Sketching stochastic valuation functions

Milan Vojnović and Yiliu Wang
Department of Statistics, London School of Economics
{m.vojnovic,y.wang295}@lse.ac.uk

Abstract

We consider the problem of sketching a set valuation function, which is defined as the expectation of a valuation function of independent random item values. We show that for monotone subadditive or submodular valuation functions satisfying a weak homogeneity condition, or certain other conditions, there exist discretized distributions of item values with $O(k \log(k))$ support sizes that yield a sketch valuation function which is a constant-factor approximation, for any value query for a set of items of cardinality less than or equal to $k$. The discretized distributions can be efficiently computed by an algorithm for each item’s value distribution separately. Our results hold under conditions that accommodate a wide range of valuation functions arising in applications, such as the value of a team corresponding to the best performance of a team member, constant elasticity of substitution production functions exhibiting diminishing returns used in economics and consumer theory, and others. Sketch valuation functions are particularly valuable for finding approximate solutions to optimization problems such as best set selection and welfare maximization. They enable computationally efficient evaluation of approximate value oracle queries and provide an approximation guarantee for the underlying optimization problem.

1 Introduction

Evaluating the valuation of a set of items is essential in various applications, including ranking or selecting items in recommender systems, information retrieval, assortment optimization, team selection in online gaming, freelancing and other online platforms. In these scenarios, efficiently computing value query answers is crucial, all while maintaining a certain level of approximation error tolerance. This task poses a challenge due to the uncertainty of individual item’s values and the non-linear relations between the values of input items and the corresponding set value. A solution to this problem is to use compact summaries of individual item value distributions to approximate the expected valuation of a set of items. The primary challenge lies in constructing such summaries and determining how to use them to evaluate a valuation set function that provides a reliable approximation and can be efficiently computed for any queried set of items.

We consider the class of the stochastic valuation functions, which for a ground set of items $\Omega = \{1, \ldots, n\}$, are of the following form

$$u(S) = \mathbb{E}[f(X_S)] \text{ for } S \subseteq \Omega$$

(1)

where $f : \mathbb{R}^n_+ \rightarrow \mathbb{R}$ is some given function and $X_S$ is a $n$-dimensional vector with the $i$-th element equal to $X_i$ for
\( i \in S \) and is equal to a neutral element 0, otherwise. Here \( X_1, \ldots, X_n \) represent individual items’ values, assumed to be independent random variables with distributions \( P_1, \ldots, P_n \), respectively. By convention, we assume that \( P_1, \ldots, P_n \) are cumulative distribution functions, so that for every \( i \in \Omega \) and \( x \in \mathbb{R} \), \( P_i(x) \) is the probability that item \( i \) has value smaller than or equal to \( x \).

Set functions of the form given by (1) have been studied in prior literature under various assumptions on the function \( f \), such as monotone submodular (diminishing returns) functions Asadpour and Nazerzadeh (2016), maximum value, and some other submodular functions Kleinberg and Raghu (2018), submodular functions satisfying an extended diminishing returns condition Sekar et al. (2021); Lee et al. (2023), and maximum value and some other order statistics Mehta et al. (2020). This prior work primarily focuses on optimization problems, such as best set selection subject to cardinality or more general budget constraints, and welfare maximization problems. All these optimization problems require efficient evaluation of value oracle queries, i.e. evaluating the value of \( u(S) \) for given input set \( S \) in response to a value oracle query. Computing the value of a set according to (1) requires multi-dimensional integration to evaluate the expected value, which can be prohibitively expensive in practice. For discrete item value distributions with a support of size \( m \), evaluating (1) for a set of cardinality \( k \) has a computation complexity of \( \Theta(m^k) \), which can be overly expensive even for sets of small cardinality \( k \) for distributions with a large support size \( m \).

We study the problem of approximating a set function \( u \) of the form (1) with a sketch set function \( v \) such that \( v(S) \leq u(S) \leq \alpha v(S) \), for every \( S \in \mathcal{F} \), for some \( \alpha \geq 1 \), given \( \mathcal{F} \subseteq 2^n \). A set function \( v \) satisfying these conditions is said to be an \( \alpha \)-approximate sketch valuation function (or simply an \( \alpha \)-sketch) on \( \mathcal{F} \). If \( \mathcal{F} = 2^n \), then we say that \( v \) is an \( \alpha \)-approximation of \( u \) everywhere. We focus on representing each item’s value distribution \( P_i \) with a summary \( Q_i \) such that the sketch valuation function \( v \) can be evaluated by only having access to summaries \( Q_1, \ldots, Q_n \). It is desired for these summaries to be compact (i.e. having a small representation size) while guaranteeing that the sketch function is an \( \alpha \)-approximation and that it can be efficiently computed. Having an \( \alpha \)-approximate sketch valuation function is useful for different optimization problems. A canonical example is the best set selection problem, which asks to find a set \( S^* \) that maximizes \( u(S) \) over \( S \subseteq \Omega \) subject to the cardinality constraint \( |S| \leq k \). If there exists an algorithm that provides a \( c \)-approximation for the best selection problem with respect to an \( \alpha \)-sketch function \( v \), then using the output of this algorithm provides an \( \alpha c \)-approximation for the original best set selection problem. Another example is a welfare maximization problem that asks to find disjoint sets of items that maximize a welfare function defined as the sum of valuations of sets of items subject to cardinality constraints.

There are several motivating applications for the problem we study. In e-commerce, information retrieval, and recommender systems, items such as shopping products, documents, movies, and other media items are recommended based on their predicted individual relevance scores. Relevance scores, computed using machine learning algorithms based on item features, user interests, and contextual information, are intrinsically uncertain Cohen et al. (2021); Zhu et al. (2009). In digital advertising, ads are selected based on uncertain click-through-rate predictions He et al. (2014). In online gaming, freelancing, and other online platforms, items correspond to players or online workers and item values represent their individual performances or skills in solving particular tasks. Individual performances naturally tend to be random depending on internal state of players or online workers and external
factors. Skill estimates of players are typically uncertain, computed using a statistical inference procedure based on limited historical data, e.g. by using approximate Bayesian inference Herbrich et al. (2006).

In the aforementioned application scenarios, for item values defined as outputs of classification or regression machine learning algorithms, values are uncertain due to aleatoric (data) and epistemic (model) uncertainty Hüllermeier and Waegeman (2021). Point predictors, such as those defined by neural networks, may lead to over-confident predictions Lakshminarayanan et al. (2017). Various approaches have been proposed for estimating uncertainty of prediction models, such as approximate Bayesian methods allowing for drawing samples from a posterior distribution using methods such as MCDropout Gal and Ghahramani (2016), model parameter regularization Blundell et al. (2015), and Bayesian ensembles of decision trees Malinin et al. (2021). The underlying posterior distributions of item values defined as outputs of machine learning algorithms can have large supports, and thus evaluating set functions of the form (1) can be prohibitively expensive. Our work explores the properties of these posterior distributions that are sufficient for approximating the underlying set valuation function.

We address the problem of finding $\alpha$-approximate sketch valuation functions within a computation model where the algorithm has value oracle access to value $P_i(x)$ for every given input $i \in \Omega$ and $x \in \mathbb{R}$. This computation model has been used in the line of work on finding approximate solutions for best set selection or welfare maximization problems using scalar-valued representations of individual items (so called test scores), e.g. Kleinberg and Raghu (2018); Sekar et al. (2021); Lee et al. (2023), as well as vector-valued representations of individual items, e.g. Sekar et al. (2021); Mehta et al. (2020). It is noteworthy that we consider the problem of finding a sketch valuation function using a deterministic algorithm that guarantees an $\alpha$-approximation with probability 1. This is in contrast to using a randomized algorithm, which can only guarantee an approximation guarantee to hold with a certain probability. For example, estimating a stochastic valuation function by the sample average approximation method Kleywegt et al. (2002) can only provide a guarantee that holds with a certain probability. A deterministic $\alpha$-approximation guarantee for value oracle calls and a deterministic approximation algorithm for the underlying discrete optimization problem allow us to find a deterministic approximate solution to the underlying discrete optimization problem.

We consider summaries $Q_1, \ldots, Q_n$ defined as discrete distributions with finite supports. It is important to note that input distributions $P_1, \ldots, P_n$ can have infinite supports, corresponding to discrete distributions on an infinite support, continuous distributions, or mixed discrete/continuous distributions. Even when $P_1, \ldots, P_n$ correspond to discrete distributions with finite but large supports, there may be a desire to use discretized distributions with smaller support sizes. We present an algorithm that independently computes discrete distribution $Q_i$ for each item $i \in \Omega$.

The sketch valuation function $v$ is defined as the expected valuation of a set of items according to function $f$ with distributions of items corresponding to discrete distributions $Q_1, \ldots, Q_n$. We establish approximation guarantees for monotone subadditive or submodular functions under certain conditions, specifically for weakly homogeneous functions $f$, among other function classes. The weakly homogeneous functions satisfy $(1/\eta)f(x) \leq f(\theta x) \leq \theta^d f(x)$ for every $x$ in the domain of $f$ and all $\theta \in [0, 1]$, where $\eta$ is a tolerance parameter and $d$ is a degree parameter. Many valuation functions of interest in practice, such as maximum value, sum of a fixed number of top highest values, and others, are accommodated by these conditions. Various examples of valuation functions meeting these conditions are discussed throughout the paper.
1.1 Summary of contributions

For weakly homogeneous valuation functions $f$ with degree $d$ and tolerance $\eta$, we demonstrate the existence of discretized distributions with support sizes $O((1/d) \log(k))$ which ensure that the sketch function $v$ is $\alpha$-approximate for any set of cardinality at most $k$, where $\alpha$ can be arbitrarily close to $1/(4\eta)$. Consequently, a constant-factor approximation holds everywhere within the underlying class of stochastic valuation functions with $O(n \log(n))$ support size of each item’s discretized value distribution. To our knowledge, this result represents the first instance of a constant-factor sketch for the underlying class of stochastic valuation functions. In contrast, a previously known sketch achieved a $O(\log(k))$ approximation with a $O(k)$ sketch size for submodular stochastic valuation functions satisfying the extended diminishing returns property Sekar et al. (2021).

Several commonly studied valuation functions, including the maximum value and, more generally, the sum of a fixed number of top highest values, are weakly homogeneous with degree $d = 1$ and tolerance $\eta = 1$. Consequently, for these valuation functions, we achieve an $\alpha$-sketch with $\alpha$ arbitrarily close to $1/4$, and each discretized distribution has a support size of $O(k \log(k))$.

We demonstrate similar approximation guarantees for other classes of valuation functions, such as a subset of concave valuation functions (referred to as extendable concave functions), functions that are weakly homogeneous coordinate-wise, and those obtained by using certain univariate transformations of items’ values.

Our approximation results for stochastic valuation functions are shown to hold using a distribution discretization algorithm. This algorithm, for any given distribution $P$ of an item’s values, outputs a discrete distribution $Q$ with a finite support using a value oracle access to $P$ (refer to Figure 1 for an illustration). The algorithm involves two parameters, $\epsilon$ and $a$, controlling the support size of the output distribution $Q$. The support size $s$ of $Q$ satisfies $s = O((1/\epsilon) \log(1/a))$.

Using an approximate value oracle along with an approximation algorithm for a discrete optimization problem yields a computationally less expensive approximation guarantee. Specifically, for weakly homogeneous valuation functions with tolerance $\eta$, we demonstrate a $4/\eta(1 - 1/e)$-approximation for the best set selection problem and an $8\eta$-approximation for the welfare maximization problem. These constant-factor approximations hold with
polynomial-time evaluation of approximate value oracle calls, provided that the cardinality constraint is sufficiently small relative to the total number of items—specifically, \( k = O(\log(n)/\log(\log(n))) \).

Finally, we present numerical results obtained using both randomly generated and real-world data to validate various hypotheses suggested by our theoretical analysis. The results demonstrate that accurate function approximations can be obtained using our proposed method.

### 1.2 Related work

Reference Goemans et al. (2009) was the first to formulate the problem of approximating a submodular function everywhere, i.e. approximating its value for points in the domain. Given a value oracle access to a function \( u \) on a ground set of size \( n \), the goal is to design an algorithm that performs a polynomial number in \( n \) of value queries to the oracle to construct an oracle for a function \( v \) such that, for every set \( S \), \( v(S) \) approximates \( u(S) \) within an approximation factor \( \alpha \). The authors demonstrated that there exists an algorithm that, for any non-negative, monotone, submodular function \( u \), achieves approximation factor \( \alpha = O(\sqrt{n} \log n) \). They also established that no algorithm can achieve a factor better than \( \Omega(\sqrt{n}/\log n) \). The approximation function they proposed is the root-linear function \( v(S) = \sqrt{\sum_{i \in S} c_i} \) for some coefficients \( c_1, \ldots, c_n \) in \( \mathbb{R}_+ \). Balcan and Harvey (2011) showed that for some matroid rank functions, a subclass of submodular set functions, every sketch fails to provide an approximation factor better than \( n^{1/3} \). Furthermore, Badanidiyuru et al. (2012) demonstrated that every subadditive set function \( u \) has an \( \alpha \)-sketch where \( \alpha = O(\sqrt{n} \text{polylog}(n)) \), and there is an algorithm that can achieve this with a polynomial number of demand queries. They also showed that every deterministic algorithm that only has access to a value oracle cannot guarantee a sketching ratio better than \( n^{1-\epsilon} \). The sketches discussed in the references discussed so far used geometric constructions, finding an ellipsoid that approximates well a polymatroid associated with \( u \). In contrast, Cohavi and Dobzinski (2017) showed how to obtain faster and simpler sketches for valuation functions. They introduced an algorithm that finds a \( \tilde{O}(\sqrt{n}) \) sketch of a submodular set function with only \( \tilde{O}(n^{3/2}) \) value queries and another algorithm that finds a \( \tilde{O}(\sqrt{n}) \) sketch of a subadditive function with \( O(n) \) demand and value queries.

The problem of approximating the expected value of a function of independent random variables was studied as early as by Klass (1981). The focus was on approximating the expected value of a function (satisfying certain conditions) of a sum of independent random variables using a function that involves expectations only with respect to univariate marginal distributions.

Reference Asadpour and Nazerzadeh (2016) investigated the problem of maximizing a monotone submodular function, defined as the expected value of a monotone submodular value function, subject to a matroid constraint. In a related vein, Kleinberg and Raghu (2018) explored this problem specifically for cardinality constraints, focusing on the class of test score algorithms that employ a one-dimensional representation of each item’s value distribution. They demonstrated that for a sum of top-order statistics objective function, certain test scores guarantee a constant-factor approximation. Subsequently, within a framework based on sketch functions, Sekar et al. (2021) established the existence of test scores that guarantee a constant-factor approximation for a subset of monotone submodular functions satisfying an extended diminishing returns property. Notably, they identified a \( O(\log n) \)-approximate
sketch function using a $k$-dimensional test score.

To the best of our knowledge, this represents the best-known sketch among monotone stochastic valuation functions that satisfy the extended diminishing returns property. Further extending the framework of test scores for solving the stochastic valuation maximization problem subject to more general budget constraints, Lee et al. (2023) contributed to this field. Additionally, Mehta et al. (2020) presented a Polynomial Time-Approximation Scheme (PTAS) for the stochastic valuation maximization problem based on the maximum value function and subject to a cardinality constraint. They achieved this by representing each item’s distribution with a histogram of size $O(k \log(k))$. Our distribution discretization algorithm employs a strategy similar to that in Mehta et al. (2020), involving truncating tails of the input distribution and using exponential binning for the remaining part. An important distinction is that our algorithm computes the output discrete distribution for each item independently. In contrast, Mehta et al. (2020) requires using the same binning boundaries for the output discrete distributions of all items. Neither Lee et al. (2023) nor Mehta et al. (2020) provided results on sketching for approximating a stochastic valuation function everywhere.

Finally, we highlight the line of work on data summaries, which encompasses sketching various properties of sets, multisets, ordered data, vectors, matrices, and graph data. Interested readers can find a comprehensive overview in the book Cormode and Yi (2020) and the references therein. Our work adds to this line of research by focusing on sketching a collection of distributions with the goal of approximating expected values with respect to arbitrary subsets of these distributions.

1.3 Organization of the paper

In Section 2, we present a formal problem formulation and provide background information along with preliminary results. Section 3 introduces our sketching algorithm and outlines guarantees for set function approximation under various assumptions. The implications of using our sketch functions for approximately solving best set selection and welfare maximization problems are discussed in Section 4. Our numerical results are presented in Section 5. Finally, we offer concluding remarks in Section 6. Proofs can be found in the Appendix.

2 Problem formulation

Let $\Omega = \{1, \ldots, n\}$ be a ground set of items. Each item $i \in \Omega$ has a value according to the random variable $X_i$ with distribution $P_i$. The variables $X_1, \ldots, X_n$ are assumed to be independent. The value of a set of items $S \subseteq \Omega$ corresponds to the value $u(S)$, where $u$ is a set valuation function defined as

$$u(S) = E[f((X_i, i \in S))]$$

where $f: \mathbb{R}^n \rightarrow \mathbb{R}^+$ is a monotone function. Hereinafter, for any set $S \subseteq \Omega$ and $x_1, \ldots, x_n$, we use the notation $(x_i, i \in S)$ to denote $z \in \mathbb{R}^n$ such that $z_i = x_i$ if $i \in S$ and $z_i = 0$ otherwise.

We compute $Q_1, \ldots, Q_n$ as representations of $P_1, \ldots, P_n$ corresponding to the distributions of some discrete random variables $Y_1, \ldots, Y_n$. We define the sketch set function $v$ as the expected value of function $f$ with respect to item value distributions $Q_1, \ldots, Q_n$. The sketch function $v$ should approximate the set valuation function $u$ within
a multiplicative approximation error tolerance on some given \( F \subseteq 2^{\Omega} \), i.e., for some \( \alpha \geq 1 \),

\[
v(S) \leq u(S) \leq \alpha v(S), \quad \text{for every } S \in F.
\]

When this guarantee holds we say that \( v \) is an \( \alpha \)-approximation of \( u \), or that \( v \) is an \( \alpha \)-sketch of \( u \), on \( F \). If \( F = 2^{\Omega} \), then \( v \) is an \( \alpha \)-approximation of \( u \) everywhere. For a positive integer \( k \), we let \( F_k \) be defined as \( F_k = \{ S \subseteq \Omega : |S| \leq k \} \).

The problem we consider is to find an algorithm that outputs an \( \alpha \)-sketch valuation function on \( F_k \), for any given \( k \in [n] \). This is considered in the computation model under which the algorithm has access to a value oracle that outputs value of \( P_i(x) \) for any input \( x \in \mathbb{R} \) and \( i \in \Omega \). Preferably, a good sketch has a small representation size, and the approximation factor \( \alpha \) should not depend on \( k \). We aim to address the challenge of finding good sketches for a wide range of function classes, defined as the expected value with respect to discretized item value distributions.

**Function classes and their properties** In what follows, we present definitions for various valuation function classes and some of their properties. These definitions and properties are used in the paper.

A function \( f \) is monotone if \( f(x) \leq f(y) \) for every \( x, y \) in the domain of \( f \) such that \( x \leq y \), where the last relation holds coordinate-wise.

Submodular functions are defined on subsets of \( \mathbb{R}^n \), \( \mathcal{X} = \mathcal{X}_1 \times \cdots \times \mathcal{X}_n \), where each \( \mathcal{X}_i \) is a compact subset of \( \mathbb{R} \). A function \( f \) is submodular if for every \( x, y \in \mathcal{X} \),

\[
f(x \wedge y) + f(x \vee y) \leq f(x) + f(y)
\]

where \( \wedge \) and \( \vee \) denote the coordinate-wise minimum and maximum operations, respectively. If \( \mathcal{X} = \{0, 1\} \), then \( f \) is a set function and the definition corresponds to the standard notion of a submodular set function. By Topkis (1978), if \( f \) is twice-differentiable on its domain, then \( f \) is submodular if, and only if, all off-diagonal elements of the Hessian matrix of \( f \) are nonpositive, i.e. \( \partial^2 f(x)/\partial x_i \partial x_j \leq 0 \), for all \( i \neq j \), for every \( x \) in the domain of \( f \). Submodular functions may be concave, convex, or neither.

A function \( f \) is said to be convex on \( \mathcal{X} \) if

\[
f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y)
\]

for all \( x, y \in \mathcal{X} \) and \( \lambda \in [0, 1] \). Concavity requires the inequality to hold in the reverse direction.

A function \( f \) is said to satisfy the weak DR (diminishing returns) property if, for every \( x, y \in \mathcal{X} \) such that \( x \leq y \), \( i \in [n] \) such that \( x_i = y_i \), and \( z \in \mathbb{R}_+ \) such that \( x + ze_i \in \mathcal{X} \) and \( y + ze_i \in \mathcal{X} \), the following holds:

\[
f(x + ze_i) - f(x) \geq f(y + ze_i) - f(y)
\]

where \( e_i \) is a standard basis vector.

By Bian et al. (2017), a function \( f \) is submodular if, and only if, it satisfies the weak DR property.

A subclass of submodular functions is DR (diminishing returns)-submodular functions Bian et al. (2017); Soma and Yoshida (2015). A function \( f \) is said to be DR-submodular if, for all \( x, y \in \mathcal{X} \) such that \( x \leq y \) and any \( e_i \) and a non-negative number \( z \) such that \( x + ze_i \in \mathcal{X} \) and \( y + ze_i \in \mathcal{X} \), the diminishing returns property (4) holds.
A function $f$ is said to be coordinate-wise concave if, for every $x \in \mathcal{X}$, $i \in [n]$, and $u, v \in \mathbb{R}_+$ such that $x + ue_i \in \mathcal{X}$, $x + ve_i \in \mathcal{X}$, and $x + (u + v)e_i \in \mathcal{X}$, it holds

$$f(x + ue_i) - f(x) \geq f(x + (u + v)e_i) - f(x + ve_i). \tag{5}$$

If $f$ is a twice-differentiable function, then the coordinate-wise concave property is equivalent to $\frac{\partial^2 f(x)}{\partial x_i^2} \leq 0$, for all $x \in \mathcal{X}$ and $i \in [n]$ Bian et al. (2017). Therefore, if $f$ is twice-differentiable, the coordinate-wise concave property corresponds to the standard notion of concave functions holding for each coordinate.

By Bian et al. (2017), a function $f$ is DR-submodular if, and only if, it is submodular and coordinate-wise concave.

The following lemma establishes a known relation between a function $f$ and the set function $u$ defined as $u(S) = \mathbb{E}[f((X_i, i \in S))]$, where $X_1, \ldots, X_n$ are some random variables.

**Lemma 2.1** (Lemma 3 Asadpour and Nazerzadeh (2016)). Assuming that $f$ is a monotone submodular function, it follows that $u$ is a monotone submodular set function.

A function $f$ is said to be subadditive if $f(x + y) \leq f(x) + f(y)$ for all $x$ and $y$ in the domain of $f$. A set function $u$ is said to be subadditive if $u(S \cup T) \leq u(S) + u(T)$, for every $S, T \subseteq \Omega$. It is evident that any non-negative submodular set function is subadditive.

The following lemma, which is utilized in our analysis, establishes a relationship between $f$ and $u$:

**Lemma 2.2.** Assume that $f$ is a monotone function that is either subadditive or submodular, then $u$ is a monotone subadditive set function.

Proof is provided in Appendix A.2.

Additionally, we introduce two noteworthy facts:

**Lemma 2.3.** If $f$ is a DR-submodular function on $\mathcal{X} \subseteq \mathbb{R}_+^n$, $0 \in \mathcal{X}$, and $f(0) \geq 0$, then $f$ is subadditive on $\mathcal{X}$.

Proof is provided in Appendix A.3.

**Lemma 2.4.** If $f$ is a monotone and subadditive function, with $\mathcal{X} \subseteq \mathbb{R}_+^n$, then for every $\lambda \in (0, 1]$, and $x \in \mathcal{X}$,

$$f(x) \leq \lceil 1/\lambda \rceil f(\lambda x). \tag{6}$$

Proof is provided in Appendix A.4. Note that Lemma 2.4 implies $\lambda f(x) \leq f(\lambda x)$ for every $x \in \mathcal{X}$ and $\lambda \in (0, 1]$, which may be interpreted as a weak homogeneity condition.

### 3 Algorithm and approximation guarantees

In this section, we first present an algorithm for representing an item’s value distribution with a discrete distribution having a finite support. Subsequently, we provide results on the approximation guarantees of sketch valuation functions defined by the expectation with respect to these discrete distributions.
3.1 Distribution discretization algorithm

3.1.1 Algorithm

We present a distribution discretization algorithm that, with value oracle access to distribution $P$, outputs a discrete distribution $Q$ with a finite support. The algorithm is defined in Algorithm 1. For simplicity of presentation, the algorithm specifies the output distribution $Q$ at a set of points covering all its jump points, with an understanding that $Q$ is constant between jump points and is right-continuous.

The algorithm employs a binning that assigns all values smaller than a lower-end threshold value to zero, all values larger than an upper-end threshold value to a specific value, and has bins of exponentially increasing widths covering values between the lower-end and upper-end thresholds. Specifically, the algorithm uses two parameters $a \in (0, 1)$ and $\epsilon \in (0, 1]$ to determine the discretization bins. The lower-end threshold is set to $a\tau$ and the upper-end threshold is set to $\tau$, where $\tau$ is the $(1 - \epsilon)$-quantile of distribution $P$. The bins of exponentially increasing widths start at the lower-end threshold $a\tau$ and have increasing width for a factor $1/(1 - \epsilon)$. The smallest number of such bins to cover the interval $[a\tau, \tau]$ is the smallest integer $J$ such that $a\tau/(1 - \epsilon)^J \geq \tau$. The boundary values of bins are
\[ x_j = a \tau / (1 - \epsilon)^{-1}, \text{ for } j = 1, \ldots, J \text{ and } x_{J+1} = \tau. \]

The distribution \( Q \) is defined by relocating all the mass of \( P \) on \([0, a \tau]\) to the value 0, shifting all the mass of \( P \) on \((\tau, \infty)\) to the value \( f^{-1}(H) \) where \( H = \mathbb{E}[f(X) \mid X > \tau] \) whenever \( P(\tau) < 1 \), and transferring all the mass of \( P \) on \((x_j, x_{j+1}]\) to value \( x_j \), for \( j = 1, \ldots, J \). Refer to Figure 2 for an illustration.

The output distribution \( Q \) has its support contained in
\[
Q = \{0\} \cup \left\{ a \tau, \frac{1}{1 - \epsilon} a \tau, \ldots, \frac{1}{(1 - \epsilon)^{J-1}} a \tau \right\} \cup Q^* \]
where \( Q^* = \{f^{-1}(H)\} \) when \( P(\tau) < 1 \) and \( Q^* = \emptyset \) otherwise.

It is noteworthy that the support size \( s \) of \( Q \) is such that
\[
s = O\left(\frac{1}{\epsilon} \log(1/a)\right). \]

3.1.2 Comparison of the input and the output distributions

We present some comparison properties for distributions \( P \) and \( Q \), where \( Q \) is the output of Algorithm 1 for the input distribution \( P \) in the following lemma.

**Lemma 3.1.** For two distributions \( P \) and \( Q \), where \( Q \) is the output of Algorithm 1 for the input distribution \( P \), the following properties hold:

(i) \( Q(x) \geq P(x) - \epsilon \) for all \( x \),

(ii) \( Q(x) \geq P(x) \) for all \( x \leq \tau \), and

(iii) \( Q(x) \leq P(x) + \epsilon \), for all \( x \geq \tau \).

The proof is provided in Appendix A.6. The proof follows from the quantization transformation steps performed by Algorithm 1. Note that the properties asserted in Lemma 3.1 depend only on parameter \( \epsilon \) and \((1 - \epsilon)\)-quantile \( \tau \) of distribution \( P \), and not on parameter \( a \).
As an example, we consider the case when \( P \) is the Pareto distribution, i.e. \( P(x) = 1 - (x_m/x)\beta \) for \( x \geq x_m \) and \( P(x) = 0 \) for \( x < x_m \), where \( x_m > 0 \) and \( \beta > 1 \). The smaller the value of \( \beta \), the heavier the tail of the complementary cumulative distribution function \( 1 - P(x) \). Note that \( \beta > 1 \) is a necessary and sufficient condition for \( \mathbb{E}[X] \) to be finite. It can be readily shown that \( \tau = x_m (1/e)^{1/\beta} \) and \( \mathbb{E}[X \mid X > \tau] = (\beta/(\beta - 1))\tau \). Note that \( \mathbb{E}[X \mid X > \tau] \) can be larger than \( \tau \) for an arbitrarily large factor by taking a small enough \( \beta > 1 \).

**Lemma 3.2.** Assume that \( P \) is the Pareto distribution with \( x_m > 0 \) and \( \beta > 1 \), and \( f \) is such that \( f(x,0,\ldots,0) = x \) for \( x \in \mathbb{R}_+ \). Then, the following properties hold:

(i) For all \( x \), \( Q(x) - P(x) \geq -(1 - (1/\beta)\beta)e \), which is increasing in \( \beta \) on \( (1,\infty) \) from \( -e \) to \( -(1 - 1/e)e \).

(ii) For all \( x \geq \tau \), \( Q(x) - P(x) \leq (1 - 1/\beta)\beta e \), which is increasing in \( \beta \) on \( (1,\infty) \), from \( 0 \) to \( (1/e)e \).

(iii) For all \( x_m \leq x < \tau \), \( Q(x) - P(x) \leq \min\{e/a^\beta,1\} \).

(iv) \( Q(0) = \max\{1 - e/a^\beta,0\} \). Note that \( Q(0) > 0 \) if, and only if, \( \alpha \tau > x_m \), i.e. \( \epsilon < a^\beta \).

### 3.1.3 Discussion of computation aspects

Algorithm 1 produces the discrete distribution \( Q \) for input distribution \( P \) by computing the following properties of distribution \( P \):

(P1) Values of distribution \( P \) at \( O((1/e) \log(1/a)) \) points,

(P2) Conditional expected value \( \mathbb{E}[f(X) \mid X > \tau] \), if \( P(\tau) < 1 \), and

(P3) \((1 - \varepsilon)\)-quantile value \( \tau \) of \( P \).

For a discrete distribution that has a support of size \( m \), the evaluation of \( P(x) \) for every \( x \) can be done in \( O(m) \) time. Similarly, for given \( \tau \), computing \( \mathbb{E}[f(X) \mid X > \tau] \) can be achieved in \( O(m) \) time. Computing the quantile value \( \tau \) is a selection problem. To compute a \( \phi \)-quantile, for some \( \phi \in (0,1) \), we may sort all elements in increasing order, let the \( i \)-th smallest element have rank \( i \), and then return the element of rank \( \lceil \phi m \rceil \). This has time complexity of sorting \( m \) elements, which is \( O(m \log(m)) \). Faster, linear-time deterministic algorithms, are known for selection problem Blum et al. (1973); Schönhage et al. (1976).

It is noteworthy that the above computations need to be performed only once for each item. Evaluating \( u(S) \) directly for any set \( S \) of cardinality at most \( k \) can be computed in \( O(m^k) \) time, which in practice can be prohibitively expensive for large values of \( m \) even for small values of \( k \).

Properties (P1), (P2) and (P3) are amenable to distributed computation when \( P \) is the distribution of points in a multiset of values partitioned across nodes in a distributed computing system. Computing (P1) and (P2) involves resolving simple count aggregate queries. Computing (P3) corresponds to the distributed selection problem. Properties (P1), (P2) and (P3) can be estimated by drawing samples from distribution \( P \) or computed in a stream computing setting. Further discussion on this is provided in Appendix A.5.
Table 1: Properties of some functions $f$.

| $f(x)$ | subadditive | submodular | convex | concave | $d$ | $\eta$ |
|--------|-------------|------------|--------|--------|-----|-------|
| $\max\{x_1, \ldots, x_n\}$ | ✓ | ✓ | ✓ | ✓ | 1 | 1 |
| $x_1 + \cdots + x_k$ | ✓ | ✓ | ✓ | ✓ | 1 | 1 |
| $(\sum_{i=1}^n x_i^r)^{1/r}, r \geq 1$ | ✓ | ✓ | ✓ | ✓ | 1 | 1 |
| $g(\sum_{i=1}^n x_i)$, concave $g$ | ✓ | ✓ | ✓ | ✓ | min elasticity of $g$ | 1 |
| $1 - \prod_{i=1}^n (1 - x_i)$ | ✓ | ✓ | ✓ | ✓ | $\leq 1/2$, for $n \geq 2$ | 1 |

* $x_{(i)}$ denotes the $i$-th element of a sequence corresponding to values $x_1, \ldots, x_n$ sorted in decreasing order.

3.2 Approximation guarantees for stochastic set valuation functions

We present approximation guarantees for approximating the set function $u(S) = \mathbb{E}[f((X_i, i \in S))]$ with the sketch set function $v(S) = \mathbb{E}[f((Y_i, i \in S))]$ where $X_1, \ldots, X_n$ are independent random variables with distributions $P_1, \ldots, P_n$ and $Y_1, \ldots, Y_n$ are independent random variables with distributions $Q_1, \ldots, Q_n$, respectively. Here each $Q_i$ is the output of Algorithm 1 for the input distribution $P_i$.

We assume that distributions $P_1, \ldots, P_n$ have all atoms (if any) of mass at most $\Delta \in [0, 1]$, i.e. for all $i \in \Omega$,

$$P_i(x) - \lim_{z \uparrow x} P_i(z) \leq \Delta, \text{ for all } x \in \mathbb{R}. \quad (7)$$

In fact, it suffices for $(7)$ to hold only for $x = \tau_i$ where $\tau_i$ is the $(1 - \epsilon)$-quantile of $P_i$. If, for each $i \in \Omega$, $P_i$ is continuous and strictly increasing on its support, then $\Delta = 0$.

3.2.1 Weakly homogeneous functions

We will show approximation guarantees for the class of functions that satisfy a weak homogeneous condition. Recall that a function $f$ is said to be homogeneous of degree $d$ over some set $\Theta \subseteq \mathbb{R}$, if $f(\theta x) = \theta^d f(x)$ for all $x$ in the domain of $f$, and all $\theta \in \Theta$. We consider the following relaxed notion of function homogeneity.

**Definition 3.1.** A function $f$ is said to be weakly homogeneous of degree $d$ and tolerance $\eta$ over a set $\Theta \subseteq \mathbb{R}$ if $(1/\eta) \theta f(x) \leq f(\theta x) \leq \theta^d f(x)$, for every $x$ in the domain of $f$ and all $\theta \in \Theta$.

Many functions are weakly homogeneous with a positive degree and tolerance equal to 1. In Table 1 we show several examples of weakly homogeneous functions $f$. In the table, elasticity of a differentiable function $g : \mathbb{R} \rightarrow \mathbb{R}$ at a point $z$ is defined as $zg'(z)/g(z)$. We provide further discussion on the class of weakly homogeneous functions later in this section.

**Theorem 3.3.** Assume that $f$ is a monotone subadditive or submodular function, and is weakly homogeneous with degree $d$ and tolerance $\eta$ over $[0, 1]$, and $\epsilon \in (\Delta, 1)$. Then, for every set $S \subseteq \Omega$ such that $|S| \leq k$, we have

$$\frac{1}{2} (1 - \epsilon)^{k-1} (1 - \Delta/\epsilon) v(S) \leq u(S) \leq 2\eta \frac{1 + adk/\epsilon - \Delta/\epsilon}{(1 - \epsilon)^k (1 - \Delta/\epsilon)^2} v(S).$$

Proof of Theorem 3.3 is provided in Appendix A.8. Here we only discuss the main steps of the proof. Recall that Algorithm 1 performs three transformations: limiting the upper-end of the support, limiting the lower-end of the
support, and exponential binning of the middle part. The proof goes through this sequence of transformations and establishes the effect on the approximation factors by each of them.

At different steps in the proof, we leverage properties of the class of valuation functions that we consider to derive approximation bounds. The middle part is easiest to address. By the definition of exponential binning, the range of each bin is such that each point in a bin is mapped to a point that is at least \((1 - \epsilon)\) factor of the original point. Hence, this results in at most factor \((1 - \epsilon)\) loss of each item’s original value. The proof arguments for the upper and lower ends are more delicate. For the upper-end, each item’s value is split into two parts based on whether or not it exceeds the upper-end threshold \(\tau_i\) and we analyze the contribution of each. For the lower-end part, we split each item’s value into two parts based on whether its value exceeds the lower-end threshold \(a\tau_i\) and evaluate the contribution of each.

The approximation factors in Theorem 3.3 depend on the weak homogeneity degree \(d\) and tolerance \(\eta\) parameters, which specify a subset of monotone functions that are either subadditive or submodular, to which the theorem applies. The approximation factors also depend on the parameters of the distribution discretization algorithm, namely \(a\) and \(\epsilon\), as well as on the set cardinality \(k\), and property \(\Delta\) of item value distributions.

It is noteworthy that the lower bound in Theorem 3.3 holds for any monotone function \(f\) that is either subadditive or submodular. Hence, the approximation factor for the lower bound depends only on \(\epsilon\), \(k\), and \(\Delta\), and not on the weak homogeneity parameters \(d\) and \(\eta\). This approximation factor originates solely from bounding the upper ends of item value distributions. The transformations performed for discretization of the lower and middle parts of an item’s value distribution are monotonic, making discretized distribution stochastically smaller than the original distribution for values up to the upper-end threshold.

In Theorem 3.3, the factors \((1/2)(1 - \epsilon)^{k-1}\) and \(2/(1 - \epsilon)^{k-1}\), appearing in the lower and upper bounds, respectively, come from transformations of the upper ends of item value distributions. The factor \((1 + a^d k - \Delta/\epsilon)/(1 - \Delta/\epsilon)\) comes from the transformation of lower ends of item value distributions, and factor \(\eta/(1 - \Delta/\epsilon)\) comes from the transformation of middle parts of item value distributions. These two parts only affect the upper bound. Note that the approximation factor in the upper bound increases in both \(d\) and \(\eta\).

Theorem 3.3 implies a constant-factor approximation guarantee for any valuation function \(f\) that is weakly homogeneous with strictly positive degree and constant tolerance \(\eta\), under condition, \(\Delta = o(1/k)\), by appropriately choosing input parameters \(a\) and \(\epsilon\) of Algorithm 1.

**Corollary 3.1.** Assume that \(d > 0\) and \(\Delta k < 1\). Under same conditions as in Theorem 3.3 and taking \(a = \epsilon^{2/d}\) and \(\epsilon = c/k\), for some constant \(c \in (\Delta k, 1)\), for every set \(S \subseteq \Omega\) such that \(|S| \leq k\),

\[
\frac{1}{2} e^{-\frac{1}{\Delta k/c}} (1 - \Delta k/c) v(S) \leq u(S) \leq 2\eta e^{-\frac{1}{\Delta k/c}} \frac{1 + c}{(1 - \Delta k/c)^2} v(S).
\]

The proof of Corollary 3.1 is provided in Appendix A.9.

Under the conditions of Corollary 3.1, the discretization algorithm yields an \(\alpha\)-sketch on \(\mathcal{F}_k\) with

\[
\alpha = 4\eta \frac{1 + c}{(1 - \Delta k/c)^3} \epsilon^{2/d}.
\]
Figure 3: Approximation factor \((1 + c)e^{2c/(1-c)}\) versus \(c\).

Note that when \(\Delta = 0\), \(\alpha\) can be made arbitrarily close to \(4\eta\), by taking \(c\) small enough. See Figure 3 for a graph of function \(c \mapsto (1 + c)e^{2c/(1-c)}\).

If \(\Delta = o(1/k)\), each discretized distribution has a support of size at most \(s\), where

\[
s = O\left(\frac{1}{d}k \log(k)\right).
\]

If the degree parameter \(d\) is lower bounded by a positive constant, we have \(s = O(k \log(k))\).

**Approximation results for distributions with arbitrary point masses** The result in Theorem 3.3 is under the assumption that, for every item value distribution, every atom (if any) has a mass of value at most \(\Delta\) and \(\Delta < \epsilon < 1\). Corollary 3.1 holds under the same assumption on the item value distributions and further assumes \(\Delta < c/k\) for some positive constant \(c\). Furthermore, a constant-factor approximation holds when \(\Delta = o(1/k)\). These results thus hold only when \(\Delta\) is small enough relative to either \(\epsilon\) or \(1/k\). This restriction can be alleviated by redefining the sketch valuation function as follows.

For any given \(\Delta \in (0, 1)\), we can represent each item value distribution by a mixture of distributions, where one of them has every atom (if any) with mass at most \(\Delta\) and all others are point mass distributions, and there are at most \(1/\Delta\) of them. This allows us to express the stochastic valuation function as a weighted sum of stochastic valuation functions with the expectation over the distributions of mixture components. It suffices to discretize only mixture component distributions that are not point masses by using Algorithm 1. The redefined sketch valuation function is defined by approximating each stochastic valuation function in the weighted sum by a sketch valuation function. This allows us to extend the results in Theorem 3.3 and Corollary 3.1 and other approximation results that follow to item value distributions with arbitrary jump sizes. The above-described approach increases the representation size of an item for at most \(1/\Delta\) elements (defining point mass distributions). The computation complexity of the sketch valuation function is upper bounded by the product of \((1/\Delta + 1)^k\) and \(s^k\) where \(s\) is an upper bound on the support size of every discretized distribution. We provide a more detailed discussion in Appendix A.10.

**Discussion of weakly homogeneous functions** Clearly, any homogeneous function \(f\) of degree 1 over \(\Theta\) is weekly homogeneous of degree 1 and tolerance \(\eta = 1\) over \(\Theta\). For example, \(f(x) = \max\{x_1, \ldots, x_n\}\) and \(f(x) = (\sum_{i=1}^n x_i^r)^{1/r}\) are homogeneous functions of degree 1 over \(\mathbb{R}\). Note that any function that is convex on a
domain that includes 0 and is such that \( f(0) \leq 0 \) is weakly homogeneous of degree 1 over \([0, 1]\). Some concave functions are weakly homogeneous with a strictly positive degree. For example, \( f(x) = (\sum_{i=1}^n x_i)^r \) with domain \( \mathbb{R}_+^n \), for \( r \in (0, 1] \), is weakly homogeneous of degree \( r \) over \( \mathbb{R}_+ \). A differentiable function \( f \) is weakly homogeneous of degree \( d \) over \([0, 1]\) if, and only if,

\[
x^T \nabla f(x) \geq d f(x) \text{ for every } x \in \text{dom}(f).
\]  

(8)

For example, consider \( f(x) = g(\sum_{i=1}^n x_i) \) where \( g \) is an increasing, differentiable and concave function on \( \mathbb{R}_+ \). Then, the second inequality in (8) is equivalent to \( \eta(z) \geq d \) for all \( z \in \mathbb{R}_+ \), where \( \eta(z) \) is the elasticity of function \( g \), defined as \( \eta(z) = z g'(z)/g(z) \), which is always less than or equal to 1 for any increasing, differentiable and concave function \( g \). Function \( g \) has a constant elasticity \( r \) if, and only if, \( g(z) = cz^r \) for an arbitrary constant \( c > 0 \).

Some concave functions have zero minimum elasticity, e.g. \( g(z) = 1 - e^{-\lambda z} \), for parameter \( \lambda > 0 \), has decreasing elasticity from value 1 at \( z = 0 \) to value 0 as \( z \) goes to infinity.

Many functions are weakly homogeneous over \([0, 1]\) with a constant tolerance parameter \( \eta \), which we discuss next.

Any monotone subadditive function \( f : \mathbb{R}^n \rightarrow \mathbb{R}_+ \) is weakly homogeneous over \([0, 1]\) with tolerance \( \eta = 2 \). To see this, note that by Lemma 2.4, for any monotone subadditive function \( f(\theta x) \geq (1/[1/\theta]) f(x) \), which combined with the fact \( 1/[1/\theta] \geq 1/(1/\theta + 1) \geq \theta / 2 \), implies \( (1/2) \theta f(x) \leq f(\theta x). \)

Any function \( f \) that is subadditive and convex on a domain that includes 0, and is such that \( f(0) = 0 \), is weakly homogeneous over \([0, 1]\) with tolerance \( \eta = 1 \). If \( f \) is a subadditive and convex function on a domain that includes 0 and \( f(0) \leq 0 \), then it is weakly homogeneous with tolerance \( \eta = 1 \). This follows from

\[
f(\theta x) \geq f(x) - f((1 - \theta)x) \\
\geq f(x) - (1 - \theta) f(x) \\
= \theta f(x)
\]

where the first inequality is by subadditivity and the second inequality is by convexity.

Finally, any concave function on a domain that includes 0 such that \( f(0) \geq 0 \) is weakly homogeneous with tolerance \( \eta = 1 \). This follows straightforwardly from the definition of concave functions.

**The importance of the degree of homogeneity**  In general, the dependence of the approximation factor on the degree parameter \( d \) is unavoidable due to assigning values smaller than \( \alpha \tau_i \) to 0 for item \( i \in \Omega \). This can cause an excessive loss of the approximation accuracy for functions with a small degree of weak homogeneity, especially when distributions of item values have sufficient mass near zero. We demonstrate this by the following example.

Let \( f(x) = x^r \) on the domain \( \mathbb{R}_+ \), for a parameter \( r \in (0, 1] \). Let \( X \) be a random variable with cumulative distribution \( P \) with support in \( \mathbb{R}_+ \), and let \( P(r) = 1 - e^{-r} \). By truncating the lower tail of distribution \( P \), we would like \( \mathbb{E}[X^r \mathbf{1}_{\{X \geq \alpha \tau\}}] \) to be a good approximation of \( \mathbb{E}[X^r] \). For any \( a \in [0, 1] \), we can write

\[
\mathbb{E}[(X^r \mathbf{1}_{\{X \geq \alpha \tau\}})] = \rho \mathbb{E}[X^r]
\]

where

\[
\rho = \frac{\mathbb{E}[X^r \mathbf{1}_{\{X \geq \alpha \tau\}}] \mathbb{E}[X^r]}{\mathbb{E}[X^r]} = \frac{\int_{\alpha \tau}^{\infty} x^r dP(x)}{\int_0^{\infty} x^r dP(x)}
\]

15
Obviously, \( \rho \in [0, 1] \). We would like \( \rho \) to be as close to 1 as possible. However, there exist instances for which \( \rho \) can be arbitrarily close to 0 as demonstrated by the following example.

Consider an instance where \( P(x) = x^d \) for \( x \in [0, 1] \), for a parameter \( d > 0 \). Note that \( \tau^d = 1 - \epsilon \). By simple calculus, we have

\[
\rho = 1 - (a(1 - \epsilon)^{1/d})^{r+d}.
\]

Assuming \( a = \epsilon^c \) for a fixed constant \( c > 0 \), and \( r = d = \epsilon \), we have

\[
\rho = 1 - \epsilon^{2\epsilon c}(1 - \epsilon)^{2\epsilon} \downarrow 0 \quad \text{as} \quad \epsilon \downarrow 0.
\]

This shows that for the given family of instances, \( \mathbb{E}[(X \mathbb{1}_{(X \geq \alpha)})^r] \) can be an arbitrarily bad approximation of \( \mathbb{E}[X^r] \) when \( \epsilon \) is small enough.

### 3.2.2 Extendable concave functions

The weak homogeneity condition in Corollary 3.1 restricts the approximation guarantee to functions with a strictly positive degree of homogeneity. Here we show that approximation guarantees can be provided for some valuation functions that have a zero degree of homogeneity.

**Definition 3.2.** A monotone subadditive and concave function \( f \) on \( \mathbb{R}^n_+ \) is said to have an extension on \( \mathbb{R}^n_+ \) if there exists a function \( f^* \) that is monotone subadditive and concave on \( \mathbb{R}^n \) and \( f^*(x) = f(x) \) for all \( x \in \mathbb{R}^n_+ \).

For example, consider \( f(x) = g(\sum_{i=1}^n x_i) \) on \( \mathbb{R}^n_+ \) with \( g(z) = 1 - e^{-\lambda z} \) for parameter \( \lambda > 0 \), and \( z \in \mathbb{R}_+ \). Recall that this function has the weak homogeneity degree of value 0 and hence Corollary 3.1 cannot be applied. However, note that \( f \) has an extension on \( \mathbb{R}^n \), e.g. given as \( f^*(x) = g^*(\sum_{i=1}^n x_i) \) where

\[
g^*(z) = \begin{cases} 
1 - e^{-\lambda z} & \text{if } z \geq 0 \\
\lambda z & \text{otherwise}.
\end{cases}
\]

In the next theorem we show an approximation guarantee for functions \( f \) that have extensions on \( \mathbb{R}^n \).

**Theorem 3.4.** Assume that \( f \) is a monotone subadditive, concave function on \( \mathbb{R}^n_+ \) that has an extension on \( \mathbb{R}^n \), and \( \epsilon \in (\Delta, 1) \). Then, the discretization algorithm guarantees that for every set \( S \subseteq \Omega \) such that \( |S| \leq k \), we have

\[
\frac{1}{2}(1 - \epsilon)^{k-1}(1 - \Delta/\epsilon)v(S) \leq u(S) \leq 2\frac{1 + ak/\epsilon - \Delta/\epsilon}{(1 - \epsilon)^k(1 - \Delta/\epsilon)^2}v(S).
\]

We also have the following corollary.

**Corollary 3.2.** Assume that \( \Delta k < 1 \). Under same conditions as in Theorem 3.4 and taking \( a = \epsilon^2 \) and \( \epsilon = c/k \), for some constant \( c \in (\Delta k, 1) \), for every set \( S \subseteq \Omega \) such that \( |S| \leq k \),

\[
\frac{1}{2}e^{-\frac{c}{\epsilon^2}}(1 - \Delta k/c)v(S) \leq u(S) \leq 2e^{-\frac{c}{\epsilon^2}}\frac{1 + c}{(1 - \Delta k/c)^2}v(S).
\]
Under conditions of Corollary 3.2, we have an $\alpha$-sketch with

$$\alpha = 4 \frac{1 + c}{(1 - \Delta k/c)^3 e^{2/(1 - \epsilon)}}$$

which can be made arbitrarily close to 4 by taking $c$ small enough when $\Delta = 0$.

Theorem 3.4 alleviates the need for the weak homogeneity condition for some concave functions, and covers some concave functions which do not satisfy this condition with a positive degree.

Note that not all concave functions have an extension on $\mathbb{R}^n$. Consider again the example $f(x) = g(\sum_{i=1}^{n} x_i)$ such that $g(z)$ has a vertical tangent at $z = 0$. If $g(z)$ is differentiable at $z = 0$, then this is equivalent to $\lim_{z \to 0} dg(z)/dz = \infty$. In this case, $f$ does not have an extension on $\mathbb{R}^n$. An example of when this is the case is when $g$ is a power function $g(z) = z^r$, with $r \in (0, 1)$.

### 3.2.3 Coordinate-wise conditions

There exist functions that are not weakly homogeneous, but satisfy a weaker definition of coordinate-wise weak homogeneity. As an example, consider $x \in \mathbb{R}^2$ and $f(x) = x_1 + x_1 x_2$. Note that

$$f(\theta x) = \theta x_1 + \theta^2 x_1 x_2 \geq \theta f(x) \cdot \theta.$$ 

Clearly, we cannot find a positive $\eta$ for this function to satisfy the definition of weak homogeneity. However, we can easily show that this function is coordinate-wise weakly homogeneous.

**Definition 3.3.** A function $f$ is said to be coordinate-wise weakly homogeneous of degree $d$ and tolerance $\eta$ over a set $\Theta \in \mathbb{R}$ if for every $i \in [n]$,

$$\frac{1}{\eta} \theta f(x) \leq f \left( \sum_{j \neq i} x_j e_j + \theta x_i e_i \right) \leq \theta^d f(x),$$

for every $x$ in the domain of $f$ and all $\theta \in \Theta$.

In this section, we provide approximation guarantees when a weak homogeneous condition holds coordinate-wise.

**Theorem 3.5.** Assume that $f$ is a monotone subadditive or submodular function, and is coordinate-wise weakly homogeneous with degree $d$ and tolerance $\eta$ over $[0, 1]$ and $\epsilon \in (\Delta, 1)$. Then, the discretization algorithm guarantees that for every set $S \subseteq \Omega$ such that $|S| \leq k$, we have

$$\frac{1}{2} (1 - \epsilon)^{k-1} (1 - \Delta/\epsilon) v(S) \leq u(S) \leq 2\eta^k \frac{1 + a^d k/\epsilon - \Delta/\epsilon}{(1 - \epsilon)^{2k-1}(1 - \Delta/\epsilon)^2} v(S).$$

Proof is provided in Appendix A.12.

From Theorem 3.5, we have the following corollary.

**Corollary 3.3.** Assume that $\Delta k < 1$. Under same conditions as in Theorem 3.5 and such that $\eta = 1$, by taking $a = \epsilon^{2/d}$ and $\epsilon = c/k$, for some constant $c \in (\Delta k, 1)$, for every set $S \subseteq \Omega$ such that $|S| \leq k$,

$$\frac{1}{2} e^{-\frac{1}{1+k}} (1 - \Delta k/c) v(S) \leq u(S) \leq 2 \frac{1 + c}{(1 - \Delta k/c)^2} e^{\frac{1}{1+k}} v(S).$$
Any function $f$ that is subadditive and coordinate-wise convex on a domain that includes 0 and is such that $f(0) = 0$, is weakly homogeneous over $[0, 1]$ with tolerance $\eta = 1$. Any function $f$ that is coordinate-wise concave on a domain that includes 0 and is such that $f(0) \geq 0$, is weakly homogeneous over $[0, 1]$ with tolerance $\eta = 1$.

### Univariate transformations

For any given function $f$, we may establish approximation guarantees by validating conditions of the theorems in previous sections for a function $f^*$ such that $f^*(x_1, \ldots, x_n) = f(\phi_1(x_1), \ldots, \phi_n(x_n))$ for some continuous and strictly increasing functions $\phi_1, \ldots, \phi_n$. The univariate transformations $\phi_1, \ldots, \phi_n$ correspond to a change of variables that only affects the input distributions. Using univariate transformations can be useful in some cases. We illustrate this by two examples.

First, let us consider the case when $f(x) = (\sum_{i=1}^{n} x_i)^r$, with $r \in (0, 1)$. This function is weakly homogeneous over $[0, 1]$ with a degree of $r$. We can apply Corollary 3.1 to obtain a constant-factor approximation, with discretized distributions having supports of size $O((1/r)k \log(k))$. We can avoid having this dependence on $r$ by using univariate transformations $\phi_i(z) = z^{1/r}$. We thus need to validate conditions for $f^*(x) = (\sum_{i=1}^{n} x_i^{1/r})^r$, with $r \in (0, 1)$. Function $f^*$ is subadditive, submodular, convex, and weakly homogeneous over $[0, 1]$ with degree 1 and tolerance 1. Thus, by Corollary 3.1, we have a constant-factor approximation with discretized distributions having supports of size $O(k \log(k))$.

Second, consider the case when $f(x) = 1 - \prod_{i=1}^{n} (1 - x_i)$ on $[0, 1]^n$. This function is submodular and is weakly homogeneous over $[0, 1]$ with degree $d \leq 1/2$ and tolerance 1. Further details on these properties are provided in Appendix A.1.4. Again, we can apply Corollary 3.1, which gives a constant-factor approximation with $O((1/d)k \log(k))$ support size of discretized distributions. We can remove this dependence on $d$ by considering the transformations $\phi_i(z) = 1 - e^{-z}$. Hence, we have $f^*(x) = 1 - e^{-\sum_{i=1}^{n} x_i}$. We can apply Corollary 3.2 to show that a constant-factor approximation holds with $O(k \log(k))$ support size of discretized distributions.

### Using sketch value oracles in discrete optimization

In this section, we discuss application of our function approximation results to the best set selection and submodular welfare maximization problems.

The **best set selection** problem asks to find a set $S^* \subseteq \Omega$ such that $S^* \in \arg \max_{S \subseteq \Omega: |S|=k} u(S)$, given the cardinality constraint parameter $k$. A set $S$ is said to be a $\rho$-approximate solution for the best set selection problem if $u(S^*) \leq \rho u(S)$. If $v$ is a $\alpha$-sketch of $u$ and $S$ is a $\rho$-approximation solution for the best set selection problem with the objective function $v$, then $S$ is a $\rho \alpha$-approximate solution for the best set selection problem with the objective function $u$. This guarantee holds, even more generally, for the submodular welfare maximization problem. In this problem, given a positive integer $m$ and cardinality constraint parameters $k_1, \ldots, k_m$, the goal is to find disjoint sets $S_1, \ldots, S_m \subseteq \Omega$ of cardinalities $k_1, \ldots, k_m$ that maximize $\sum_{j=1}^{m} u_j(S_j)$, where $u_1, \ldots, u_m$ are monotone submodular set functions.

It is well-known that a greedy algorithm provides a $1/(1 - 1/e)$-approximation for the best set selection problem.
for any monotone submodular objective function Nemhauser et al. (1978). This greedy algorithm starts with an empty set and adds one item at each step to this set, choosing an item in each step that maximizes the marginal value gain. An allocation (winner determination) greedy algorithm provides a 2-approximation for the submodular welfare maximization problem Lehmann et al. (2006).

For a set valuation function of the form (2), with probability distributions of item values having a finite support, each of size at most $s$, evaluating $u(S)$ for a set $S$ of cardinality $k$ has $s^k$ computation complexity. Hence, it follows that the computation complexity of the greedy algorithm using value oracle calls for the set function of the form (2) with distributions of item values with supports of size at most $s$ is $O(ns^k)$. This is easily seen as follows. In each step $t \in \{1, \ldots, k\}$, the algorithm needs to compute values of $n - (t - 1)$ set functions, each for a set of cardinality $t$. Hence, the total computation complexity is $O(ns^k)$. Clearly, if $s = O(1)$ and $k = O(1)$, then the greedy algorithm has $O(n)$ complexity. The greedy algorithm has a polynomial complexity $O(n^{1+\gamma})$ for a positive constant $\gamma$, if, and only if, $s^k = O(n^\gamma)$. For example, this holds if $s = O(k \log(k))$ and $k \leq \gamma \log(n)/\log(\log(n))$.

We have the following implication of Corollary 3.1.

**Corollary 4.1.** Assume that $\Delta_k < 1$. For the class of functions satisfying conditions of Theorem 3.3, and by taking $a = c^{2/d}$ and $\epsilon = c/k$, for some $c \in (\Delta_k, 1)$, greedy algorithms for best set selection and submodular welfare maximization problems guarantee the approximation factor

$$4\eta\rho \frac{1 + c}{(1 - \Delta_k/c)^3} e^{2\frac{c}{1 - c}}$$

where $\rho$ is a constant, which for the best set selection problem is equal to $1/(1 - 1/e)$, and for the submodular welfare maximization problem is equal to 2.

If $\Delta = 0$, this approximation factor in Corollary 4.1 can be made arbitrarily close to $4\eta\rho$ by taking $c$ small enough. Recall that for any input item value distributions, we can choose $\Delta$ arbitrarily small by extending the definition of the sketch value functions as discussed in Section 3.2.

## 5 Numerical results

In this section, we present the results of our numerical experiments. The goal is to assess the performance of a sketch valuation function under various assumptions on item value distributions, set utility functions, set size, and parameters of the discretization algorithm. We also compare the performance against the baseline method based on test scores proposed in Sekar et al. (2021) and demonstrate that our sketch outperforms this baseline in terms of approximation accuracy. We have performed our experiments on both synthetic and real-world data sets. The code we used is available on GitHub: [https://github.com/Sketch-EXP/Sketch](https://github.com/Sketch-EXP/Sketch).

### 5.1 Synthetic data

We consider a ground set of $n = 50$ items. For each item, we generate $N = 500$ training samples of random performance values and estimate the value of the set utility function $u(S)$. We then choose the parameter $\epsilon$ of our discretization algorithm and compute the value of the sketch $v(S)$. To assess the performance of our algorithm,
we randomly generate 50 sets of size $k$ from the ground set of $n$ elements and estimate the ratio $v(S)/u(S)$. We examine three types of set utility functions: the maximum value, the CES function with degree 2, and the square root function of sum. Additionally, we consider two parametric families of distributions, the exponential and Pareto distribution. For the exponential distribution, we sample the mean value of each item uniformly from the unit interval [0, 1]. For the Pareto distribution, we sample the shape parameter of each item uniformly from the interval [1.1, 3] and fix the scale parameter to 1.5.

For each setting of set utility function and item value distributions, we test various values from 1 to 20 for the set size $k$. For each value of set size $k$, we set $\epsilon = c/k$ where $c$ is some constant between 0.1 and 10. Based on our theoretical results, we expect good approximation ratios when $c$ is small enough.

In Figure 4, boxplots aggregate the results from all settings of set size $k$ for different set utility functions and item
value distributions with $c = 0.1$. It can be observed that the ratio values are concentrated around 1, indicating that our sketch approximates the original set utility function well for most instances. Figure 5, shows the results for different values of $\epsilon$, aggregated over all settings of set size $k$. It can be seen that the ratio starts to deteriorate around the value $\epsilon = 1/k$, regardless of the set size $k$. In Figure 6, we compare the performance with the test score benchmark. The plot suggests that our sketch outperforms the test score sketch in terms of approximation accuracy.

5.2 Real-world data

We tested our method on three real-world datasets: YouTube, StackExchange, and New York Times data. In the YouTube data, we consider items as content publishers, and their performance is measured by the number of views their content pieces receive. For the StackExchange data, items are considered as experts, and their performance is quantified by the rate of upvotes received for their answers to questions. In the New York Times data, items are regarded as news sections, and their performance is assessed by the number of comments per news piece. All datasets used in our study are publicly available.

For each item, we compute the empirical distribution of its performance based on the given data. Subsequently, we generate $N = 100$ training samples for each item’s performance to estimate the set utility functions. We set
\( \epsilon = 0.1 \) and the set size \( k \) to 5 and 10. In Figure 7, we present a summary of our results. It is observed that our sketch provides a good approximation in most cases. Additional details about our experiments, discussions, and further results are provided in Section B.

6 Conclusion

In this paper, we addressed the problem of finding a good sketch of a stochastic set valuation function, which is defined as the expectation of a valuation function of independent random item values. We introduced an efficient algorithm to compute the sketch valuation function, representing it as the expectation of the valuation function with respect to discretized item value distributions. Our analysis reveals that, for a wide class of monotone subadditive or submodular valuation functions that satisfy specific conditions, our algorithm achieves a constant-factor approximation for any value query related to a set of items of size at most \( k \). Notably, our approach uses discretized distributions with support sizes of \( O(k \log(k)) \).

Our work provides the first positive results on function approximation for a class functions that can accommodate a wide-range of valuation functions studied in the existing literature. The results are also of relevant for applications in best set selection and welfare maximization problems. As part of future work, it may be interesting to consider other systematic discretization strategies and investigate the trade-off between approximation accuracy and the complexity of the sketch.

References

N. Alon, Y. Matias, and M. Szegedy. The space complexity of approximating the frequency moments. In Proceedings of the Twenty-Eighth Annual ACM Symposium on Theory of Computing, STOC ’96, page 20–29, New York, NY, USA, 1996. Association for Computing Machinery.

A. Asadpour and H. Nazerzadeh. Maximizing stochastic monotone submodular functions. Management Science, 62 (8):2374–2391, 2016.

A. Badanidiyuru, S. Dobzinski, H. Fu, R. Kleinberg, N. Nisan, and T. Roughgarden. Sketching valuation functions. In Proceedings of the Twenty-Third Annual ACM-SIAM Symposium on Discrete Algorithms, SODA ’12, page 1025–1035, USA, 2012. Society for Industrial and Applied Mathematics.

M.-F. Balcan and N. J. Harvey. Learning submodular functions. In Proceedings of the Forty-Third Annual ACM Symposium on Theory of Computing, STOC ’11, page 793–802, New York, NY, USA, 2011. Association for Computing Machinery.

A. A. Bian, B. Mirzasoleiman, J. Buhmann, and A. Krause. Guaranteed Non-convex Optimization: Submodular Maximization over Continuous Domains. In A. Singh and J. Zhu, editors, Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, volume 54 of Proceedings of Machine Learning Research, pages 111–120, 20–22 Apr 2017.
C. Blair. Problem complexity and method efficiency in optimization (a. s. nemirovsky and d. b. yudin). *SIAM Review*, 27(2):264–265, 1985.

M. Blum, R. W. Floyd, V. Pratt, R. L. Rivest, and R. E. Tarjan. Time bounds for selection. *J. Comput. Syst. Sci.*, 7(4):448–461, aug 1973.

C. Blundell, J. Cornebise, K. Kavukcuoglu, and D. Wierstra. Weight uncertainty in neural network. In F. Bach and D. Blei, editors, *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pages 1613–1622, Lille, France, 07–09 Jul 2015. PMLR.

O. Catoni. Challenging the empirical mean and empirical variance: A deviation study. *Annales de l’Institut Henri Poincaré, Probabilités et Statistiques*, 48(4):1148 – 1185, 2012.

K. Cohavi and S. Dobzinski. Faster and simpler sketches of valuation functions. *ACM Trans. Algorithms*, 13(3), Mar. 2017.

D. Cohen, B. Mitra, O. Lesota, N. Rekabsaz, and C. Eickhoff. Not all relevance scores are equal: Efficient uncertainty and calibration modeling for deep retrieval models. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR ’21, page 654–664, New York, NY, USA, 2021. Association for Computing Machinery.

G. Cormode and K. Yi. *Small Summaries for Big Data*. Cambridge University Press, 2020.

G. Cormode, Z. Karnin, E. Liberty, J. Thaler, and P. Veselý. Relative error streaming quantiles. In *Proceedings of the 40th ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems*, PODS’21, page 96–108, New York, NY, USA, 2021. Association for Computing Machinery. ISBN 9781450383813.

Y. Gal and Z. Ghahramani. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In M. F. Balcan and K. Q. Weinberger, editors, *Proceedings of The 33rd International Conference on Machine Learning*, volume 48 of *Proceedings of Machine Learning Research*, pages 1050–1059, New York, New York, USA, 20–22 Jun 2016. PMLR.

M. X. Goemans, N. J. A. Harvey, S. Iwata, and V. Mirrokni. Approximating submodular functions everywhere. In *Proceedings of the 2009 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 535–544, 2009.

M. Greenwald and S. Khanna. Space-efficient online computation of quantile summaries. In *Proceedings of the 2001 ACM SIGMOD International Conference on Management of Data*, SIGMOD ’01, page 58–66, New York, NY, USA, 2001. Association for Computing Machinery.

X. He, J. Pan, O. Jin, T. Xu, B. Liu, T. Xu, Y. Shi, A. Atallah, R. Herbrich, S. Bowers, and J. Q. n. Candela. Practical lessons from predicting clicks on ads at facebook. In *Proceedings of the Eighth International Workshop on Data Mining for Online Advertising*, ADKDD’14, page 1–9, New York, NY, USA, 2014. Association for Computing Machinery. ISBN 9781450329996.

R. Herbrich, T. Minka, and T. Graepel. Trueskill™: A bayesian skill rating system. In B. Schölkopf, J. Platt, and T. Hoffman, editors, *Advances in Neural Information Processing Systems*, volume 19. MIT Press, 2006.
E. Hüllermeier and W. Waegeman. Aleatoric and epistemic uncertainty in machine learning: an introduction to concepts and methods. *Machine Learning*, 110(3):457–506, 2021.

M. R. Jerrum, L. G. Valiant, and V. V. Vazirani. Random generation of combinatorial structures from a uniform distribution. *Theoretical Computer Science*, 43:169–188, 1986.

Kaggle.com. New york times articles & comments (2020), 2020. URL https://www.kaggle.com/benjaminawd/new-york-times-articles-comments-2020.

Kaggle.com. Youtube dislikes dataset, 2021. URL https://www.kaggle.com/dmitrynikolaev/youtube-dislikes-dataset.

M. J. Klass. A Method of Approximating Expectations of Functions of Sums of Independent Random Variables. *The Annals of Probability*, 9(3):413 – 428, 1981.

J. Kleinberg and M. Raghu. Team performance with test scores. *ACM Trans. Econ. Comput.*, 6(3–4), Oct. 2018. ISSN 2167-8375.

A. J. Kleywegt, A. Shapiro, and T. Homem-de Mello. The sample average approximation method for stochastic discrete optimization. *SIAM Journal on Optimization*, 12(2):479–502, 2002.

B. Lakshminarayanan, A. Pritzel, and C. Blundell. Simple and scalable predictive uncertainty estimation using deep ensembles. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS’17, page 6405–6416, Red Hook, NY, USA, 2017. Curran Associates Inc. ISBN 9781510860964.

D. Lee, M. Vojnovic, and S.-Y. Yun. Test score algorithms for budgeted stochastic utility maximization. *INFORMS Journal on Optimization*, 5(1):27–67, 2023.

J. C. Lee and P. Valiant. Optimal sub-gaussian mean estimation in R. In 2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS), pages 672–683, 2022.

B. Lehmann, D. Lehmann, and N. Nisan. Combinatorial auctions with decreasing marginal utilities. *Games and Economic Behavior*, 55(2):270–296, 2006.

G. Lugosi and S. Mendelson. Mean estimation and regression under heavy-tailed distributions: A survey. *Foundations of Computational Mathematics*, 19(5):1145–1190, 2019.

A. Malinin, L. Prokhorenkova, and A. Ustimenko. Uncertainty in gradient boosting via ensembles. In *International Conference on Learning Representations*, 2021.

A. Mehta, U. Nadav, A. Psomas, and A. Rubinstein. Hitting the high notes: Subset selection for maximizing expected order statistics. In H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 15800–15810. Curran Associates, Inc., 2020.

G. L. Nemhauser, L. A. Wolsey, and M. L. Fisher. An analysis of approximations for maximizing submodular set functions—i. *Mathematical Programming*, 14(1):265–294, 1978.
T. Popoviciu. Sur les équations algébriques ayant toutes leurs racines réelles. *Mathematica (Cuj)*, 9:129–145, 1935.

A. Schönhage, M. Paterson, and N. Pippenger. Finding the median. *J. Comput. Syst. Sci.*, 13:184–199, 1976.

S. Sekar, M. Vojnovic, and S. Yun. A test score-based approach to stochastic submodular optimization. *Manag. Sci.*, 67(2):1075–1092, 2021.

T. Soma and Y. Yoshida. A generalization of submodular cover via the diminishing return property on the integer lattice. In C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 28. Curran Associates, Inc., 2015.

D. M. Topkis. Minimizing a submodular function on a lattice. *Operations Research*, 26(2):305–321, 1978.

J. Zhu, J. Wang, I. J. Cox, and M. J. Taylor. Risky business: Modeling and exploiting uncertainty in information retrieval. In *Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR ’09, page 99–106, New York, NY, USA, 2009. Association for Computing Machinery. ISBN 9781605584836.
A Additional results and proofs

A.1 Further properties of valuation functions

A.1.1 Not all submodular functions are subadditive

Consider, for example, the success-probability value function:

\[ f(x) = 1 - \prod_{i=1}^{n} (1 - p(x_i)). \]

Here \( p : \mathbb{R} \rightarrow [0, 1] \) is an increasing function. The submodularity of this function can be verified by checking its satisfaction of the weak DR property. For any \( x, y \in \mathbb{R}^n \) such that \( x \leq y \), the increasing nature of \( p \) implies that \( f(x) \leq f(y) \). To confirm the weak DR property, consider adding \( z \) to the \( j \)-th basis direction of both \( x \) and \( y \), such that \( x \leq y \) and \( x_j = y_j \). The weak-DR condition becomes:

\[ \prod_{i \neq j} (1 - p(x_i))(1 - p(x_j + z)) \geq \prod_{i \neq j} (1 - p(y_i))(1 - p(y_j + z)). \]

This condition clearly holds since \( x_j = y_j \) and \( f(x) \leq f(y) \). However, it is important to note that for certain choices of function \( p \), \( f \) may not be subadditive. For instance, when \( n = 1 \), \( f \) is subadditive if and only if \( p \) is subadditive.

A.1.2 Coordinate-wise concave functions

There exist functions that are coordinate-wise concave but are not coordinate-wise concave according to the classical notion of concave functions. An example is the maximum value function \( f(x) = \max\{x_1, \ldots, x_n\} \) for \( n > 1 \) (this function is, in fact, coordinate-wise convex according to classical notion of convex functions). An example of a function that is concave according to the classical notion of concave functions (and thus is coordinate-wise concave according to the classical notion of concave functions) is \( f(x) = g(\sum_{i=1}^{n} x_i) \), where \( g \) is a concave function.

A.1.3 Extended diminishing returns property

A function \( f \) is said to satisfy the extended diminishing returns property Sekar et al. (2021) if for any \( i \in [n] \) and \( v \geq 0 \) that has a non-empty preimage under \( f \), there exists \( y \in \mathbb{R}^n_+ \) with \( y_i = 0 \) such that (a) \( f(y) = v \) and (b) \( f(x + ze_i) - f(x) \geq f(y + ze_i) - f(y) \) for any \( z \in \mathbb{R} \) and \( x \) such that \( f(x) \leq f(y) = v \) and \( x_i = 0 \). A simpler but stronger property is that \( f \) is such that \( f(x + ze_i) - f(x) \geq f(y + ze_i) - f(y) \) for every \( z \in \mathbb{R} \) and \( x, y \) such that \( f(x) \leq f(y) \) and \( x_i = y_i = 0 \).

There are functions that satisfy the extended diminishing returns property but that are not DR-submodular. Consider, for example, \( f(x) = (\sum_{i=1}^{n} x_i^r)^{1/r} \), for \( r > 1 \). Function \( f \) satisfies the extended diminishing returns property, as shown in Sekar et al. (2021). However, \( f \) is not DR-submodular. To see this note that \( f \) is twice-differentiable and is a convex function; hence it is coordinate-wise convex according to standard notion of convex functions. On the
We next show that $1 - f$ is subadditive.

We consider the function $f$ which shows that $f$ is subadditive. For the base case $j = 1$, function $f_1(x) = x_1$ is clearly subadditive. For the induction step, assume that $f_j$ is subadditive, for an arbitrary $1 < j < n$. Then, we have

$$f_{j+1}(x + y) = x_{j+1} + y_{j+1} + f_j(x + y) - (x_{j+1} + y_{j+1})f_j(x + y)$$

$$= x_{j+1} + y_{j+1} + (1 - x_{j+1} - y_{j+1})f_j(x + y)$$

$$\leq x_{j+1} + y_{j+1} + (1 - x_{j+1} - y_{j+1})(f_j(x) + f_j(y))$$

$$= f_{j+1}(x) + f_{j+1}(y) - x_{j+1}f_j(y) - y_{j+1}f_j(x)$$

$$\leq f_{j+1}(x) + f_{j+1}(y)$$

which shows that $f_{j+1}$ is subadditive.

We next show that $f$ is weakly homogeneous over $[0, 1]$ with tolerance $\eta = 1$ by induction as follows. For the base case $j = 1$, $f_1(x) = x_1$, so clearly it holds $f_1(\theta x) \geq \theta f_1(x)$. For the induction step, assume that $f_j(\theta x) \geq \theta f_j(x)$, for an arbitrary $1 \leq j < n$. Then, we have

$$f_{j+1}(\theta x) = \theta x_{j+1} + f_j(\theta x) - \theta x_{j+1}f_j(\theta x)$$

$$= \theta x_{j+1} + (1 - \theta x_{j+1})f_j(\theta x)$$

$$\geq \theta x_{j+1} + (1 - \theta x_{j+1})\theta f_j(x)$$

$$= \theta x_{j+1} + \theta f_j(x) - \theta^2 x_{j+1}f_j(x)$$

$$\geq \theta x_{j+1} + \theta f_j(x) - \theta x_{j+1}f_j(x)$$

$$= \theta f_{j+1}(x)$$

which shows that $f_{j+1}$ is weakly homogeneous over $[0, 1]$ with tolerance $\eta = 1$.

We next show that $f$ is weakly homogeneous over $[0, 1]$ with degree $d \leq 1/2$. To show this, let us consider the case when $n = 2$. We then have $f(x) = x_1 + x_2 - x_1x_2$. The condition $f(\theta x) \leq \theta^d f(x)$ can be written as follows

$$(1 - \theta^2 - d)x_1x_2 \leq (1 - \theta^1 - d)(x_1 + x_2)$$
for all $x_1, x_2 \in [0, 1]$. Clearly the last inequality holds when either $x_1 = 0$ or $x_2 = 0$. Hence, the condition is equivalent to

$$1 - \theta^{2-d} \leq (1 - \theta^{1-d}) \left( \frac{1}{x_1} + \frac{1}{x_2} \right)$$

for all $x_1, x_2 \in (0, 1]$.

This is clearly equivalent to $1 - \theta^{2-d} \leq 2(1 - \theta^{1-d})$ which can be written as

$$\theta^{1-d} (2 - \theta) \leq 1. \tag{9}$$

The left-hand side is increasing in $d$ and achieves the maximum value at $\theta^* = 1/(2(1 - d))$. Hence, equality in (9) is achieved at $\theta^*$ when $d = 1/2$.

### A.2 Proof of Lemma 2.2

If $f$ is a monotone submodular function, then by Lemma 2.1, $u$ is a monotone submodular set function; hence, $u$ is a monotone subadditive set function. Now, consider the case when $f$ is a monotone subadditive function. In this case, for any $S, T \subseteq \Omega$,

$$u(S) + u(T) = \mathbb{E}[f((X_i, i \in S))] + \mathbb{E}[f((X_i, i \in T))]$$

$$= \mathbb{E} \left[ f \left( \sum_{i \in S} X_ie_i \right) \right] + \mathbb{E} \left[ f \left( \sum_{i \in T} X_ie_i \right) \right].$$

By the monotonicity and subadditivity of $f$, for every $x$ in the domain of $f$, we have:

$$f \left( \sum_{i \in S} x_ie_i \right) + f \left( \sum_{i \in T} x_ie_i \right) \geq f \left( \sum_{i \in S} x_ie_i + \sum_{i \in T} x_ie_i \right)$$

$$\geq f \left( \sum_{i \in S \cup T} x_ie_i \right).$$

Thus, it follows:

$$u(S) + u(T) \geq \mathbb{E}[f((X_i, i \in S \cup T))] = u(S \cup T).$$
A.3 Proof of Lemma 2.3

For any $x, y \in \mathcal{X}$, we have

$$f(x + y) - f(x) = f\left(x + \sum_{i=1}^{n} y_i e_i\right) - f\left(x + \sum_{i=2}^{n} y_i e_i\right) + f\left(x + \sum_{i=2}^{n} y_i e_i\right) - f\left(x + \sum_{i=3}^{n} y_i e_i\right) + \ldots + f\left(x + e_n y_n\right) - f\left(x\right) \leq f\left(y_n e_n\right) - f\left(0\right) = f(y) - f(0)$$

where the inequalities hold by the DR-submodular property. Combining with $f(0) \geq 0$, we have $f(x + y) - f(x) \leq f(y)$, which is equivalent to saying that $f$ is subadditive on $\mathcal{X}$.

A.4 Proof of Lemma 2.4

By monotonicity, we observe that $f(x) \leq f\left(\lceil 1/\lambda \rceil \lambda x\right)$, and through subadditivity, we have $f\left(\lceil 1/\lambda \rceil \lambda x\right) \leq \lceil 1/\lambda \rceil f(\lambda x)$. Combining these inequalities, we can conclude that $f(x) \leq \lceil 1/\lambda \rceil f(\lambda x)$.

A.5 On the estimation of distribution properties (P1), (P2) and (P3)

Properties (P1), (P2) and (P3), stated in Section 3.1.2, can be evaluated within an approximation error in a stream computation setting. An $\epsilon$-approximate quantile up to additive error $\pm \epsilon N$ for a dataset of $N$ elements can be computed in one pass with a worst-case space requirement of $O((1/\epsilon) \log(\epsilon N))$ Greenwald and Khanna (2001). A randomized algorithm exists that returns an approximate quantile up to additive error $\pm \epsilon N$ with probability at least $1 - \delta$ with space complexity $O((1/\epsilon) \log(1/\delta))$. An $\epsilon$-approximate quantile up to multiplicative $(1 \pm \epsilon)$ with probability at least $1 - \delta$ can be computed by a randomized algorithm with space complexity $(1/\epsilon) \log(\epsilon N)^{3/2} \log(1/\delta)^{1/2}$ Cormode et al. (2021).

In cases when one can only draw samples from distribution $P$, properties (P1), (P2) and (P3) can be estimated by using statistical estimators. For every $x$, $P(x)$ can be estimated up to additive error $\epsilon$ with probability at least $1 - \delta$ by drawing $O((1/\epsilon^2) \log(2/\delta))$ samples from $P$. An approximate $\phi$-quantile up to additive $\pm \epsilon$ error with probability
at least $1 - \delta$ can be computed by using $O((1/\epsilon^2) \log(2/\delta))$ samples from $P$. Estimating $\mathbb{E}[f(X) \mid X > \tau]$ corresponds to the mean estimation problem. There is a long history of work on mean estimation. For distributions with finite variance, the median-of-means algorithm Alon et al. (1996); Jerrum et al. (1986); Blair (1985) is known to have sample complexity which is tight within a constant factor. Catoni (2012) improved the sample complexity to essentially optimal, i.e. with $1 + o(1)$ factor, for the cases when the variance of the distribution is known or the fourth moment is finite and bounded. Lee and Valiant (2022) found an estimator that is essentially optimal for any distribution with a finite variance. For a distribution with finite mean $\mu$ and finite variance $\sigma^2$, mean value can be estimated up to additive error $\pm \epsilon$ with probability at least $1 - 1/\delta$, provided that the number of samples is at least $(2\sigma^2/\epsilon^2) \log(1/\delta)/(1 + o(1))$. By Popoviciu’s inequality Popoviciu (1935), for any distribution with support contained in $[a, b]$, the variance $\sigma^2$ satisfies $\sigma^2 \leq (b - a)^2/4$ where the equality is achieved for uniform distribution on $\{a, b\}$. Sample mean estimation has also been studied for distributions with infinite variance; for example, see Lugosi and Mendelson (2019) and the references therein.

A.6 Proof of Lemma 3.1

To show property (i), note that on $(\tau, \infty)$, $Q$ has only one atom at $f^{-1}(H)$ at which it achieves a value of 1, and $P$ is an increasing function. Hence, it follows that, for all $x$,

$$Q(x) - P(x) \geq P(\tau) - P(f^{-1}(H))$$

$$= -\epsilon + (1 - P(f^{-1}(H)))$$

$$\geq -\epsilon.$$

This shows that property (i) holds.

Property (ii) holds straightforwardly because Algorithm 1 transfers any mass of $P$ on $[0, \tau]$ to smaller or equal values. Hence, $Q(x) < P(x)$ can only hold for some $x > \tau$.

Finally, property (iii) holds because, for all $x \geq \tau$,

$$Q(x) - P(x) \leq Q(f^{-1}(H)) - P(f^{-1}(H))$$

$$= 1 - P(f^{-1}(H))$$

$$\leq 1 - P(\tau) = \epsilon.$$

A.7 Proof of Lemma 3.2

Property (i) follows from the facts that for all $x$,

$$Q(x) - P(x) \geq P(\tau) - P(\mathbb{E}[X \mid X > \tau])$$

and

$$P(\tau) - P(\mathbb{E}[X \mid X > \tau]) = -(1 - (1 - 1/\beta)^2)\epsilon.$$

Property (ii) follows from the facts that for all $x \geq \tau$,

$$Q(x) - P(x) \leq 1 - P(\mathbb{E}[X \mid X > \tau]).$$
and \[ 1 - P(E[X \mid X > \tau]) = (1 - 1/\beta)^\epsilon. \]

Property (iii) follows by the following facts. For every \( x \in [x_m, \tau) \), for some \( 1 \leq j \leq J \),

\[
Q(x) - P(x) \leq Q(x_j) - P(x_j) = P(x_{j+1}) - P(x_j) \leq P(\tilde{x}_{j+1}) - P(\tilde{x}_j)
\]

where \( \tilde{x}_j = a\tau/(1 - \epsilon)^{j-1} \), for \( 1 \leq j \leq J + 1 \).

Under the condition \( \epsilon < a^\beta \),

\[
P(\tilde{x}_{j+1}) - P(\tilde{x}_j) = (\epsilon/a^\beta)(1 - \epsilon)^\beta(1 - (1 - \epsilon)^\beta) \leq \epsilon/a^\beta.
\]

Finally, property (iv) holds because by Algorithm 1, \( Q(0) = P(a\tau) \), assuming that \( P \) is the Pareto distribution \( P(a\tau) = 1 - 1/\beta \) if \( a\tau \geq x_m \) and \( P(a\tau) = 0 \) otherwise, and \( a\tau \geq x_m \) is equivalent to \( \epsilon/a^\beta \leq 1 \). Hence, it holds \( Q(0) = \max\{1 - \epsilon/a^\beta, 0\} \).

### A.8 Proof of Theorem 3.3

We separately consider the effects of three different transformations performed by Algorithm 1, namely, for the upper end, the lower end, and the middle part of the support of an item’s value distribution, on the approximation factors. We first introduce notation to define these transformations as follows.

Let \( \tau_i \) be the \((1 - \epsilon)\)-quantile of distribution \( P_i \), i.e.

\[
\tau_i = \inf\{x \in \mathbb{R} : P_i(x) \geq 1 - \epsilon\},
\]

where \( \epsilon \) is a parameter with a value in \((0, 1]\).

The algorithm limits the upper end of the support of distribution \( P_i \). Define \( H_i = E[f(X_i) \mid X_i > \tau_i] \) if \( P_i(\tau_i) < 1 \) and \( H_i = 0 \) otherwise. Let \( \tilde{X}_i \) be a random variable equal to \( X_i \) if \( X_i \leq \tau_i \) and equal to \( f^{-1}(H_i) \), otherwise. Here \( f^{-1} \) denotes the inverse of the function \( f(x, 0, \ldots, 0) \) with respect to \( x \). Note that \( \tilde{X}_i \) has a distribution with support contained in \([0, \tau_i] \cup \{f^{-1}(H_i)\}\).

The algorithm limits the lower end of the support by ensuring that values of \( \tilde{X}_i \) smaller or equal to \( a\tau_i \) are assigned to 0. This results in a new random variable \( \tilde{X}_i = \tilde{X}_i \mathbb{1}_{\{\tilde{X}_i > a\tau_i\}} \), which has a distribution with the support contained in \( \{0\} \cup [a\tau_i, \tau_i] \cup \{f^{-1}(H_i)\}\).

Each \( \tilde{X}_i \) is transformed into a random variable \( Y_i \) using exponential binning of the interval \([a\tau_i, \tau_i]\), where each value in a bin is mapped to the lower boundary of that bin. Formally, let \( q \) be the quantization function defined as
follows. Let $J = \lceil \log_{1/(1-\epsilon)}(1/a) \rceil$. Then, define
\[ q(x; \tau, \epsilon, a) = \frac{1}{(1-\epsilon)^{j-1} a\tau}, \quad \text{for } x \in I_j(\tau, \epsilon, a), \ 1 \leq j \leq J \]
and $q(x; \tau, \epsilon, a) = x$, for $x \geq \tau$, where
\[ I_j(\tau, \epsilon, a) = \left( \frac{1}{(1-\epsilon)^{j-1} a\tau}, \frac{1}{(1-\epsilon)^{j} a\tau} \right), \quad \text{for } 1 \leq j < J \]
and
\[ I_J(\tau, \epsilon, a) = \left( \frac{1}{(1-\epsilon)^{J-1} a\tau}, \tau \right). \]
Then, $Y_i = q(\tilde{X}_i; \tau_i, \epsilon, a)$.

Note that under the assumption that for all $i \in \Omega$, $P_i(\tau_i) - \lim_{z \uparrow x} P_i(z) \leq \Delta$, for all $x \in \mathbb{R}$, we have for all $i \in \Omega$,
\[ 1 - \epsilon \leq P(X_i \leq \tau_i) \leq 1 - \epsilon + \Delta. \quad (10) \]

Next, we will consider the effect of these three transformations on the approximation factors.

**Upper end** For each $i \in \Omega$ such that $P_i(\tau_i) < 1$, by definition of $\tilde{X}_i$, we have
\[ \mathbb{E}[f(\tilde{X}_i) \mid \tilde{X}_i > \tau_i] = \mathbb{E}[f(X_i) \mid X_i > \tau_i]. \]
Let
\[ w(S) = \mathbb{E} \left[ f \left( (X_i \mathbb{1}_{X_i \leq \tau_i}, i \in S) \right) \right]. \]
We have the following lemma.

**Lemma A.1.** Assume that $f$ is a monotone function that is either subadditive or submodular on its domain. Then, for every $S \subseteq \Omega$ such that $|S| \leq k$,
\[ u(S) \geq (1-\epsilon)^{k-1} (1-\Delta/\epsilon) \max \left\{ \epsilon \sum_{i \in S} H_i, w(S) \right\} \]
and
\[ u(S) \leq 2 \max \left\{ \epsilon \sum_{i \in S} H_i, w(S) \right\}. \]

**Proof** We first prove the upper bound. Let $T$ be the subset of $S$ containing every $i \in S$ such that $X_i$ exceeds the threshold value $\tau_i$, i.e. $T = \{i \in S \mid X_i > \tau_i\}$.

By Lemma 2.2, under the condition that $f$ is monotone and either subadditive or submodular, $u$ is a monotone subadditive function. Hence, we have
\[ \mathbb{E}[f((X_i, i \in S))] \]
\[ \leq \mathbb{E}[f((X_i, i \in T))] + \mathbb{E}[f((X_i, i \in S \setminus T))]. \]
Now, note
\[
E[f((X_i, i \in S))] 
\leq 2 \max \{E[f((X_i, i \in T))], E[f((X_i, i \in S \setminus T))]) \}
\leq 2 \max \{E[f((X_i, i \in T))], w(S)\}.
\]

By subadditivity of \( u \), we have
\[
E[f((X_i, i \in T))] \leq \sum_{i \in T} f(X_i) \leq \epsilon \sum_{i \in S} H_i
\] (11)

where recall \( H_i = E[f(X_i) \mid X_i > \tau_i] \) if \( P_i(\tau_i) < 1 \) and \( H_i = 0 \) otherwise.

Thus, it follows
\[
E[f((X_i, i \in S))] \leq 2 \max \left\{ \epsilon \sum_{i \in S} H_i, w(S) \right\}
\]
which proves the upper bound stated in the lemma.

We next prove the lower bound. Since \( f \) is a monotone function,
\[
E[f((X_i, i \in S))] 
\geq \max \{E[f((X_i, i \in T))], E[f((X_i, i \in S \setminus T))]\}.
\]

Thus, we have
\[
u(S) \geq \max \{E[f((X_i, i \in T))], w(S)\}. \quad (12)
\]

Now, note
\[
E[f((X_i, i \in T))]
= \sum_{U \subseteq S} P(T = U)E[f((X_i, i \in T)) \mid T = U]
\geq \sum_{U \subseteq S : |U| = 1} P(T = U)E[f((X_i, i \in T)) \mid T = U]
= \sum_{i \in S} P(X_i > \tau_i)P(X_j \leq \tau_j, \forall j \neq i) \notag
\]
\[
E[f(X_i) \mid X_i > \tau_i] \geq (\epsilon - \Delta)(1 - \epsilon)^{k-1} \sum_{i \in S} H_i
\]

where we used the facts \( P(X_j > \tau_j) \geq \epsilon - \Delta \) and \( P(X_j \leq \tau_j) \geq 1 - \epsilon \) for all \( j \in \Omega \) that follow from (10), and assumption that set \( S \) is such that \(|S| \leq k\).

Combining with (12), we have
\[
u(S) \geq \max \left\{ (\epsilon - \Delta)(1 - \epsilon)^{k-1} \sum_{i \in S} H_i, w(S) \right\}
\]
from which the lower bound stated in the lemma follows.

In the following lemma, we consider how the set function \( v_1(S) = \mathbb{E}[f((\hat{X}_i, i \in S))] \) compares to the set function \( u(S) = \mathbb{E}[f((\tilde{X}_i, i \in S))] \).

**Lemma A.2.** Assume that \( f \) is a monotone function that is either subadditive or submodular on its domain. Then, for every \( S \subseteq \Omega \) such that \( |S| \leq k \),

\[
\frac{1}{2} (1 - \epsilon)^{k-1} (1 - \Delta/\epsilon) v_1(S) \leq u(S) \leq 2 \frac{1}{(1 - \epsilon)^{k-1} (1 - \Delta/\epsilon)} v_1(S).
\]

**Proof** Note that

\[
u(S) \leq 2 \max \left\{ \epsilon \sum_{i \in S} H_i, w(S) \right\} = \frac{2}{(1 - \epsilon)^{k-1} (1 - \epsilon)^{k-1}} \max \left\{ \epsilon \sum_{i \in S} H_i, w(S) \right\} \leq \frac{2}{(1 - \epsilon)^{k-1} (1 - \Delta/\epsilon)} v_1(S)
\]

where the first inequality is by the upper bound in Lemma A.1 and the last inequality is by the lower bound in Lemma A.1.

This shows the upper bound in the statement of the lemma. The lower bound in the statement of the lemma follows by similar arguments as follows,

\[
u(S) \geq (1 - \epsilon)^{k-1} \left(1 - \frac{\Delta}{\epsilon}\right) \max \left\{ \epsilon \sum_{i \in S} H_i, w(S) \right\} = \frac{1}{2} (1 - \epsilon)^{k-1} \left(1 - \frac{\Delta}{\epsilon}\right) \cdot 2 \max \left\{ \epsilon \sum_{i \in S} H_i, w(S) \right\} \geq \frac{1}{2} (1 - \epsilon)^{k-1} \left(1 - \frac{\Delta}{\epsilon}\right) v_1(S)
\]

where the first inequality is by the lower bound and the last inequality is by the upper bound in Lemma A.1.

**Lower end** We next consider random variables defined as \( \tilde{X}_i := \hat{X}_i \mathbb{1}_{\tilde{X}_i > a \tau_i} \), for some \( a \in [0, 1] \). We compare the set function \( v_2(S) = \mathbb{E}[f((\tilde{X}_i, i \in S))] \) and the set function \( v_1(S) = \mathbb{E}[f((\hat{X}_i, i \in S))] \) in the following lemma.

**Lemma A.3.** Assume that \( f \) is a monotone function that is either subadditive or submodular, and is weekly homogeneous of degree \( d \) over \([0, 1]\). Then, for every \( S \subseteq \Omega \) such that \(|S| \leq k\), we have

\[
v_2(S) \geq \frac{1}{1 + a^d k/(\epsilon - \Delta)} v_1(S).
\]
Proof. For any monotone submodular function $f$ or any monotone subadditive function $f$, it holds, for any $a \in [0, 1]$,

$$v_1(S) = \mathbb{E}[f((\hat{X}_i, i \in S))]$$

$$= \mathbb{E}[f((\hat{X}_i \mathbb{I}_{\hat{X}_i \leq a\tau_i}, i \in S)]]$$

$$\leq \mathbb{E}[f((\hat{X}_i \mathbb{I}_{\hat{X}_i \leq a\tau_i}, i \in S))] + \mathbb{E}[f((\hat{X}_i \mathbb{I}_{\hat{X}_i > a\tau_i}, i \in S))]$$

$$\leq f((a\tau_i, i \in S)) + \mathbb{E}[f((\hat{X}_i, i \in S))]$$

Combining with the condition that $f$ is weakly homogeneous of degree $d$ over $[0, 1]$, we have

$$v_1(S) \leq a^d f((\tau_i, i \in S)) + v_2(S).$$

Now, note that for any monotone, subadditive or submodular function $f$,

$$\mathbb{E}[f((\hat{X}_i, i \in S))] \geq \frac{\epsilon - \Delta}{k} \mathbb{E}[f((\tau_i, i \in S))].$$

This can be shown as follows. Let $j \in \arg \max_{i \in S} \tau_i$. Then, we have

$$\mathbb{E}[f((\hat{X}_i, i \in S))] \geq \mathbb{P}(\hat{X}_j > \tau_j) f(\tau_j e_j)$$

$$\geq (\epsilon - \Delta) f(\tau_j e_j)$$

$$\geq \frac{\epsilon - \Delta}{k} \sum_{i \in S} \tau_j e_i$$

$$\geq \frac{\epsilon - \Delta}{k} f(\sum_{i \in S} \tau_i e_i)$$

$$= \frac{\epsilon - \Delta}{k} f((\tau_i, i \in S))$$

where we used the fact $\mathbb{P}(\hat{X}_j > \tau_j) = \mathbb{P}(X_j > \tau_j) \geq \epsilon - \Delta$, with the last inequality following from (10).

Putting the pieces together, we have

$$v_2(S) \geq \frac{1}{1 + a^d k/(\epsilon - \Delta)} v_1(S).$$

Middle part. We next consider the approximation error due to the last step of the algorithm. Recall that for each $i \in \Omega$, $Y_i = q(\hat{X}_i; \tau_i, \epsilon)$. We compare the set functions $v_2(S) = \mathbb{E}[f((\hat{X}_i, i \in S))]$ and $v(S) = \mathbb{E}[f((Y_i, i \in S))]$.

Lemma A.4. Assume that $f$ is monotone and weakly homogeneous with tolerance $\eta$. Then, we have

$$v(S) \leq \frac{1 - \epsilon}{\eta} v_2(S).$$
which is satisfied for \( \theta \) where
\[
(1 - \varepsilon) x.<.i,j
\]
This combined with monotonicity of \( f \) immediately yields
\[
v(S) \geq E[f((1 - \varepsilon)X_i, i \in S)].
\]
Combining with the condition that \( f \) is weakly homogeneous with tolerance \( \eta \) yields \( v(S) \geq ((1 - \varepsilon)/\eta)v_2(S) \).

**Putting the pieces together**  
The lower bound in the theorem follows from Lemma A.2. The upper bound in the theorem follows by combining Lemmas A.3 and A.4.

**A.9 Proof of Corollary 3.1**

Note that \( 1 - x \geq e^{-\theta x} \) for all \( x \leq 1 - 1/\theta \) and \( \theta \geq 1 \). Now, consider the case where \( \varepsilon = c/k \), with \( c \) being a positive constant in the range \((0, 1)\). For this choice of \( \varepsilon \), we have \( (1 - \varepsilon)^k \geq e^{-\theta c} \) under the condition \( c/k \leq 1 - 1/\theta \), which is satisfied for \( \theta \geq 1 \).

Taking \( \theta = 1/(1 - c) \) ensures that the condition holds, and we obtain \( (1 - \varepsilon)^k \geq e^{-\theta x} \). This inequality establishes the result stated in the corollary.

**A.10 Approximation results for distributions with arbitrary point masses**

For any fixed \( \Lambda \in (0, 1] \), let \( \Omega^* \) be the subset of items in \( \Omega \) containing those with value distributions having point masses larger than \( \Lambda \). For each \( i \in \Omega^* \), let \( x_{i,1}^*, \ldots, x_{i,m_i}^* \) be the locations of these point masses, and \( \alpha_{i,j} = P_i(x_{i,j}^*) - \lim_{x \rightarrow x_{i,j}^*} P_i(z) \) be the corresponding point masses. For each \( i \in \Omega \setminus \Omega^* \), define \( P_{i,j} \) and \( P_{i,0} \) such that \( P_{i,j} \) is the point mass distribution with all the mass on \( x_{i,j}^* \), i.e. \( P_{i,j}(x) = 0 \) if \( x < x_{i,j}^* \) and \( P_{i,j}(x) = 1 \) otherwise, and
\[
P_{i,0}(x) = \frac{P_i(x) - \sum_{j=1}^{m_i} \alpha_{i,j} P_{i,j}(x)}{1 - \sum_{j=1}^{m_i} \alpha_{i,j}}
\]
if \( \sum_{j=1}^{m_i} \alpha_{i,j} < 1 \), and \( P_{i,0}(x) = 0 \), otherwise. Note that, for every \( i \in \Omega^* \), we can write
\[
P_i(x) = \sum_{j=0}^{m_i} \alpha_{i,j} P_{i,j}(x) \tag{13}
\]
where \( \alpha_{i,0} := 1 - \sum_{j=1}^{m_i} \alpha_{i,j} \).

Now, using (13), we express \( u(S) \) as a weighted sum of stochastic valuation functions for any \( S \subseteq \Omega \). Consider \( S = \{i_1, \ldots, i_k\} \subseteq \Omega \). For each item \( i_\ell \), define virtual items \( i_{\ell,0}, \ldots, i_{\ell,m_\ell} \). Using (13), we can write
\[
u(S) = \sum_{j_1=0}^{m_1} \cdots \sum_{j_k=0}^{m_k} (\alpha_{i_1,j_1} \cdots \alpha_{i_k,j_k}) u(\{i_1,j_1, \ldots, i_k,j_k\}).
\]
This weighted sum has \((1/\Lambda + 1)^k\) elements. We approximate each \( u(\{i_1,j_1, \ldots, i_k,j_k\}) \) using sketch valuation functions. If \( j_\ell = 0 \), we use the discretization of \( P_{i_\ell,0} \) by Algorithm 1; otherwise, we use \( P_{i_\ell,j_\ell} \).
The results of Theorem 3.3 hold for any item value distributions $P_1, \ldots, P_n$ by choosing $\Delta \in (\epsilon, 1)$ in the decomposition.

Similarly, the results of Corollary 3.1 apply by choosing $\Lambda = 1/(kh(k))$, where $h(k) > 1$. If $h(k) = \omega(k)$, we achieve a constant-factor approximation with each discretized distribution having support size $s = O((1/d)k \log(k))$. Each item is represented by a discretized distribution with support size $O((1/d)k \log(k))$ and at most $1/\Delta = kh(k)$ point mass distributions.

The sketch valuation function defined by this decomposition has computation complexity $O((skh(k))^k)$.

A.11 Proof of Theorem 3.4

The proof for the upper end remains the same as in the proof of Theorem 3.3. We thus only need to address the lower end and middle part of the proof.

**Lower end** Let $f^*$ be a concave extension of $f$. Since $f^*(x) = f(x)$ for all $\mathbb{R}_n^+$ and we consider item value distributions with positive supports, we can consider $v_1(S) = \mathbb{E}_i[f^*((\hat{X}_i, i \in S))]$ and $v_2(S) = \mathbb{E}_i[f^*((\tilde{X}_i, i \in S))]$.

Recall that it holds

$$\mathbb{E}_i[f^*((\tilde{X}_i, i \in S))] \geq \frac{\epsilon - \Delta}{k} \mathbb{E}_i[f^*((\tau_i, i \in S))]. \quad (14)$$

Let $Z_i = \hat{X}_i - a\tau_i$ and note that $\hat{X}_i = \hat{X}_i 1_{\{\hat{X}_i > a\tau_i\}} \geq Z_i$. Note that we can write,

$$Z_i = (1 - a)\hat{X}_i + a(\hat{X}_i - \tau_i).$$

Since $f^*$ is monotone, concave and subadditive, we have the following inequalities.

$$v_2(S) \geq \mathbb{E}_i[f^*((Z_i, i \in S))]$$
$$\geq (1 - a)\mathbb{E}_i[f^*((\hat{X}_i, i \in S))]$$
$$+ a\mathbb{E}_i[f^*((\hat{X}_i - \tau_i, i \in S))]$$
$$\geq (1 - a)\mathbb{E}_i[f^*((\hat{X}_i, i \in S))]$$
$$+ a\mathbb{E}_i[f^*((\hat{X}_i, i \in S))] - a \cdot f^*((\tau_i, i \in S))$$
$$\geq v_1(S) - (ak/(\epsilon - \Delta))v_2(S)$$

where the first inequality is by monotonicity, the second inequality is by concavity, the third inequality is by subadditivity, and the last inequality is by the definition of $v_1(S)$ and the inequality in (14).

**Middle part** This follows by the same arguments as in the proof of Theorem 3.3 and making use of the fact that any concave function is weakly homogeneous with tolerance 1.

A.12 Proof of Theorem 3.5

The proof of the upper end remains the same as in the proof of Theorem 3.3. In what follows, we show the proof for the lower end part and the middle part of the proof.

**Lower end** This part is shown by the following lemma.
Lemma A.5. Assume that $f$ is a monotone function that is either subadditive or submodular and is coordinate-wise weakly homogeneous of degree $d$ over $[0, 1]$. Then, we have

$$v_2(S) \geq \frac{1}{1 + \alpha d k/(\epsilon - \Delta)} v_1(S).$$

Proof The proof can be established by similar steps as in the proof of Lemma A.3 and making use of the following fact: under coordinate-wise weakly homogeneous condition, $f((a \tau_i, i \in S)) \leq a^{dk} f((\tau_i, i \in S)) \leq a^d f((\tau_i, i \in S))$.

Middle part This follows by the same arguments as in the proof of Theorem 3.3 combined with repeated application of the weak homogeneity property that holds coordinate-wise which yields

$$E[f(((1 - \epsilon) \tilde{X}_i, i \in S))] \geq (1/\eta)^k (1 - \epsilon)^k E[f((\tilde{X}_i, i \in S))].$$

B Supplementary results for real-world data experiments

B.1 Information about datasets

The YouTube dataset Kaggle.com (2021) contains information about 37422 unique videos, including publication date, view counts, number of likes and dislikes, for the period from August 2020 to December 2021, for the USA, Canada and Great Britain. For our experiments, we filtered out YouTubers with fewer than 50 uploads. In the main experiment, we took the view counts per day as the measure for video performance. We also tested other metrics and the results can be found in Section B.2.

The StackExchange dataset contains information about 35218 questions and 88584 answers on Academia.StackExchange platform. The dataset is retrieved on Jan 20, 2022 from the official StackExchange data dump at [https://archive.org/details/stackexchange](https://archive.org/details/stackexchange). Each answer receives up-votes and down-votes from users of the platform, indicating quality of the answer. For our experiments, we took only the users who have submitted at least 100 answers. If an answer $a$ to question $q$ receives $u(a, q)$ up-votes and $d(a, q)$ down-votes, then we define

$$s(a, q) = \frac{u(a, q) + c_1}{u(a, q) + d(a, q) + c_2}$$

as the quality value of the answer, where $c_1$ and $c_2$ are positive-valued parameters. This metric is motivated by Bayesian estimation, and was used in Sekar et al. (2021). The ratio increases with the number of up-votes and decreases with the number of down-votes. It is called balanced when $c_1/c_2 = 1/2$, conservative if $c_1/c_2 < 1/2$. We took the conservative choice $(c_1, c_2) = (2, 8)$ for the main experiment. Results for other value pairs can be found in Section B.3.

The New York Times dataset Kaggle.com (2020) contains information about 16570 articles and comments on New York Times from January 2020 to December 2020. Each article belongs to one section. We took all articles and their comment numbers for our experiment.
We show the empirical CDFs for performance values aggregated over all data points in a dataset, for all three datasets in Figure 8. The performance values are calculated using the metrics specified in the main text. We observe that these three datasets have very different distributions, which implies that our method works well for different data distributions.

B.2 YouTube dataset: other performance metrics

In this section, we illustrate and compare the results using different measures for the performance value of YouTube video uploads. Specifically, if a video \( j \) uploaded \( \tau \) days ago received \( n(j) \) views, \( l(j) \) likes and \( d(j) \) dislikes, we calculate six measures for its performance as follows.

- **View counts per day (view ratio):** \( n(j)/\tau \)
- **Log of the view counts:** \( \log(n(j) + 1) \)
- **Standard like ratio:** \( l(j)/(l(j) + d(j)) \)
- **Video Power Index (VPI):** view ratio \( \times \) standard like ratio
- **Bayesian like ratio:** \((l(j) + c_1)/(l(j) + d(j) + c_2)\)
  - Conservative case where \( c_1 = 0.01n(j) \) and \( c_2 = 0.1n(j) \)
  - Balanced case where \( c_1 = 0.05n(j) \) and \( c_2 = 0.1n(j) \)

The Bayesian like ratio has a similar interpretation as for the StackExchange dataset. The difference is that in this setting the ratio factors in the effect of view counts. Note that \( c_1 \) can be seen as a threshold value needed for the number of likes to have an effect on the performance value.

We computed the performance values for all videos submitted by qualified YouTubers with more than 50 uploads. Figure 9 shows how the values are distributed for all six measures. Note that the distributions under the measures view counts per day, VPI and conservative Bayesian like ratio are more heavy-tailed compared to others.

Figure 10 shows the results for three objective functions (max, CES of degree 2 and square root function of sum) under the six measures. We can observe that all the approximation ratios for the measures log of the view counts, standard like ratio, and balanced Bayesian like ratio are more concentrated around 1 compared to the results using
view counts per day and VPI. This result is not surprising, as we noted that the value distributions under these two measures are centered and light-tailed.

**B.3 StackExchange dataset: other parameter settings**

We tested the following values of \((c_1, c_2)\): \((2, 8)\), \((8, 32)\) and \((10, 10)\). As explained previously, the first and second cases correspond to a conservative choice, and the last one is a balanced ratio.

Figure 11 shows the empirical CDF for performance values, for the three settings of \((c_1, c_2)\) parameters. We can see that as the values of \(c_1\) and \(c_2\) increase, the values are more concentrated around \(c_1/c_2\).

Figure 12 shows the corresponding results under the different value pairs for \(c_1\) and \(c_2\). From the plots, we observe that all the approximation ratios are highly concentrated around 1. Thus we can conclude that the choices of \(c_1\) and
Figure 11: Empirical CDFs of performance values for the StackExchange dataset, for three different parameter settings.

Figure 12: The approximation ratio for different valuation functions for the StackExchange dataset, for different parameter settings.

\(c_2\) have no particular effects on the approximation ratio.