contrast, the RMSE of both ASSS-A and ASSS-B remains consistently small, indicating notable robustness. The relative superiority is context-dependent. Specifically, ASSS-A demonstrates a slight advantage in Beta$(1, 1/4)$-Shapley value computation, whereas ASSS-B slightly outperforms ASSS-A in Beta$(1, 4)$-Shapley value computation.

6.6 Summary and Guidelines

In summary, both ASSS-A and ASSS-B consistently deliver superior performance compared to existing methods. Leveraging adaptive stratification, both methods effectively harness the variance-reduction properties of stratified sampling while successfully circumventing the instability caused by empty strata. Specifically, ASSS-A generally exhibits a slight edge in overall estimation accuracy, benefiting from its fine-grained stratification. ASSS-B distinguishes itself through superior stability and offers an advantage in memory efficiency. Based on the comprehensive analysis of effectiveness,



(a) Beta(4, 1). (b) Beta(1, 4). (c) Beta(1/4, 1). (d) Beta(1, 1/4).

Fig. 11. Extreme Beta Shapley values effectiveness.

efficiency, scalability, and robustness, we propose the following practical guidelines for choosing between the two adaptive strategies:

ASSS-A (Prioritize Accuracy). For small to medium-scale problems where memory resources are sufficient, ASSS-A is recommended. Its fine-grained stratification maximizes sample efficiency, thereby achieving lower estimation error with fewer samples. However, practitioners should account for its memory scaling; for instance, in our experimental configuration, ASSS-A incurs an additional memory overhead of approximately $2n^2 \times 10^4$ bytes (e.g., 8×10^7 bytes when $n = 2,000$).

ASSS-B (Prioritize Scalability). For large-scale tasks (large n) where memory footprint is a critical constraint, ASSS-B is the optimal choice. It effectively avoids the memory overflow risks associated with the aforementioned quadratic growth, reducing space complexity while still maintaining competitive accuracy and theoretical guarantees.

These guidelines provide a clear path for practitioners to deploy the appropriate ASSS variant based on specific constraints regarding dataset scale and available hardware resources.

7 Conclusions

In this paper, we adapt and refine stratified sampling for Shapley-like value estimation to tackle the "empty stratum" problem and achieve variance reduction. We propose a novel adaptive stratification strategy to address the fundamental limitations of static stratified sampling when sample sizes are insufficient. By setting appropriate sampling probability distributions and reducing the number of strata, we avoid the potential bias issues that may arise in stratified sampling while preserving the advantage of variance reduction. Additionally, we introduce a more generalized concept of stratification, which not only enhances the current methodology but also lays the groundwork for future research on stratified sampling techniques in Shapley-like value estimation. This paper opens new avenues for exploring innovative approaches in this area. The superiority of our method is validated through both mathematical analysis and empirical experiments. For future work, it would be interesting and valuable to design a sampling approach that optimally combines the number of strata and the sample size within each stratum to achieve maximum variance reduction, particularly when prior information about utility distributions is available.

Acknowledgment

This work was supported in part by NSFC (U23A20306, 62472378, U25A20430), the Zhejiang Provincial Natural Science Foundation for Distinguished Young Scholars (LR25F020001), and the Zhejiang Province Pioneer Plan (2024C01074, 2025C01084).

References

[1] Omer Abramovich, Daniel Deutch, Nave Frost, Ahmet Kara, and Dan Olteanu. 2024. Banzhaf Values for Facts in Query Answering. *SIGMOD* 2, 3 (2024), 123. doi:10.1145/3654926

[2] Anish Agarwal, Munther A. Dahleh, and Tuhin Sarkar. 2019. A marketplace for data: An algorithmic solution. In *Proceedings of the 2019 ACM Conference on Economics and Computation, EC 2019, Phoenix, AZ, USA, June 24-28, 2019*, Anna Karlin, Nicole Immorlica, and Ramesh Johari (Eds.). ACM, 701–726. doi:10.1145/3328526.3329589

[3] Dana Arad, Daniel Deutch, and Nave Frost. 2024. Predicting Fact Contributions from Query Logs with Machine Learning. In *Proceedings 27th International Conference on Extending Database Technology, EDBT 2024, Paestum, Italy, March 25 - March 28*, Letizia Tanca, Qiong Luo, Giuseppe Polese, Loredana Caruccio, Xavier Oriol, and Donatella Firmani (Eds.). OpenProceedings.org, 704–716. doi:10.48786/EDBT.2024.60

[4] Santiago Andrés Azcoitia, Costas Iordanou, and Nikolaos Laoutaris. 2023. Understanding the Price of Data in Commercial Data Marketplaces. In *39th IEEE International Conference on Data Engineering, ICDE 2023, Anaheim, CA, USA, April 3-7, 2023*. IEEE, 3718–3728. doi:10.1109/ICDE55515.2023.00300

[5] John F Banzhaf III. 1964. Weighted voting doesn't work: A mathematical analysis. *Rutgers L. Rev.* 19 (1964), 317.

[6] Leopoldo E. Bertossi, Benny Kimelfeld, Ester Livshits, and Mikaël Monet. 2023. The Shapley Value in Database Management. *SIGMOD Rec.* 52, 2 (2023), 6–17. doi:10.1145/3615952.3615954

[7] Meghyn Bienvenu, Diego Figueira, and Pierre Lafourcade. 2024. When is Shapley Value Computation a Matter of Counting? *PODS* 2, 2 (2024), 105. doi:10.1145/3651606

[8] Javier Castro, Daniel Gómez, Elisenda Molina, and Juan Tejada. 2017. Improving polynomial estimation of the Shapley value by stratified random sampling with optimum allocation. *Comput. Oper. Res.* 82 (2017), 180–188. doi:10.1016/J.COR.2017.01.019

[9] Javier Castro, Daniel Gómez, and Juan Tejada. 2009. Polynomial calculation of the Shapley value based on sampling. *Comput. Oper. Res.* 36, 5 (2009), 1726–1730. doi:10.1016/J.COR.2008.04.004

[10] Lingjiao Chen, Hongyi Wang, Leshang Chen, Paraschos Koutris, and Arun Kumar. 2019. Demonstration of Nimbus: Model-based Pricing for Machine Learning in a Data Marketplace. In *Proceedings of the 2019 International Conference on Management of Data, SIGMOD'19, Amsterdam, The Netherlands, June 30 - July 5, 2019*, Peter A. Boncz, Stefan Manegold, Anastasia Ailamaki, Amol Deshpande, and Tim Kraska (Eds.). ACM, 1885–1888. doi:10.1145/3299869.3320231

[11] Yiwei Chen, Kaiyu Li, Guoliang Li, and Yong Wang. 2024. Contributions Estimation in Federated Learning: A Comprehensive Experimental Evaluation. *Proceedings of the VLDB Endowment* 17, 8 (2024), 2077–2090.

[12] Ian Covert and Su-In Lee. 2021. Improving KernelSHAP: Practical Shapley Value Estimation Using Linear Regression. In *The 24th International Conference on Artificial Intelligence and Statistics, AISTATS 2021, April 13-15, 2021, Virtual Event (Proceedings of Machine Learning Research, Vol. 130)*, Arindam Banerjee and Kenji Fukumizu (Eds.). PMLR, 3457–3465. http://proceedings.mlr.press/v130/covert21a.html

[13] Susan B. Davidson, Daniel Deutch, Nave Frost, Benny Kimelfeld, Omer Koren, and Mikaël Monet. 2022. ShapGraph: An Holistic View of Explanations through Provenance Graphs and Shapley Values. In *SIGMOD '22: International Conference on Management of Data, Philadelphia, PA, USA, June 12 - 17, 2022*, Zachary G. Ives, Angela Bonifati, and Amr El Abbadi (Eds.). ACM, 2373–2376. doi:10.1145/3514221.3520172

[14] Xiaotie Deng and Christos H. Papadimitriou. 1994. On the Complexity of Cooperative Solution Concepts. *Math. Oper. Res.* 19, 2 (1994), 257–266. doi:10.1287/MOOR.19.2.257

[15] Daniel Deutch, Nave Frost, Benny Kimelfeld, and Mikaël Monet. 2022. Computing the Shapley value of facts in query answering. In *Proceedings of the 2022 International Conference on Management of Data*. 1570–1583.

[16] Zhenan Fan, Huang Fang, Xinglu Wang, Zirui Zhou, Jian Pei, Michael Friedlander, and Yong Zhang. 2024. Fair and Efficient Contribution Valuation for Vertical Federated Learning. In *The Twelfth International Conference on Learning Representations*. https://openreview.net/forum?id=sLQb8q0sUi

[17] Zhenan Fan, Huang Fang, Zirui Zhou, Jian Pei, Michael P. Friedlander, Changxin Liu, and Yong Zhang. 2022. Improving Fairness for Data Valuation in Horizontal Federated Learning. In *38th IEEE International Conference on Data Engineering, ICDE 2022, Kuala Lumpur, Malaysia, May 9-12, 2022*. IEEE, 2440–2453. doi:10.1109/ICDE53745.2022.00228

[18] Eitan Farchi, Ramasuri Narayanam, and Lokesh Nagalapatti. 2021. Ranking Data Slices for ML Model Validation: A Shapley Value Approach. In *37th IEEE International Conference on Data Engineering, ICDE 2021, Chania, Greece, April 19-22, 2021*. IEEE, 1937–1942. doi:10.1109/ICDE51399.2021.00180

[19] Raul Castro Fernandez. 2022. Protecting Data Markets from Strategic Buyers. In *SIGMOD '22: International Conference on Management of Data, Philadelphia, PA, USA, June 12 - 17, 2022*, Zachary G. Ives, Angela Bonifati, and Amr El Abbadi (Eds.). ACM, 1755–1769. doi:10.1145/3514221.3517855

[20] R. A. Fisher. 1988. Iris. UCI Machine Learning Repository. DOI: https://doi.org/10.24432/C56C76.

[21] Amirata Ghorbani and James Y. Zou. 2019. Data Shapley: Equitable Valuation of Data for Machine Learning. In *Proceedings of the 36th International Conference on Machine Learning, ICML 2019, 9-15 June 2019, Long Beach, California, USA (Proceedings of Machine Learning Research, Vol. 97)*, Kamalika Chaudhuri and Ruslan Salakhutdinov (Eds.). PMLR,

2242–2251. http://proceedings.mlr.press/v97/ghorbani19c.html

[22] Stefan Grafberger, Shubha Guha, Paul Groth, and Sebastian Schelter. 2023. MLWHATIF: What If You Could Stop Re-Implementing Your Machine Learning Pipeline Analyses Over and Over? *Proc. VLDB Endow.* 16, 12 (2023), 4002–4005. doi:10.14778/3611540.3611606

[23] Ben Halstead, Yun Sing Koh, Patricia Riddle, Mykola Pechenizkiy, Albert Bifet, and Russel Pears. 2021. Fingerprinting Concepts in Data Streams with Supervised and Unsupervised Meta-Information. In *37th IEEE International Conference on Data Engineering, ICDE 2021, Chania, Greece, April 19-22, 2021*. IEEE, 1056–1067. doi:10.1109/ICDE51399.2021.00096

[24] Wassily Hoeffding. 1994. Probability inequalities for sums of bounded random variables. *The collected works of Wassily Hoeffding* (1994), 409–426.

[25] D. Holt and T. M. F. Smith. 1979. Post Stratification. *Journal of the Royal Statistical Society. Series A (General)* 142, 1 (1979), 33–46. http://www.jstor.org/stable/2344652

[26] Ruoxi Jia, David Dao, Boxin Wang, Frances Ann Hubis, Nezihe Merve Gürel, Bo Li, Ce Zhang, Costas J. Spanos, and Dawn Song. 2019. Efficient Task-Specific Data Valuation for Nearest Neighbor Algorithms. *Proc. VLDB Endow.* 12, 11 (2019), 1610–1623. doi:10.14778/3342263.3342637

[27] Kevin Fu Jiang, Weixin Liang, James Y. Zou, and Yongchan Kwon. 2023. OpenDataVal: a Unified Benchmark for Data Valuation. In *Advances in Neural Information Processing Systems 36: Annual Conference on Neural Information Processing Systems 2023, NeurIPS 2023, New Orleans, LA, USA, December 10 - 16, 2023*, Alice Oh, Tristan Naumann, Amir Globerson, Kate Saenko, Moritz Hardt, and Sergey Levine (Eds.). http://papers.nips.cc/paper_files/paper/2023/hash/5b047c7d862059a5df623c1ce2982fca-Abstract-Datasets_and_Benchmarks.html

[28] Ahmet Kara, Dan Olteanu, and Dan Suciu. 2024. From Shapley Value to Model Counting and Back. *PODS* 2, 2 (2024), 79. doi:10.1145/3651142

[29] Pratik Karmakar, Mikaël Monet, Pierre Senellart, and Stéphane Bressan. 2024. Expected Shapley-Like Scores of Boolean functions: Complexity and Applications to Probabilistic Databases. *SIGMOD* 2, 2 (2024), 92. doi:10.1145/3651593

[30] Patrick Kolpaczki, Viktor Bengs, Maximilian Muschalik, and Eyke Hüllermeier. 2024. Approximating the Shapley Value without Marginal Contributions. In *Thirty-Eighth AAAI Conference on Artificial Intelligence, AAAI 2024, Thirty-Sixth Conference on Innovative Applications of Artificial Intelligence, IAAI 2024, Fourteenth Symposium on Educational Advances in Artificial Intelligence, EAAI 2014, February 20-27, 2024, Vancouver, Canada*, Michael J. Wooldridge, Jennifer G. Dy, and Sriraam Natarajan (Eds.). AAAI Press, 13246–13255. doi:10.1609/AAAI.V38I12.29225

[31] Yongchan Kwon, Eric Wu, Kevin Wu, and James Zou. 2024. DataInf: Efficiently Estimating Data Influence in LoRA-tuned LLMs and Diffusion Models. In *The Twelfth International Conference on Learning Representations, ICLR 2024, Vienna, Austria, May 7-11, 2024*. OpenReview.net. https://openreview.net/forum?id=9m02ib92Wz

[32] Yongchan Kwon and James Zou. 2022. Beta Shapley: a Unified and Noise-reduced Data Valuation Framework for Machine Learning. In *International Conference on Artificial Intelligence and Statistics, AISTATS 2022, 28-30 March 2022, Virtual Event (Proceedings of Machine Learning Research, Vol. 151)*, Gustau Camps-Valls, Francisco J. R. Ruiz, and Isabel Valera (Eds.). PMLR, 8780–8802. https://proceedings.mlr.press/v151/kwon22a.html

[33] Yongchan Kwon and James Zou. 2023. Data-OOB: Out-of-bag Estimate as a Simple and Efficient Data Value. In *International Conference on Machine Learning, ICML 2023, 23-29 July 2023, Honolulu, Hawaii, USA (Proceedings of Machine Learning Research, Vol. 202)*, Andreas Krause, Emma Brunskill, Kyunghyun Cho, Barbara Engelhardt, Sivan Sabato, and Jonathan Scarlett (Eds.). PMLR, 18135–18152. https://proceedings.mlr.press/v202/kwon23e.html

[34] Feifei Li. 2023. Modernization of Databases in the Cloud Era: Building Databases that Run Like Legos. *Proc. VLDB Endow.* 16, 12 (2023), 4140–4151. doi:10.14778/3611540.3611639

[35] Jinyang Li, Yuval Moskovitch, and H. V. Jagadish. 2023. Detection of Groups with Biased Representation in Ranking. In *39th IEEE International Conference on Data Engineering, ICDE 2023, Anaheim, CA, USA, April 3-7, 2023*. IEEE, 2167–2179. doi:10.1109/ICDE55515.2023.00168

[36] Weida Li and Yaoliang Yu. 2024. Faster Approximation of Probabilistic and Distributional Values via Least Squares. In *The Twelfth International Conference on Learning Representations, ICLR 2024, Vienna, Austria, May 7-11, 2024*. OpenReview.net. https://openreview.net/forum?id=lvSMIsztka

[37] Yifan Li, Xiaohui Yu, and Nick Koudas. 2021. Data Acquisition for Improving Machine Learning Models. *Proc. VLDB Endow.* 14, 10 (2021), 1832–1844. doi:10.14778/3467861.3467872

[38] Yifan Li, Xiaohui Yu, and Nick Koudas. 2024. Data Acquisition for Improving Model Confidence. *SIGMOD* 2, 3 (2024), 131. doi:10.1145/3654934

[39] Stephen C Littlechild and Guillermo Owen. 1973. A simple expression for the Shapley value in a special case. *Management Science* 20, 3 (1973), 370–372.

[40] Stephen C Littlechild and GF Thompson. 1977. Aircraft landing fees: a game theory approach. *The Bell Journal of Economics* (1977), 186–204.

[41] Jinfei Liu, Qiongqiong Lin, Jiayao Zhang, Kui Ren, Jian Lou, Junxu Liu, Li Xiong, Jian Pei, and Jimeng Sun. 2021. Demonstration of Dealer: An End-to-End Model Marketplace with Differential Privacy. *Proc. VLDB Endow.* 14, 12

(2021), 2747–2750. doi:10.14778/3476311.3476335

[42] Jinfei Liu, Jian Lou, Junxu Liu, Li Xiong, Jian Pei, and Jimeng Sun. 2021. Dealer: An End-to-End Model Marketplace with Differential Privacy. *Proc. VLDB Endow.* 14, 6 (2021), 957–969. doi:10.14778/3447689.3447700

[43] Ester Livshits, Leopoldo E. Bertossi, Benny Kimelfeld, and Moshe Sebag. 2020. The Shapley Value of Tuples in Query Answering. In *23rd International Conference on Database Theory, ICDT 2020, March 30-April 2, 2020, Copenhagen, Denmark (LIPIcs, Vol. 155)*, Carsten Lutz and Jean Christoph Jung (Eds.). Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 20:1–20:19. doi:10.4230/LIPICS.ICDT.2020.20

[44] Scott M. Lundberg and Su-In Lee. 2017. A Unified Approach to Interpreting Model Predictions. In *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA*, Isabelle Guyon, Ulrike von Luxburg, Samy Bengio, Hanna M. Wallach, Rob Fergus, S. V. N. Vishwanathan, and Roman Garnett (Eds.). 4765–4774. https://proceedings.neurips.cc/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html

[45] Xuan Luo and Jian Pei. 2024. Applications and Computation of the Shapley Value in Databases and Machine Learning. In *Companion of the 2024 International Conference on Management of Data, SIGMOD/PODS 2024, Santiago AA, Chile, June 9-15, 2024*, Pablo Barceló, Nayat Sánchez Pi, Alexandra Meliou, and S. Sudarshan (Eds.). ACM, 630–635. doi:10.1145/3626246.3654680

[46] Xuan Luo, Jian Pei, Zicun Cong, and Cheng Xu. 2022. On Shapley Value in Data Assemblage Under Independent Utility. *Proc. VLDB Endow.* 15, 11 (2022), 2761–2773. doi:10.14778/3551793.3551829

[47] Xuan Luo, Jian Pei, Cheng Xu, Wenjie Zhang, and Jianliang Xu. 2024. Fast Shapley Value Computation in Data Assemblage Tasks as Cooperative Simple Games. *SIGMOD* 2, 1 (2024), 56:1–56:28. doi:10.1145/3639311

[48] Shuaicheng Ma, Yang Cao, and Li Xiong. 2021. Transparent Contribution Evaluation for Secure Federated Learning on Blockchain. In *37th IEEE International Conference on Data Engineering Workshops, ICDE Workshops 2021, Chania, Greece, April 19-22, 2021*. IEEE, 88–91. doi:10.1109/ICDEW53142.2021.00023

[49] Irwin Mann and Lloyd S Shapley. 1960. *Values of large games, IV: Evaluating the electoral college by Montecarlo techniques*. Rand Corporation.

[50] Rory Mitchell, Joshua Cooper, Eibe Frank, and Geoffrey Holmes. 2022. Sampling Permutations for Shapley Value Estimation. *J. Mach. Learn. Res.* 23 (2022), 43:1–43:46. http://jmlr.org/papers/v23/21-0439.html

[51] Nikolaos Myrtakis, Ioannis Tsamardinos, and Vassilis Christophides. 2021. PROTEUS: Predictive Explanation of Anomalies. In *37th IEEE International Conference on Data Engineering, ICDE 2021, Chania, Greece, April 19-22, 2021*. IEEE, 1967–1972. doi:10.1109/ICDE51399.2021.00185

[52] Guillermo Owen. 2013. *Game theory*. Emerald Group Publishing.

[53] Romila Pradhan, Aditya Lahiri, Sainyam Galhotra, and Babak Salimi. 2022. Explainable AI: Foundations, Applications, Opportunities for Data Management Research. In *38th IEEE International Conference on Data Engineering, ICDE 2022, Kuala Lumpur, Malaysia, May 9-12, 2022*. IEEE, 3209–3212. doi:10.1109/ICDE53745.2022.00300

[54] Alon Reshef, Benny Kimelfeld, and Ester Livshits. 2020. The Impact of Negation on the Complexity of the Shapley Value in Conjunctive Queries. In *Proceedings of the 39th ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems, PODS 2020, Portland, OR, USA, June 14-19, 2020*, Dan Suciu, Yufei Tao, and Zhewei Wei (Eds.). ACM, 285–297. doi:10.1145/3375395.3387664

[55] Lloyd S. Shapley. 1953. A value for n-person games. (1953).

[56] Qiheng Sun, Xiang Li, Jiayao Zhang, Li Xiong, Weiran Liu, Jinfei Liu, Zhan Qin, and Kui Ren. 2023. ShapleyFL: Robust Federated Learning Based on Shapley Value. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, KDD 2023, Long Beach, CA, USA, August 6-10, 2023*, Ambuj K. Singh, Yizhou Sun, Leman Akoglu, Dimitrios Gunopulos, Xifeng Yan, Ravi Kumar, Fatma Ozcan, and Jieping Ye (Eds.). ACM, 2096–2108. doi:10.1145/3580305.3599500

[57] Guan Wang. 2019. Interpret Federated Learning with Shapley Values. *CoRR* abs/1905.04519 (2019). arXiv:1905.04519 http://arxiv.org/abs/1905.04519

[58] Junhao Wang, Lan Zhang, Anran Li, Xuanke You, and Haoran Cheng. 2022. Efficient Participant Contribution Evaluation for Horizontal and Vertical Federated Learning. In *38th IEEE International Conference on Data Engineering, ICDE 2022, Kuala Lumpur, Malaysia, May 9-12, 2022*. IEEE, 911–923. doi:10.1109/ICDE53745.2022.00073

[59] Jiachen T. Wang and Ruoxi Jia. 2023. Data Banzhaf: A Robust Data Valuation Framework for Machine Learning. In *International Conference on Artificial Intelligence and Statistics, 25-27 April 2023, Palau de Congressos, Valencia, Spain (Proceedings of Machine Learning Research, Vol. 206)*, Francisco J. R. Ruiz, Jennifer G. Dy, and Jan-Willem van de Meent (Eds.). PMLR, 6388–6421. https://proceedings.mlr.press/v206/wang23e.html

[60] Tingting Wang, Shixun Huang, Zhifeng Bao, J. Shane Culpepper, Volkan Dedeoglu, and Reza Arablouei. 2024. Optimizing Data Acquisition to Enhance Machine Learning Performance. *Proc. VLDB Endow.* 17, 6 (2024), 1310–1323. https://www.vldb.org/pvldb/vol17/p1310-bao.pdf

[61] WIlliam Wolberg. 1992. Breast Cancer Wisconsin (Original). UCI Machine Learning Repository. DOI: https://doi.org/10.24432/C5HP4Z.

[62] Haocheng Xia, Xiang Li, Junyuan Pang, Jinfei Liu, Kui Ren, and Li Xiong. 2024. P-Shapley: Shapley Values on Probabilistic Classifiers. *Proc. VLDB Endow.* 17, 7 (2024), 1737–1750. https://www.vldb.org/pvldb/vol17/p1737-liu.pdf

[63] Haocheng Xia, Jinfei Liu, Jian Lou, Zhan Qin, Kui Ren, Yang Cao, and Li Xiong. 2023. Equitable Data Valuation Meets the Right to Be Forgotten in Model Markets. *Proc. VLDB Endow.* 16, 11 (2023), 3349–3362. doi:10.14778/3611479.3611531

[64] Jiayao Zhang, Qiheng Sun, Jinfei Liu, Li Xiong, Jian Pei, and Kui Ren. 2023. Efficient Sampling Approaches to Shapley Value Approximation. *SIGMOD* 1, 1 (2023), 48:1–48:24. doi:10.1145/3588728

[65] Jiayao Zhang, Haocheng Xia, Qiheng Sun, Jinfei Liu, Li Xiong, Jian Pei, and Kui Ren. 2023. Dynamic Shapley Value Computation. In *39th IEEE International Conference on Data Engineering, ICDE 2023, Anaheim, CA, USA, April 3-7, 2023*. IEEE, 639–652. doi:10.1109/ICDE55515.2023.00055

[66] Xinyi Zhang, Zhuo Chang, Yang Li, Hong Wu, Jian Tan, Feifei Li, and Bin Cui. 2022. Facilitating Database Tuning with Hyper-Parameter Optimization: A Comprehensive Experimental Evaluation. *Proc. VLDB Endow.* 15, 9 (2022), 1808–1821. doi:10.14778/3538598.3538604

[67] Shuyuan Zheng, Yang Cao, and Masatoshi Yoshikawa. 2023. Secure Shapley Value for Cross-Silo Federated Learning. *Proc. VLDB Endow.* 16, 7 (2023), 1657–1670. doi:10.14778/3587136.3587141

[68] Yuqing Zhu, Jing Tang, Xueyan Tang, and Lei Chen. 2021. Analysis of Influence Contribution in Social Advertising. *Proc. VLDB Endow.* 15, 2 (2021), 348–360. doi:10.14778/3489496.3489514

Received October 2025; revised January 2026; accepted February 2026