Estimation for Monotone Sampling: Competitiveness and Customization

Edith Cohen
Microsoft Research
Mountain View, CA, USA
editco@microsoft.com

ABSTRACT

[1] Random samples are lossy summaries which allow queries posed over the data to be approximated by applying an appropriate estimator to the sample. The effectiveness of sampling, however, hinges on estimator selection. The choice of estimators is subjected to global requirements, such as unbiasedness and range restrictions on the estimate value, and ideally, we seek estimators that are both efficient to derive and apply and *admissible* (not dominated, in terms of variance, by other estimators). Nevertheless, for a given data domain, sampling scheme, and query, there are many admissible estimators.

We define monotone sampling, which is implicit in many applications of massive data set analysis, and study the choice of admissible nonnegative and unbiased estimators. Our main contribution is general derivations of admissible estimators with desirable properties. We present a construction of *order-optimal* estimators, which minimize variance according to *any* specified priorities over the data domain. Order-optimality allows us to customize the derivation to common patterns that we can learn or observe in the data. When we prioritize lower values (e.g., more similar data sets when estimating difference), we obtain the L* estimator, which is the unique monotone admissible estimator and dominates the classic Horvitz-Thompson estimator. We show that the L* estimator is 4-competitive, meaning that the expectation of the square, on any data, is at most 4 times the minimum possible for that data. These properties make the L* estimator a natural default choice. We also present the U* estimator, which prioritizes large values (e.g., less similar data sets). Our estimator constructions are general, natural, and practical, allowing us to make the most from our summarized data.

1. INTRODUCTION

Random sampling is a common tool in the analysis of massive data. Sampling is highly suitable for parallel or distributed platforms. The samples facilitate scalable approximate processing of queries posed over the original data, when exact processing is too resource consuming or when the original data is no longer available. Random samples have a distinct advantage over other synopsis in their flexibility. In particular, they naturally support *domain* (subset) queries, which specify a selected set of records. Moreover, the same sample can be used for basic statistics, such as sums, moments, and averages, and more complex relations: distinct counts, size of set intersections, and difference norms.

The value of a sample hinges on the accuracy within which we can estimate query results. In turn, this boils down to the *estimators* we use, which are the functions we apply to the sample to produce the estimate. As a rule, we are interested in estimators that satisfy desirable *global* properties, which must hold for *all possible data* in our data domain. Common desirable properties are:

• *Unbiasedness*, which means that the expectation of the estimate is equal to the estimated value. Unbiasedness is particularly important when we are ultimately interested in estimating a sum aggregate, and our estimator is applied to each summand. Typically, the estimate for each summand has high variance, but with unbiasedness (and pairwise independence), the relative error decreases with aggregation.

• *Range restriction* of estimates: since the estimate is often used as a substitute of the true value, we would like it to be from the same domain as the query result. Often, the domain is *nonnegative* and we would like the estimate to be nonnegative as well. Another natural restriction is *boundedness* which means that for each given input, the set of possible estimate values is bounded.

• *Finite variance* (implied by boundedness but less restrictive)

Perhaps the most ubiquitous quality measure of an estimator is its variance. The variance, however, is a function of the input data. An important concept in estimation theory is a Uniform Minimum Variance Unbiased (UMVUE) estimator [30], that is, a single estimator which attains the minimum possible variance for all inputs in our data domain [27]. A UMVUE estimator, however, generally does not exist. We instead seek an *admissible* (Pareto variance optimal) estimator [30] – meaning that strict improvement is not possible without violating some global properties. More precisely, an estimator is admissible if there is no other estimator that satisfies the global properties with at most the variance of our estimator on all data and strictly lower variance on some data. A UMVUE must be admissible, but when one does not exist, there is typically a full Pareto front of admissible estimators. We recently proposed *variance competitiveness* [15], as a robust "worst-case" performance measure when there is no UMVUE. The variance competitive ratio is the maximum, over data, of the ratio of the expectation of the square of our estimator to the minimum possible for the data subject to the global properties. A small ratio means that variance on each input in the data domain is not too far off the minimum variance attainable on this data by an estimator which satisfies the global properties. In particular, the UMVUE, when it exists, is 1-competitive.

We propose the following definition of monotone sampling. In the sequel we show how it incorporates common sampling schemes.

> A *monotone sampling scheme* $(\mathbf{V}, S^*)$ is specified by a *data domain* $\mathbf{V}$ and a mapping $S^* : \mathbf{V} \times (0,1] \rightarrow 2^{\mathbf{V}}$. The mapping is such that the set $S^*(\boldsymbol{v}, u)$ for a fixed v is monotone non-decreasing with u.

The sampling interpretation is that a *sample* $S(\boldsymbol{v}, u)$ of the *input* $\boldsymbol{v}$ (which we also refer to as the *data vector*) is obtained by drawing a *seed* $u \sim U[0,1]$, uniformly at random from $[0,1]$. The sample deterministically depends on $\boldsymbol{v}$ and the (random) *seed* u. The mapping $S^*(\boldsymbol{v}, u)$ is the set of all data vectors that are consistent with S (which we assume includes the seed value u). It represents all the information we can glean from the sample on the input. In

[1] This is a full version of a PODC 2014 paper

particular, we must have $\boldsymbol{v} \in S^*(\boldsymbol{v}, u)$ for all $\boldsymbol{v}$ and u. The sampling scheme is monotone in the randomization: When fixing $\boldsymbol{v}$, the set $S^*(\boldsymbol{v}, u)$ is non-decreasing with u, that is, the smaller u is, the more information we have on the data $\boldsymbol{v}$.

In the applications we consider, the (expected) representation size of the sample $S(\boldsymbol{v}, u)$ is typically much smaller size than $\boldsymbol{v}$. The set S^* can be very large (or infinite), and our estimators will only depend on performing certain operations on it, such as obtaining the infimum of some function. Monotone sampling can also be interpreted as obtaining a "measurement" $S(\boldsymbol{v}, u)$ of the data $\boldsymbol{v}$, where u determines the granularity of our measuring instrument. Ultimately, the goal is to recover some function of the data from the sample (the outcome of our measurement):

> A *monotone estimation* problem is specified by a monotone sampling scheme and a nonnegative function $f : \mathbf{V} \geq 0$. The goal is to specify an *estimator*, which is a function of all possible outcomes $\hat{f} : \mathcal{S} \geq 0$, where $\mathcal{S} = \{S(\boldsymbol{v}, u) | \boldsymbol{v} \in \mathbf{V}, u \in (0, 1]$. The estimator should be unbiased $\forall \boldsymbol{v}$, $\mathsf{E}_{u \sim U[0,1]} \hat{f}(S(\boldsymbol{v}, u)) = f(\boldsymbol{v})$ and satisfy some other desirable properties.

The interpretation is that we obtain a query, specified in the form of a nonnegative function $f : \mathbf{V} \geq 0$ on all possible data vectors $\boldsymbol{v}$. We are interested in knowing $f(\boldsymbol{v})$, but we can not see $\boldsymbol{v}$ and only have access to the sample S. The sample provides us with little information on $\boldsymbol{v}$, and thus on $f(\boldsymbol{v})$. We approximate $f(\boldsymbol{v})$ by applying an *estimator*, $\hat{f}(S) \geq 0$ to the sample. The monotone estimation problem is a bundling of a function f and a monotone sampling scheme. We are interested in estimators $\hat{f}$ that satisfy properties. We always require nonnegativity and unbiasedness and consider admissibilitiy, variance competitiveness, and what we call customization (lower variance on some data patterns).

Our formulation departs from traditional estimation theory. We view the data vectors in the domain as the possible inputs to the sampling scheme, and we treat estimator derivation as an optimization problem. The variance of the estimator parallels the "performance" we obtain on a certain input. The work horse of estimation theory, the maximum likelihood estimator, is not even applicable here as it does not distiguish between the different data vectors in S^*. Instead, the random "coin flips," in the form of the seed u, that are available to the estimator are used to restrict the set S^* and obtain meaningful estimates.

We next show how monotone sampling relates to the well-studied model of coordinated sampling, that has extensive applications in massive data analysis. In particular, estimator constructions for monotone estimation can be applied to estimate functions over coordinated samples.

Coordinated shared-seed sampling

In this framework our data has a matrix form of two or more ($r > 1$) *instances*, where each instance (row) has the form of a weight assignment to the (same) set of items (columns). Different instances may correspond to snapshots, activity logs, measurements, or repeated surveys that are taken at different times or locations. When instances correspond to documents, items can correspond to features. When instances are network neighborhoods, items can correspond to members or objects they store.

Over such data, we are interested in queries which depend on two or more of the instances and a subset D of the items. Some examples are Jaccard similarity, distance norms, or the number of distinct items with positive entry in at least one instance (distinct count).

Example 1 Dataset with 3 instances and queries

Instances $i \in \{1, 2, 3\}$ and items $k \in \{a, b, c, d, e, f, g, h\}$:

	a	b	c	d	e	f	g	h
v_1	0.95	0	0.23	0.70	0.10	0.42	0	0.32
v_2	0.15	0.44	0	0.80	0.05	0.50	0.20	0
v_3	0.25	0	0	0.10	0	0.22	0	0

Example queries over $D \subset [a\text{-}h]$. L_p difference, L_p^p: the pth power of L_p difference and a sum aggregate, L_{p+}^p: asymmetric (increase only) L_p^p. G: example "arbitrary" sum aggregate.

$$L_p^p(D) = \sum_{k \in D} |v_1^{(k)} - v_2^{(k)}|^p \qquad L_p(D) = (L_p^p(D))^{1/p}$$

$$L_{p+}^p(D) = \sum_{k \in D} \max\{0, v_1^{(k)} - v_2^{(k)}\}^p$$

$$G(D) = \sum_{k \in D} |v_1^{(k)} - 2v_2^{(k)} + v_3^{(k)}|^2$$

sum aggregate	*item function*
L_p^p	$\mathrm{RG}_p(\boldsymbol{v}) = (\max(\boldsymbol{v}) - \min(\boldsymbol{v}))^p$
L_{p+}^p	$\mathrm{RG}_{p+}(v_1, v_2) = \max\{0, v_1 - v_2\}^p$
G	$g(v_1, v_2, v_3) = \|v_1 + v_3 - 2v_2\|^2$

$$L_1(\{b, c, e\}) = |0 - 0.44| + |0.23 - 0| + |0.10 - 0.05| = 0.71$$

$$L_2^2(\{c, f, h\}) = (0.23 - 0)^2 + (0.50 - 0.42)^2 + (0.32 - 0)^2 \approx 0.16$$

$$L_2(\{c, f, h\}) = \sqrt{L_2^2(\{c, f, h\})} \approx 0.40$$

$$L_{1+}(\{b, c, e\}) = \max\{0, 0 - 0.44\} + \max\{0, 0.23 - 0\} + \\ + \max\{0, 0.10 - 0.05\} = 0.235$$

$$G(\{b, d\}) = |0 - 2 * 0.44 + 0|^2 + |0.7 - 2 * 0.8 + 0.1|^2 \approx 1.18$$

Such queries often can be expressed, or can be well approximated, by a sum over items in D of an *item function* that is applied to the tuple containing the values of the item in the different instances. Distinct count is a sum aggregate of logical OR and the L_p difference is the pth root of L_p^p, which sum-aggregates $|v_1 - v_2|^p$. For $r \geq 2$ instances, we can consider sum aggregates of the exponentiated range functions $\mathrm{RG}_p(\boldsymbol{v}) = (\max(\boldsymbol{v}) - \min(\boldsymbol{v}))^p$, where $p > 0$. This is made concrete in Example 1 which illustrates a data set of 3 instances over 8 items, example queries, specified over a selected set of items, and the corresponding item functions.

We now assume that each instance is sampled and the sample of each instance contains a subset of the items that were active in the instance (had a positive weight). Common sampling schemes for a single instance are Probability Proportional to Size (PPS) [24] or bottom-k sampling which includes Reservoir sampling [26, 36], Priority (Sequential Poisson) [31, 19], or Successive weighted sample without replacement [33, 20, 11]. The sampling of items in each instance can be completely independent or slightly dependent (as with Reservoir or bottom-k sampling, which samples exactly k items).

Coordinated sampling is a way of specifying the randomization so that the sampling of different instances utilizes the same "randomization" [2, 35, 6, 32, 34, 4, 3, 13, 28, 16]. That is, the sampling of the same item in different instances is highly correlated. This is in contrast to independent sampling. Alternative term used in the survey sampling literature is Permanent Random Numbers (PRN). Coordinated sampling is also a form of locality sensitive hashing (LSH): When the weights in two instances (rows) are similar, the samples we obtain are similar.

The method of coordinating samples had been rediscovered many times, for different applications, in both statistics and computer science. The main reason for its consideration by computer scientists is that it allows for more accurate estimates of queries that span multiple instances such as distinct counts and similarity measures [4, 3, 6, 17, 29, 21, 22, 5, 18, 1, 23, 28, 13, 16]. In some cases, such as all-distances sketches [6, 10, 29, 11, 12, 8] of neighborhoods of nodes in a graph, coordinated samples are obtained much more efficiently than independent samples. Coordination can be efficiently achieved by using a random hash function, applied to the item key k, to generate the seed $u^{(k)}$, in conjunction with the single-instance scheme of our choice (PPS or Reservoir). The use of hashing allows the sampling of different instances to be performed independently with very small shared state.

Coordinated PPS sampling of the instances in Example 1 is demonstrated in Example 2. The same random seed $u^{(k)}$ is used to determine the sampling of item k across instances.

Sum aggregates $\sum_{i\in D} f(\boldsymbol{v}^{(i)})$, such as L_p^p, over a domain of items D are estimated by summing up item function estimators over the selected items, that is $\sum_{i\in D} \hat{f}(S(\boldsymbol{v}^{(i)}, u^{(i)})$. The sampling of each item tuple is generally very sparse, with no entries or almost no entries sampled, and we therefore expect zero estimates $\hat{f} = 0$ for most items and a high variance. We therefore insist on unbiasedness and pairwise independence of the single-item estimates. That way,

$$\text{VAR}[\sum_{i\in D} \hat{f}(S(\boldsymbol{v}^{(i)}, u^{(i)}))] = \sum_{i\in D} \text{VAR}[\hat{f}(S(\boldsymbol{v}^{(i)}, u^{(i)}))] ,$$

the variance of the sum estimate is the sum over items in $i \in D$ of the variance of $\hat{f}$ for $\boldsymbol{v}^{(i)}$. Thus (assuming variance is balanced) we can expect the relative error to decrease $\propto 1/\sqrt{|D|}$. Lastly, since the functions we are interested in are nonnegative, we also require the estimates to be nonnegative.

Coordinated sampling, when projected on the tuple $\boldsymbol{v}$ of the weights of the item on the different instances (a column in our matrix) is a monotone sampling scheme [2] and the estimation problem of an item-function is a corresponding monotone estimation problem.

More precisely, the data domain in the monotone estimation problem we obtain is a subset of $r \geq 1$ dimensional vectors $\mathbf{V} \subset \mathbb{R}^r_{\geq 0}$ (where r is the number of instances in the query specification). The sampling is specified by r continuous non-decreasing functions on $(0, 1]$: $\boldsymbol{\tau} = \tau_1, \ldots, \tau_r$. The sample S includes the ith entry of $\boldsymbol{v}$ with its value v_i if and only $v_i \geq \tau_i(u)$. Note that when entry i is not sampled, we also have some information, as we know that $v_i < \tau_i(u)$. Therefore the set S^* of data vectors consistent with our sample (which we do not explicitly compute) includes the exact values of some entries and upper bounds on other entries. Since the functions τ_i are non-decreasing, the sampling scheme is monotone. In particular, PPS sampling of different instances, restricted to a single item, is expressed with $\tau_i(u)$ that are linear functions: There is a fixed vector $\boldsymbol{\tau}^*$ such that $\tau_i(u) \equiv u\tau_i^*$.

Therefore, the estimation of the sum-aggregate over coordinated samples is reduced to monotone estimation.

In [15] we provided a complete characterization of item-function estimation problems over coordinated samples for which estimators with desirable global properties exist. This characterization can be extended to monotone estimation. The properties considered were unbiasedness and nonnegativity, and together with finite variances or boundedness. We also showed that for any coordinated estimation problem for which an unbiased nonnegative estimator with finite variances exists, we can construct an estimator, which we named the J estimator, that is 84-competitive. The J estimator, however, is generally not admissible, and also, the construction was geared to establish $O(1)$ competitiveness rather than obtain a "natural" estimator or to minimize the constant.

Contributions

Our main technical contributions are the derivation of estimators for general monotone estimation problems. Our estimators are admissible, easy to apply, and satisfy desirable properties. We overview our results and provide pointers to examples and to the appropriate sections in the paper.

The optimal range: We start by studying the admissibility playing field for unbiased nonnegative estimators. We define the *optimal range* of estimates (Section 3) for each particular outcome, *conditioned* on the aggregate estimate over all "less-informative" outcomes (outcomes which correspond to larger seed value u). The range includes all estimate values that are "locally" optimal with respect to at least one data vector that is consistent with the outcome. We show that being "in range" almost everywhere is necessary for admissibility and is sufficient for unbiasedness and nonnegativity, when an unbiased nonnegative estimator exists.

The L* estimator: We study the estimator obtained when requiring the estimate on each outcome to be equal to the infimum of the optimal range. We name it the *L* estimator* and study it extensively in Section 4. The L* estimator can be expressed as the solution of a respective integral equation, which we solve to obtain a convenient form:

$$\hat{f}^{(L)}(S, \rho) = \frac{\underline{f}^{(\boldsymbol{v})}(\rho)}{\rho} - \int_\rho^1 \frac{\underline{f}^{(\boldsymbol{v})}(u)}{u^2} du , \qquad (1)$$

where ρ is the seed value used to obtain the sample S, $\boldsymbol{v} \in S^*$ is any (arbitrary) data vector consistent with S and ρ, and the *lower bound* function $\underline{f}^{(\boldsymbol{v})}(u)$ is defined as the infimum of $f(\boldsymbol{z})$ over all vectors $\boldsymbol{z} \in S^*(\boldsymbol{v}, u)$ that are consistent with the sample obtained for data $\boldsymbol{v}$ with seed u. Note that the right hand side is the same for any choice of $\boldsymbol{v} \in S^*$ and can be computed from the information in S and ρ. Therefore, the estimate is well defined and we can compute it by numeric integration or a closed form (when a respective definite integral has a closed form). The lower bound function is presented more precisely in Section 2 and an example is provided in Example 3. An example derivation of the L* estimator for the functions RG_{p+} is provided in Example 4.

We show that the L* estimator has a natural and compelling combination of properties. It satisfies both our quality measures, being both admissible and 4-competitive for any instance of the monotone estimation problem for which a bounded variance estimator exists. The competitive ratio of 4 improves over the previous upper bound of 84 [15]. We show that the ratio of 4 of the L* estimator is tight in the sense that there is a family of functions on which the supremum of the ratio, over functions and data vectors, is 4. We note however that the L* estimator has lower ratio for specific functions. For example, we computed ratios of 2 and 2.5, respectively, for exponentiated range with $p = 1, 2$ (Which facilitates estimation of L_p differences, see Example 1).

Moreover, the L* estimator is *monotone*, meaning that when fixing the data vector, the estimate value is monotone non-decreasing with the information in the outcome (the set S^* of data vectors that are consistent with our sample). In terms of our monotone sampling formulation, estimator monotonicity means that when we fix

[2] Bottom-k samples select exactly k items in each instance, hence inclusions of items are dependent. We obtain a single-item restriction by considering the sampling scheme for the item conditioned on fixing the seed values of other items. A similar situation is with all-distances sketches, where we can use the HIP inclusion probabilities [8], which are conditioned on fixing the randomization of all closer nodes.

the data $\boldsymbol{v}$, the estimate is non-increasing with the seed u. Furthermore, the L* estimator is the *unique* admissible monotone estimator and thus dominates (has at most the variance on every data vector) the Horvitz-Thompson (HT) estimator [25] (which is also unbiased, nonnegative, and monotone).

To further illustrate comparison with HT, recall that the HT estimate is positive only on outcomes that reveal $f(\boldsymbol{v})$. In this case, the inverse probability estimate $f(\boldsymbol{v})/p$, where p is the probability of an outcome which reveals $f(\boldsymbol{v})$. When we have partial information on $f(\boldsymbol{v})$, the HT estimate does not utilize that and is 0 whereas admissible estimators, such as the L* estimators, must use this information. It is also possible that the probability of an outcome that reveals $f(\boldsymbol{v})$ is 0. In this case, the HT estimator is not even applicable. One natural such example is the range $|v_1 - v_2|$ with coordinated PPS sampling using $\boldsymbol{\tau}^* = (1, 1)$. When the input is $(0.5, 0)$, the range is 0.5, but there is 0 probability of revealing $v_2 = 0$. We can obtain informative lower (and upper) bounds on the range: When $u \in (0, 0.5)$, we have a lower bound of $0.5 - u$. Nonetheless, the probability of knowing the exact value ($u = 0$) is 0. In contrast to the HT estimate, our L* estimator is defined for any monotone estimation instance for which a nonnegative unbiased estimator with finite variance exists.

Order-optimal estimators: In many situations we have information on data patterns. For example, if our data consists of hourly temperature measurements across locations or daily summaries of Wikipedia, we expect it to be fairly stable. That is, we expect instances to be very similar. That is, most tuples of values , each corresponding to a particular geographic location or Wikipedia article, would have most entries being very similar. In other cases, such as IP traffic, differences are typically larger. Since there is a choice, a Pareto front of admissible estimators, we would like to be able to select an estimator that would have lower variance on more likely patterns of data vectors, this while still providing some weaker "worst case" guarantees for all applicable data vectors in our domain.

Customization of estimators to data patterns can be facilitated through *order optimality* [14]. More precisely, an estimator is $\prec^+$-optimal with respect to some partial order $\prec$ on data vectors if any other (nonnegative unbiased) estimator with lower variance on some data $\boldsymbol{v}$ must have strictly higher variance on some data that precedes $\boldsymbol{v}$. Order-optimality implies admissibility, but not vice versa. By specifying an order which prioritizes more likely patterns in the data, we can customize the estimator to these patterns.

We show (Section 5) how to construct a $\prec^+$-optimal nonnegative unbiased estimators for *any* function and order $\prec$ for which such estimator exists. We show that when the data domain is discrete, such estimators always exist whereas continuous domains require some natural convergence properties of $\prec$.

We also show that the L* estimator is $\prec^+$-optimal with respect to the order $\prec$ such that $\boldsymbol{z} \prec \boldsymbol{v} \iff f(\boldsymbol{z}) < f(\boldsymbol{v})$. This means that when estimating the exponentiated range function, the L* estimator is optimized for high similarity (this while providing a strong 4-competitiveness guarantee even for highly dissimilar data).

The U* estimator: We also explore the upper extreme of the optimal range, that is, the solution obtained by aiming for the supremum of the range. We call this solution the *U* estimator* and we study it in Section 6. This estimator is unbiased, nonnegative, and has finite variances. We formulate some conditions on the monotone estimation problem which are satisfied by natural functions including the exponentiated range with coordinated PPS sampling , under which the estimator is admissible. The U* estimator, under some conditions, is $\prec^+$-optimal with respect to the order $\boldsymbol{z} \prec \boldsymbol{v} \iff f(\boldsymbol{z}) > f(\boldsymbol{v})$. In the context of the exponentiated range, it means that it is optimized for highly dissimilar instances.

Lastly, in Section 7 we conclude with a discussion of future work and of follow-up applications of our estimators.

2. PRELIMINARIES

Consider monotone sampling, as defined in the introduction.

LEMMA 2.1. *For any two outcomes, $S_1^* = S^*(u, \boldsymbol{v})$ and $S_2^* = S^*(u', \boldsymbol{v}')$, the sets S_1^* and S_2^* must be either disjoint or one is contained in the other.*

PROOF. If there is a common data vector $\boldsymbol{z} \in S_1^* \cap S_2^*$, then $S_1^* = S^*(u, \boldsymbol{z})$ and $S_2^* = S^*(u', \boldsymbol{z})$. From definition, if $u' > u$ then $S_1^* \subseteq S_2^*$ and vice versa. □

LEMMA 2.2. *For any $\boldsymbol{v}, \boldsymbol{z} \in \mathbf{V}$, the set of u values which satisfy $S^*(u, \boldsymbol{v}) = S^*(u, \boldsymbol{z})$ is a suffix of the interval $(0, 1]$.*

PROOF. $S^*(u, \boldsymbol{v}) = S^*(u, \boldsymbol{z})$ implies $S^*(u', \boldsymbol{v}) = S^*(u', \boldsymbol{z})$ for all $u' > u$. □

For convenience, we assume without loss of generality that this interval is open to the left. This assumption is made without loss of generality since it can be integrated while affecting at most a "zero probability" set of outcomes for any data point.

$$\forall \rho \in (0, 1]\ \forall \boldsymbol{v}, \qquad (2)$$
$$\boldsymbol{z} \in S^*(\rho, \boldsymbol{v}) \implies \exists \epsilon > 0,\ \forall x \in (\rho - \epsilon, 1],\ \boldsymbol{z} \in S^*(x, \boldsymbol{v})$$

Example 2 Coordinated PPS sampling for Example 1

We show shared-seed coordinated sampling, where each of the instances 1,2,3 is PPS sampled with $\tau^* = 1$. Therefore, each entry is sampled with probability equal to its value. We draw $u^{(k)} \in U[0, 1]$, independently for different items. An item k is sampled in instance i if and only if $v_i^{(k)} \geq u^{(k)}$. $S^{*(k)}$ contains all vectors consistent with the sampled entries and with value at most $u^{(k)}$ in unsampled entries.

item	a	b	c	d	e	f	g	h
v_1	0.95	0	0.23	0.70	0.10	0.42	0	0.32
v_2	0.15	0.44	0	0.80	0.05	0.50	0.20	0
v_3	0.25	0	0	0.10	0	0.22	0	0
$u^{(k)}$	0.32	0.21	0.04	0.23	0.84	0.70	0.15	0.64

The outcomes for the different items are: $S^{(a)} = (0.95, *, *)$, $S^{(b)} = (*, 0.44, *)$, $S^{(c)} = (0.23, *, *)$, $S^{(d)} = (0.7, 0.8, *)$, $S^{(e)} = S^{(f)} = S^{(h)} = (*, *, *)$, $S^{(g)} = (*, 0.2, *)$. The sets of vectors consistent with the outcomes are $S^{*(a)} = \{0.95\} \times [0, 0.32)^2$ and $S^{*(h)} = [0, 0.64)^3$.

Estimators: Given a monotone estimation problem, we are interested in estimating $f(\boldsymbol{v})$. An estimator $\hat{f}$ is a function of the outcome (including the seed) $S(u, \boldsymbol{v})$. We use the notation $\hat{f}(u, \boldsymbol{v}) \equiv \hat{f}(S(u, \boldsymbol{v}))$. When the domain is continuous, we only consider $\hat{f}$ that are (Lebesgue) integrable. Two estimators $\hat{f}_1$ and $\hat{f}_2$ are *equivalent* if for all data $\boldsymbol{v}$, $\hat{f}_1(u, \boldsymbol{v}) = \hat{f}_2(u, \boldsymbol{v})$ with probability 1, which is the same as

$$\hat{f}_1 \text{ and } \hat{f}_2 \text{ are equivalent} \iff \forall \boldsymbol{v} \forall \rho \in (0, 1], \qquad (3)$$
$$\lim_{\eta \to \rho^-} \frac{\int_\eta^\rho \hat{f}_1(u, \boldsymbol{v}) du}{\rho - \eta} = \lim_{\eta \to \rho^-} \frac{\int_\eta^\rho \hat{f}_2(u, \boldsymbol{v}) du}{\rho - \eta}.$$

An estimator $\hat{f}$ is *nonnegative* if $\forall S,\ \hat{f}(S) \geq 0$ and is *unbiased* if $\forall \boldsymbol{v},\ \mathsf{E}_{u\sim U[0,1]}[\hat{f}(u,\boldsymbol{v})] = f(\boldsymbol{v})$. An estimator has *finite variance* on $\boldsymbol{v}$ if $\int_0^1 \hat{f}(u,\boldsymbol{v})^2 du < \infty$ (the expectation of the square is finite) and is *bounded* on $\boldsymbol{v}$ if $\sup_{u\in(0,1]} \hat{f}(u,\boldsymbol{v}) < \infty$. If a nonnegative estimator is bounded on $\boldsymbol{v}$, it also has finite variance for $\boldsymbol{v}$. An estimator is *monotone* on $\boldsymbol{v}$ if when fixing $\boldsymbol{v}$ and considering outcomes consistent with $\boldsymbol{v}$, the estimate value is non decreasing with the information on the data that we can glean from the outcome, that is, $\hat{f}(u,\boldsymbol{v})$ is non-increasing with u. We say that an estimator is bounded, has finite variances, or is monotone, if the respective property holds for all $\boldsymbol{v} \in \mathbf{V}$.

The lower bound function. For $Z \subset \mathbf{V}$, we define $\underline{f}(Z) = \inf\{f(v) \mid v \in Z\}$ as the infimum of f on Z. We use the notation $\underline{f}(S) \equiv \underline{f}(S^*)$, $\underline{f}(\rho,\boldsymbol{v}) \equiv \underline{f}(S^*(\rho,\boldsymbol{v}))$. When $\boldsymbol{v}$ is fixed, we use $\underline{f}^{(\boldsymbol{v})}(u) \equiv \underline{f}(u,\boldsymbol{v})$. Some properties which we need in the sequel are [15]:

$$\bullet \forall \boldsymbol{v},\ \underline{f}^{(\boldsymbol{v})}(u) \text{ is monotone non increasing and left-continuous.} \tag{4}$$

$$\bullet \hat{f} \text{ is unbiased and nonnegative } \implies \tag{5}$$

$$\forall \boldsymbol{v}, \forall \rho,\ \int_\rho^1 \hat{f}(u,\boldsymbol{v})du \leq \underline{f}^{(\boldsymbol{v})}(\rho)\ . \tag{6}$$

The lower bound function $\underline{f}^{(\boldsymbol{v})}$, and its lower hull $H_f^{(\boldsymbol{v})}$, are used to characterize existence of estimators with desirable properties [15]:

$$\bullet \exists \text{ unbiased nonnegative } f \text{ estimator } \iff \tag{7}$$

$$\forall \boldsymbol{v} \in \mathbf{V},\ \lim_{u\to 0^+} \underline{f}^{(\boldsymbol{v})}(u) = f(\boldsymbol{v})\ . \tag{8}$$

$\bullet$If f satisfies (8),

$\exists$ unbiased nonnegative estimator with finite variance for $\boldsymbol{v}$

$$\iff \int_0^1 \left(\frac{dH_f^{(\boldsymbol{v})}(u)}{du}\right)^2 du < \infty\ . \tag{9}$$

$\exists$ unbiased nonnegative estimator that is bounded on $\boldsymbol{v}$

$$\iff \lim_{u\to 0^+} \frac{f(\boldsymbol{v}) - \underline{f}^{(\boldsymbol{v})}(u)}{u} < \infty\ . \tag{10}$$

Example 3 illustrates lower bound functions and respective lower hulls for RG_{p+}. **Partially specified estimators.** We use *partial specifications* $\hat{f}$ of (nonnegative and unbiased) estimators, which are specified on a set of outcomes $\mathcal{S}$ so that

$$\forall \boldsymbol{v}\ \exists \rho_v \in [0,1],\ S(u,\boldsymbol{v}) \in \mathcal{S} \textit{ almost everywhere} \text{ for } u > \rho_v \ \wedge$$
$$S(u,\boldsymbol{v}) \notin \mathcal{S} \textit{ almost everywhere} \text{ for } u \leq \rho_v\ .$$

When $\rho_v = 0$, we say that the estimator is *fully specified* for $\boldsymbol{v}$. We also require that $\hat{f}$ is nonnegative where specified and satisfies

$$\forall \boldsymbol{v},\ \rho_v > 0 \quad \implies \quad \int_{\rho_v}^1 \hat{f}(u,\boldsymbol{v})du \leq f(\boldsymbol{v}) \tag{11a}$$

$$\forall \boldsymbol{v},\ \rho_v = 0 \quad \implies \quad \int_{\rho_v}^1 \hat{f}(u,\boldsymbol{v})du = f(\boldsymbol{v})\ . \tag{11b}$$

LEMMA 2.3. *[15] If f satisfies* (8) *(has a nonnegative unbiased estimator), then any partially specified estimator can be extended to an unbiased nonnegative estimator.*

$\boldsymbol{v}$-optimal extensions and estimators. Given a partially specified estimator $\hat{f}$ so that $\rho_v > 0$ and $M = \int_{\rho_v}^1 \hat{f}(u,\boldsymbol{v})du$, a *$\boldsymbol{v}$-optimal extension* is an extension which is fully specified for $\boldsymbol{v}$ and minimizes variance for $\boldsymbol{v}$ (amongst all such extensions). The $\boldsymbol{v}$-optimal extension is defined on outcomes $S(u,\boldsymbol{v})$ for $u \in (0,\rho_v]$ and satisfies

$$\begin{aligned} \min & \int_0^{\rho_v} \hat{f}(u,\boldsymbol{v})^2 du \\ \text{s.t.} & \int_0^{\rho_v} \hat{f}(u,\boldsymbol{v})du = f(\boldsymbol{v}) - M \\ & \forall u, \int_u^{\rho_v} \hat{f}(x,\boldsymbol{v})dx \leq \underline{f}^{(\boldsymbol{v})}(u) - M \\ & \forall u, \hat{f}(u,\boldsymbol{v}) \geq 0 \end{aligned} \tag{12}$$

For $\rho_v \in (0,1]$ and $M \in [0, \underline{f}^{(\boldsymbol{v})}(\rho_v)]$, we define the function $\hat{f}^{(\boldsymbol{v},\rho_v,M)} : (0,\rho_v] \to R_+$ as the solution of

$$\hat{f}^{(\boldsymbol{v},\rho_v,M)}(u) = \inf_{0\leq\eta<u} \frac{\underline{f}^{(\boldsymbol{v})}(\eta) - M - \int_u^{\rho_v} \hat{f}^{(\boldsymbol{v},\rho_v,M)}(u)du}{\rho - \eta}\ . \tag{13}$$

Geometrically, the function $\hat{f}^{(\boldsymbol{v},\rho_v,M)}$ is the negated derivative of the lower hull of the lower bound function $\underline{f}^{(\boldsymbol{v})}$ on $(0,\rho_v)$ and the point (ρ_v, M).

THEOREM 2.1. *[15] Given a partially specified estimator $\hat{f}$ so that $\rho_v > 0$ and $M = \int_{\rho_v}^1 \hat{f}(u,\boldsymbol{v})du$, then $\hat{f}^{(\boldsymbol{v},\rho_v,M)}$ is the unique (up to equivalence) $\boldsymbol{v}$-optimal extension of $\hat{f}$.*

The *$\boldsymbol{v}$-optimal* estimates are the minimum variance extension of the empty specification. We use $\rho_v = 1$ and $M = 0$ and obtain $\hat{f}^{(\boldsymbol{v})} \equiv \hat{f}^{(\boldsymbol{v},1,0)}$. $\hat{f}^{(\boldsymbol{v})}$ is the solution of

$$\hat{f}^{(\boldsymbol{v})}(u) = \inf_{0\leq\eta<u} \frac{\underline{f}^{(\boldsymbol{v})}(\eta) - \int_u^1 \hat{f}^{(\boldsymbol{v})}(u)du}{\rho - \eta}\ , \tag{14}$$

which is the negated slope of the lower hull of the lower bound function $\underline{f}^{(\boldsymbol{v})}$. This is illustrated in Example 3.

Admissibility and order optimality. An estimator is *admissible* if there is no (nonnegative unbiased) estimator with same or lower variance on all data and strictly lower on some data. We also consider *order optimality*, specified with respect to a partial order $\prec$ on $\mathbf{V}$: An estimator $\hat{f}$ is *$\prec^+$-optimal* if there is no other nonnegative unbiased estimator with strictly lower variance on some data $\boldsymbol{v}$ and at most the variance of $\hat{f}$ on all vectors that precede $\boldsymbol{v}$. Order-optimality (with respect to some $\prec$) implies admissibility but the converse is not true in general [14].

Variance competitiveness [15] An estimator $\hat{f}$ is *c-competitive* if

$$\forall \boldsymbol{v},\ \int_0^1 \left(\hat{f}(u,\boldsymbol{v})\right)^2 du \leq c \inf_{\hat{f}'} \int_0^1 \left(\hat{f}'(u,\boldsymbol{v})\right)^2 du,$$

where the infimum is over all unbiased nonnegative estimators of f. When the estimator is unbiased, the expectation of the square is closely related to variance, and an estimator that minimizes one also minimizes the other.

$$\mathrm{VAR}[\hat{f}|\boldsymbol{v}] = \int_0^1 \hat{f}(u,\boldsymbol{v})^2 du - f(\boldsymbol{v})^2 \tag{15}$$

3. THE OPTIMAL RANGE

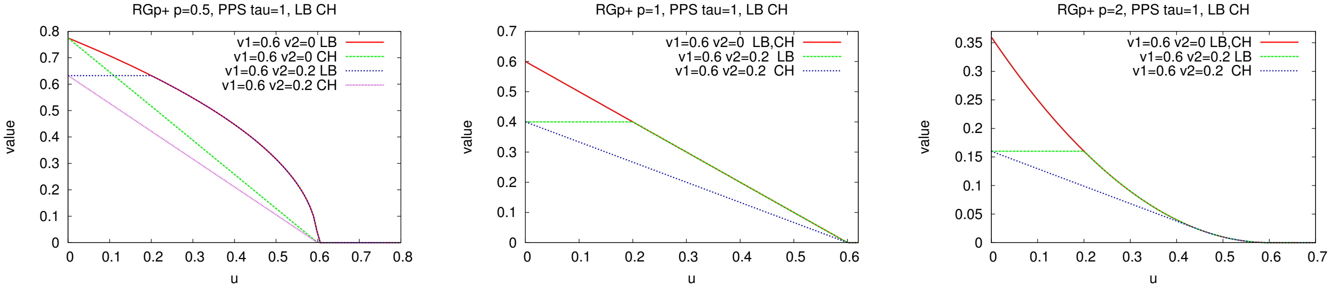
We say that an estimator $\hat{f}$ is $\boldsymbol{v}$-optimal *at an outcome* $S(u,\boldsymbol{v})$ if it satisfies (14). For an outcome $S(\rho,\boldsymbol{v})$, we are interested in the *range* of $\boldsymbol{z}$-optimal estimates at S for all $\boldsymbol{z} \in S^*$, with respect to a value M, which captures the contribution to the expectation of the estimator made by outcomes which are less informative than S.

Example 3 Lower bound function and its lower hull

Consider $\text{RG}_{p+}(v_1, v_2) = \max\{0, v_1 - v_2\}^p$ (see Example 1) over the domain $\mathbf{V} = [0,1]^2$ and PPS sampling with $\tau_1^* = \tau_2^* = 1$ (as in Example 2). The lower bound function for data $\boldsymbol{v} = (v_1, v_2)$ is

$$\underline{\text{RG}_{p+}}(u, \boldsymbol{v}) = \max\{0, v_1 - \max\{v_2, u\}\}^p \ .$$

The figures below illustrate $\underline{\text{RG}_{p+}}^{(\boldsymbol{v})}(u)$ (LB) and its lower hull (CH) for the data vectors $(0.6, 0.2)$ and $(0.6, 0)$ and $p = \{0.5, 1, 2\}$. For $u > 0.2$, the outcome when sampling both vectors is the same, and thus the lower bound function is the same. For $u \leq 0.2$, the outcomes diverge. For $p \leq 1$, $\underline{\text{RG}_{p+}}^{(\boldsymbol{v})}(u)$ is concave and the lower hull is linear on $(0, v_1]$. For $p > 1$, the lower hull coincides with $\underline{\text{RG}_{p+}}^{(\boldsymbol{v})}(u)$ on some interval $(a, v_1]$ and is linear on $(0, a]$. When $v_2 = 0$, $\underline{\text{RG}_{p+}}^{(\boldsymbol{v})}(u)$ is equal to its lower hull.



The $\boldsymbol{v}$-optimal estimates are the negated slopes of the lower hulls. They are 0 when $u \in (0.6, 1]$, since these outcomes are consistent with data on which $\underline{\text{RG}_{p+}} = 0$. They are constant for $u \in (0, v_1]$ when $p \leq 1$. Observe that for $u \in (0.2, 0.6]$, the $\boldsymbol{v}$-optimal estimates are different even though the outcome of sampling the two vectors are the same – demonstrating that it is not possible to simultaneously minimize the variance of the two vectors.

$$\lambda(\rho, \boldsymbol{v}, M) = \inf_{0 \leq \eta < \rho} \frac{\underline{f}(\eta, \boldsymbol{v}) - M}{\rho - \eta} \qquad (16)$$

$$\lambda_U(\rho, \boldsymbol{v}, M) \equiv \lambda_U(S, M) = \sup_{\boldsymbol{z} \in S^*(\rho, \boldsymbol{v})} \lambda(\rho, \boldsymbol{z}, M) \qquad (17)$$

$$\begin{aligned} \lambda_L(\rho, \boldsymbol{v}, M) \equiv \lambda_L(S, M) &= \inf_{\boldsymbol{z} \in S^*(\rho, \boldsymbol{v})} \lambda(\rho, \boldsymbol{z}, M) \\ &= \inf_{\boldsymbol{z} \in S^*(\rho, \boldsymbol{v})} \inf_{0 \leq \eta < \rho} \frac{\underline{f}(\eta, \boldsymbol{z}) - M}{\rho - \eta} \\ &= \frac{\underline{f}(\rho, \boldsymbol{v}) - M}{\rho} \qquad (18) \end{aligned}$$

To verify equality (18), observe that from left continuity of $\underline{f}(u, \boldsymbol{z})$,

$$\inf_{\eta < \rho, \ \boldsymbol{z} \in S^*} \underline{f}(\eta, \boldsymbol{z}) = \underline{f}(\rho, \boldsymbol{v})$$

and that the denominator $\rho - \eta$ is maximized at $\eta = 0$. $\lambda(\rho, \boldsymbol{v}, M)$ is the $\boldsymbol{v}$-optimal estimate at ρ, given a specification of the estimator $\hat{f}(u, \boldsymbol{v})$ for $u \in (\rho, 1]$ with $\int_\rho^1 \hat{f}(u, \boldsymbol{v}) du = M$. In short, we refer to $\lambda(\rho, \boldsymbol{v}, M)$ as the $\boldsymbol{v}$-optimal estimate at ρ given M. Geometrically, $\lambda(\rho, \boldsymbol{v}, M)$ is the negated slope of the lower hull of $\underline{f}^{(\boldsymbol{v})}$ and the point (ρ, M). $\lambda_U(S, M)$ and $\lambda_L(S, M)$, respectively, are the supremum and infimum of the *range* of $\boldsymbol{z}$-optimal estimates at S given M. Figure 1 illustrates an outcome S and the optimal range at S given M. We can see how the lower endpoint of the range is realized by a vector with f value equal to the lower bound at S, as in equality (18).

When $\hat{f}$ is provided for seed values $u \in (\rho, 1]$, we use $M = \int_\rho^1 \hat{f}(u, \boldsymbol{v}) du$. We then abbreviate the notations (we remove M) to $\lambda(\rho, \boldsymbol{v})$, $\lambda_U(S)$, and $\lambda_L(S)$.

We say that the estimator $\hat{f}$ is *in-range* (in the optimal range) at outcome $S(\rho, \boldsymbol{v})$ if

$$\lambda_L(S) \leq \hat{f}(S) \leq \lambda_U(S) \ . \qquad (19)$$

Writing (19) explicitly, we obtain

$$\begin{aligned} \hat{f}(\rho, \boldsymbol{v}) &\geq \lambda_L(\rho, \boldsymbol{v}) = \frac{\underline{f}(\rho, \boldsymbol{v}) - \int_\rho^1 \hat{f}(u, \boldsymbol{v}) du}{\rho} \qquad (20a) \\ \hat{f}(\rho, \boldsymbol{v}) &\leq \lambda_U(\rho, \boldsymbol{v}) \\ &= \sup_{\boldsymbol{z} \in S^*} \inf_{0 \leq \eta < \rho} \frac{\underline{f}(\eta, \boldsymbol{z}) - \int_\rho^1 \hat{f}(u, \boldsymbol{v}) du}{\rho - \eta} \qquad (20b) \end{aligned}$$

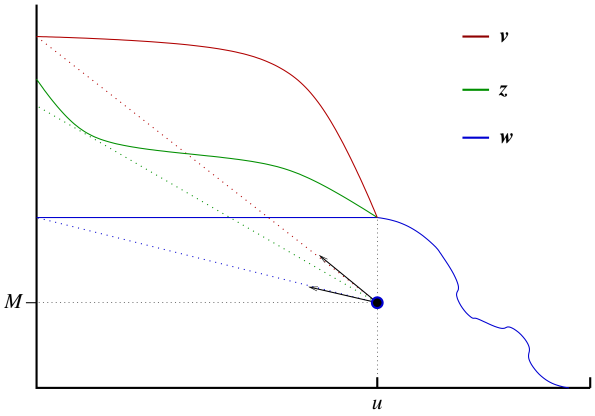


Figure 1: Lower bound functions for vectors $\boldsymbol{v}, \boldsymbol{z}, \boldsymbol{w}$. Outcomes are consistent for all $x \geq u$: $S(x, \boldsymbol{v}) = S(x, \boldsymbol{z}) = S(x, \boldsymbol{w}) \equiv S_x$. The figure illustrates the $\boldsymbol{y}$-optimal estimates $\lambda(u, \boldsymbol{y}, M)$ at u given M for y $\in \{\boldsymbol{v}, \boldsymbol{z}, \boldsymbol{w}\}$. The estimates are the negated slopes of the lower hull of the point (u, M) and the lower bound function $\underline{f}^{(\boldsymbol{y})}$. The optimal range at S_u given M is lower-bounded by $\boldsymbol{w}$, that is $\lambda_L(S_u, M) = \lambda(u, \boldsymbol{w}, M)$, and upper-bounded by $\boldsymbol{v}$, $\lambda_U(S_u, M) = \lambda(u, \boldsymbol{v}, M)$. The figure illustrates the general property that the optimal range is lower bounded by the $\boldsymbol{w}$ which satisfies $f(\boldsymbol{w}) = \underline{f}(\boldsymbol{w}, u)$.

Two special solutions that we study are the L* *estimator* ($\hat{f}^{(L)}$, see Section 4) and the U* *estimator* ($\hat{f}^{(U)}$, see Section 6), which respectively solve (20a) and (20b) with equalities. For all $\rho \in (0,1]$ and $\boldsymbol{v}$, $\hat{f}^{(L)}$ minimizes and $\hat{f}^{(U)}$ maximizes $\int_\rho^1 \hat{f}(u,\boldsymbol{v})du$ among all solutions of (19).

We show that being in-range (satisfying (19) for all outcomes S) is sufficient for nonnegativity and unbiasedness.

LEMMA 3.1. *If f satisfies* (8) *then any in-range estimator is unbiased and nonnegative.*

PROOF. For nonnegativity, it suffices to show that a solution of (19) satisfies (6), since (20a) and (6) together imply nonnegativity. Assume to the contrary that a solution $\hat{f}$ violates (6) and let ρ be the supremum of x satisfying $\int_x^1 \hat{f}(u,\boldsymbol{v})du > \underline{f}(x,\boldsymbol{v})$. From (4), which is monotonicity and left-continuity of $\underline{f}(x,\boldsymbol{v})$, we have $\int_\rho^1 \hat{f}(u,\boldsymbol{v})du = \underline{f}(\rho,\boldsymbol{v})$. Since $\int_x^1 \hat{f}(u,\boldsymbol{v})du$ is continuous in x, and $\underline{f}^{(\boldsymbol{v})}$ left-continuous, there must be $\delta > 0$ so that

$$\forall x \in [\rho-\delta,\rho), \int_x^1 \hat{f}(u,\boldsymbol{v})du > \underline{f}(x,\boldsymbol{v}) \,. \tag{21}$$

Let $x \in [\rho-\delta,\rho)$ and $M(x) = \int_x^1 \hat{f}(u,\boldsymbol{v})du$. From (21), $M(x) > \underline{f}(x,\boldsymbol{v})$. We have that

$$\begin{aligned}\hat{f}(x,\boldsymbol{v}) &\leq \sup_{\boldsymbol{z}\in S^*(x,\boldsymbol{v})} \inf_{0\leq\eta<x} \frac{\underline{f}(\eta,\boldsymbol{z})-M(x)}{x-\eta} \\ &\leq \sup_{\boldsymbol{z}\in S^*(x,\boldsymbol{v})} \inf_{0\leq\eta<x} \frac{\underline{f}(\eta,\boldsymbol{z})-\underline{f}(x,\boldsymbol{v})}{x-\eta} \\ &\leq \sup_{\boldsymbol{z}\in S^*(x,\boldsymbol{v})} \lim_{\eta\to x^-} \frac{\underline{f}(\eta,\boldsymbol{z})-\underline{f}(x,\boldsymbol{v})}{x-\eta} \\ &= \lim_{\eta\to x^-} \frac{\underline{f}(\eta,\boldsymbol{v})-\underline{f}(x,\boldsymbol{v})}{x-\eta} = -\frac{\partial \underline{f}(x,\boldsymbol{v})}{\partial x^-}\end{aligned}$$

Since this holds for all $x \in (\rho-\delta,\rho)$, we obtain that

$$\int_{\rho-\delta}^{\rho} \hat{f}(x,\boldsymbol{v})dx \leq \underline{f}(\rho-\delta,\boldsymbol{v}) - \underline{f}(\rho,\boldsymbol{v}) \,.$$

Therefore, $\int_{\rho-\delta}^1 \hat{f}(x,\boldsymbol{v})dx \leq \underline{f}(\rho-\delta,\boldsymbol{v})$, which contradicts (21).

We now establish unbiasedness. From (20a) and $\underline{f}(u,\boldsymbol{v})$ being non increasing in u, we obtain that $\forall u \forall \rho > u$,

$$\begin{aligned}\hat{f}(u,\boldsymbol{v}) &\geq \frac{\underline{f}(u,\boldsymbol{v}) - \int_u^1 \hat{f}(x,\boldsymbol{v})dx}{u} \\ &\geq \frac{\underline{f}(\rho,\boldsymbol{v}) - \int_u^1 \hat{f}(x,\boldsymbol{v})dx}{u}\end{aligned} \tag{22}$$

We argue that

$$\forall \boldsymbol{v} \forall \rho > 0, \lim_{x\to 0} \int_x^1 \hat{f}(u,\boldsymbol{v})du \geq \underline{f}(\rho,\boldsymbol{v}) \,. \tag{23}$$

To prove (23), define $\Delta(x) = \underline{f}(\rho,\boldsymbol{v}) - \int_x^1 \hat{f}(u,\boldsymbol{v})du$ for $x \in (0,\rho]$. We show that $\int_{x/2}^x \hat{f}(u,\boldsymbol{v})du \geq \Delta(x)/4$. To see this, assume to the contrary that $\int_y^x \hat{f}(u,\boldsymbol{v})du \leq \Delta(x)/4$ for all $y \in [x/2,x]$. Then from (22), the value of $\hat{f}(u,\boldsymbol{v})$ for $u \in [x/2,x]$ must be at least $(3/4)\Delta(x)/x$. Hence, the integral over the interval $[x/2,x]$ is at least $(3/8)\Delta(x)$ which is a contradiction. We can now apply this iteratively, obtaining that $\Delta(\rho/2^i) \leq (3/4)^i\Delta(\rho)$. Thus, the gap $\Delta(x)$ diminishes as $x \to 0$ and we established (23).

Since (23) holds for all $\rho \geq 0$, then $\lim_{u\to 0}\int_u^1 \hat{f}(u,\boldsymbol{v})du \geq \lim_{u\to 0} \underline{f}(u,\boldsymbol{v}) = f(\boldsymbol{v})$ (using (8)). Combining with (already established) (6) we obtain $\lim_{u\to 0}\int_u^1 \hat{f}(u,\boldsymbol{v})du = f(\boldsymbol{v})$. □

We next show that being in-range is necessary for optimality. For our analysis of order-optimality (Section 5), we need to slightly refine the notion of admissibility to be with respect to a partially specified estimator $\hat{f}$ and a subset of data vectors $Z \subset \mathbf{V}$.

An extension of $\hat{f}$ that is fully specified for all vectors in Z is admissible on Z if any other extension with strictly lower variance on at least one $\boldsymbol{v} \in Z$ has a strictly higher variance on at least one $\boldsymbol{z} \in Z$. We say that a partial specification is in-range *with respect to* Z if:

$$\forall \boldsymbol{v} \in Z, \text{ for } \rho \in (0,\rho_v] \text{ almost everywhere,}$$
$$\inf_{\boldsymbol{z}\in Z\cap S^*(\rho,\boldsymbol{v})} \lambda(\rho,\boldsymbol{z}) \leq \hat{f}(\rho,\boldsymbol{v}) \leq \sup_{\boldsymbol{z}\in Z\cap S^*(\rho,\boldsymbol{v})} \lambda(\rho,\boldsymbol{z}) \tag{24}$$

Using (3), (24) is the same as requiring that $\forall \boldsymbol{v}\, \forall \rho \in (0,\rho_v]$, when fixing the estimator on $S(u,\boldsymbol{v})$ for $u \geq \rho$, then

$$\inf_{\boldsymbol{z}\in Z\cap S^*(\rho,\boldsymbol{v})} \lambda(\rho,\boldsymbol{z}) \leq \lim_{\eta\to\rho^-} \frac{\int_\eta^\rho \hat{f}(u,\boldsymbol{v})du}{\rho-\eta} \leq \sup_{\boldsymbol{z}\in Z\cap S^*(\rho,\boldsymbol{v})} \lambda(\rho,\boldsymbol{z}) \tag{25}$$

We show that a necessary condition for admissibility with respect to a partial specification and Z is that almost everywhere, estimates for outcomes consistent with vectors in Z are in-range for Z. Formally:

THEOREM 3.1. *An extension is admissible on Z only if* (24) *holds.*

PROOF. Consider an (nonnegative unbiased) estimator $\hat{f}$ that violates (24) for some $\boldsymbol{v} \in Z$ and ρ. We show that there is an alternative estimator, equal to $\hat{f}(u,\boldsymbol{v})$ on outcomes $u > \rho$ and which satisfies (24) at ρ that has strictly lower variance than $\hat{f}$ on all vectors $Z \cap S^*(\rho,\boldsymbol{v})$. This will show that $\hat{f}$ is not admissible on Z.

The estimator $\hat{f}$ violates (25), so either

$$\lim_{\eta\to\rho^-} \frac{\int_\eta^\rho \hat{f}(u,\boldsymbol{v})du}{\rho-\eta} < \inf_{\boldsymbol{z}\in Z\cap S^*(\rho,\boldsymbol{v})} \lambda(\rho,\boldsymbol{z}) \equiv L \tag{26}$$

or

$$\lim_{\eta\to\rho^-} \frac{\int_\eta^\rho \hat{f}(u,\boldsymbol{v})du}{\rho-\eta} > \sup_{\boldsymbol{z}\in Z\cap S^*(\rho,\boldsymbol{v})} \lambda(\rho,\boldsymbol{z}) \equiv U \,. \tag{27}$$

Violation (27), for a nonnegative unbiased $\hat{f}$, means that $M \equiv \int_\rho^1 \hat{f}(u,\boldsymbol{v})du < \underline{f}(u,\boldsymbol{v})$. Consider $\boldsymbol{z} \in Z \cap S^*(\rho,\boldsymbol{v})$ and the $\boldsymbol{z}$-optimal extension, $\hat{f}^{(\boldsymbol{z},\rho,M)}$ (see Theorem 2.1). Because the point (ρ, M) lies strictly below $\underline{f}^{(\boldsymbol{z})}$, the lower hull of both the point and $\underline{f}^{(\boldsymbol{z})}$ has a linear piece on some interval with right end point ρ. More precisely, $\hat{f}^{(\boldsymbol{z},\rho,M)}(u) \equiv \lambda(\rho,\boldsymbol{z},M)$ on $S(u,\boldsymbol{z})$ at some nonempty interval $u \in (\eta_z,\rho]$ so that at the point η_z, the lower bound is met, that is, $M+(\rho-\eta_z)\lambda(\rho,\boldsymbol{z},M) = \lim_{u\to\eta_z^+} \underline{f}(u,\boldsymbol{z})$. Therefore, all extensions (maintaining nonnegativity and unbiasedness) must satisfy

$$\begin{aligned}\int_{\eta_z}^\rho \hat{f}(u,\boldsymbol{z})du &\leq \lim_{u\to\eta_z^+} \underline{f}(u,\boldsymbol{z}) - M \\ &= (\rho-\eta_z)\lambda(\rho,\boldsymbol{z},M) \leq (\rho-\eta_z)U \,.\end{aligned} \tag{28}$$

From (27), for some $\epsilon > 0$, $\hat{f}$ has average value strictly higher than U on $S(u,\boldsymbol{v})$ for all u in $(\eta,\rho]$ for $\eta \in [\rho-\epsilon,\rho)$. For each $\boldsymbol{z} \in S^*(\rho, \boldsymbol{v})$ we define ζ_z as the maximum of $\rho - \epsilon$ and $\inf\{u \mid S^*(u, \boldsymbol{v}) = S^*(u, \boldsymbol{z})\}$. From (2), $\zeta_z < \rho$. For each $\boldsymbol{z}$, the higher estimate values on $S(u, \boldsymbol{z})$ for $u \in (\zeta_z, \rho]$ must be "compensated for" by lower values on $u \in (\eta_z, \zeta_z)$ (from nonnegativity we must have the strict inequality $\eta_z < \zeta_z$) so that (28) holds. By modifying the estimator to be equal to U for all outcomes $S(u, \boldsymbol{v})$ $u \in (\rho - \epsilon, \rho]$ and correspondingly increasing some estimate values that are lower than U to U on $S(u, \boldsymbol{z})$ for $u \in (\eta_z, \zeta_z)$ we obtain an estimator with strictly lower variance than $\hat{f}$ for all $\boldsymbol{z} \in Z \cap S^*(\rho, \boldsymbol{v})$ and same variance as $\hat{f}$ on all other vectors. Note we can perform the shift consistently across all branches of the tree-like partial order on outcomes.

Violation (26) means that for some $\epsilon > 0$, $\hat{f}$ has average value strictly lower than L on $S(u, \boldsymbol{v})$ for all intervals $u \in (\eta, \rho]$ for $\eta \in [\rho - \epsilon, \rho)$. For all $\boldsymbol{z}$, the $\boldsymbol{z}$-optimal extension $\hat{f}^{(\boldsymbol{z}, \rho, M)}(u)$ has value $\lambda(\rho, \boldsymbol{z}, M) \geq L$ at ρ and (from convexity of lower hull) values that are at least that on $u < \rho$. From unbiasedness, we must have for all $\boldsymbol{z} \in Z \cap S^*(\rho, \boldsymbol{v})$, $\int_0^\rho \hat{f}(u, \boldsymbol{z})du = \int_0^\rho \hat{f}^{(\boldsymbol{z}, \rho, M)}(u)du$. Therefore, values lower than L must be compensated for in $\hat{f}$ by values higher than L. We can modify the estimator such that it is equal to L for $S(u, \boldsymbol{v})$ for $u \in (\rho - \epsilon, \rho)$ and compensate for that by lowering values at lower u values $u < \zeta_z$ that are higher than L. The modified estimator has strictly lower variance than $\hat{f}$ for all $\boldsymbol{z} \in Z \cap S^*(\rho, \boldsymbol{v})$ and same variance as $\hat{f}$ on all other vectors. □

4. THE L* ESTIMATOR

The L* estimator, $\hat{f}^{(L)}$, is the solution of (20a) with equalities, obtaining values that are minimum in the optimal range. Formally, it is the solution of the integral equation $\forall \boldsymbol{v} \in \mathbf{V}, \forall \rho \in (0, 1]$:

$$\hat{f}^{(L)}(\rho, \boldsymbol{v}) \quad = \quad \frac{\underline{f}^{(\boldsymbol{v})}(\rho) - \int_\rho^1 \hat{f}^{(L)}(u, \boldsymbol{v})du}{\rho} \tag{29}$$

Geometrically, as visualized in Figure 2, the L* estimate on an outcome $S(\rho, \boldsymbol{v})$ is exactly the slope value that if maintained for outcomes $S(u, \boldsymbol{v})$ $(u \in (0, \rho])$, would yield an expected estimate of $\underline{f}(S)$. We derive a convenient expression for the L* estimator, which enables us to derive explicit forms or compute it for any function f. We show that the L* estimator is 4-competitive and that it is the unique admissible monotone estimator. We also show it is order-optimal with respect to the natural order that prioritizes data vectors with lower $f(\boldsymbol{v})$.

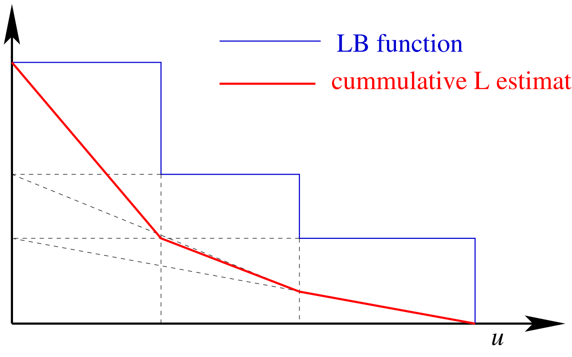


Figure 2: An example lower bound function $\underline{f}^{(\boldsymbol{v})}(u)$ with 3 steps and the respective cummulative L estimate $\int_u^1 \hat{f}^{(L)}(u, \boldsymbol{v})du$. The estimate $\hat{f}^{(L)}$ is the negated slope and in this case is also a step function with 3 steps.

Fixing $\boldsymbol{v}$, (29) is a first-order differential equation for $F(\rho) \equiv \int_\rho^1 \hat{f}^{(L)}(u, \boldsymbol{v})du$ and the initial condition $F(1) = 0$. Since the lower bound function $\underline{f}^{(\boldsymbol{v})}$ is monotonic and bounded, it is continuous (and differentiable) almost everywhere. Therefore, the equation with the initial condition has a unique solution:

LEMMA 4.1.

$$\hat{f}^{(L)}(\rho, \boldsymbol{v}) = \frac{\underline{f}^{(\boldsymbol{v})}(\rho)}{\rho} - \int_\rho^1 \frac{\underline{f}^{(\boldsymbol{v})}(u)}{u^2}du \tag{30}$$

$$\tag{31}$$

When $\underline{f}^{(\boldsymbol{v})}(1) = 0$, which we can assume without loss of generality[3]*, the solution has the simpler form:*

$$\hat{f}^{(L)}(\rho, \boldsymbol{v}) = -\int_\rho^1 \frac{\frac{d\underline{f}^{(\boldsymbol{v})}(u)}{du}}{u}du \tag{32}$$

We show a tight bound of 4 for the competitive ratio for $\hat{f}^{(L)}$, meaning that it is at most 4 for all functions f and for any $\epsilon > 0$, there exists a function f on which the ratio is no less than $4 - \epsilon$.

THEOREM 4.1.

$$\sup_{f, \boldsymbol{v} \mid \int_0^1 \hat{f}^{(\boldsymbol{v})}(u)^2 du < \infty} \frac{\int_0^1 \hat{f}^{(L)}(u, \boldsymbol{v})^2 du}{\int_0^1 \hat{f}^{(\boldsymbol{v})}(u)^2 du} = 4\,,$$

We present a family of functions for which the supermum of this ratio is 4. We use the domain $\mathbf{V} = [0, 1]$, a PPS sampling scheme with $\tau(u) = u$, and the function $f(v) = \frac{1}{1-p} - \frac{v^{1-p}}{1-p}$ for $p \in [0, 0.5)$. For the data $v = 0$ we have the following convex lower bound function

$$\underline{f}(u, 0) = \frac{1}{1-p} - \frac{u^{1-p}}{1-p}\,.$$

Being convex, this lower bound function is equal to its lower hull. Therefore, by taking its negated derivative, we get $\hat{f}^{(0)}(u) = 1/u^p$. The function $\hat{f}^{(0)}$ is square integrable when $p < 0.5$:

$$\int_0^1 \hat{f}^{(0)}(u)^2 du = \int_0^1 1/u^{2p} du = \frac{1}{1 - 2p}\,.$$

From (32), the L* estimator on outcomes consistent with $v = 0$ for $p \in (0, 0.5)$ is[4]

$$\hat{f}^{(L)}(x, 0) = \int_x^1 \frac{1}{u^{1+p}} = \frac{1}{p}\left(\frac{1}{x^p} - 1\right)\,.$$

Hence,

$$\int_0^1 \hat{f}^{(L)}(u, 0)^2 du = \frac{1}{p^2}\int_0^1 \left(\frac{1}{u^{2p}} - \frac{2}{u^p} + 1\right)du$$
$$= \frac{1}{p^2}\left(\frac{1}{1-2p} - \frac{2}{1-p} + 1\right) = \frac{2}{(1-2p)(1-p)}\,.$$

We obtain the ratio

$$\frac{\int_0^1 \hat{f}^{(L)}(u, 0)^2 du}{\int_0^1 \hat{f}^{(0)}(u)^2 du} = \frac{2}{1-p} \leq 4\,.$$

The ratio approaches 4 when $p \to 0.5^-$.

[3] Otherwise, we can instead estimate the function $f(\boldsymbol{v}) - \underline{f}^{(\boldsymbol{v})}(1)$, which satisfies this assumption, and then add a fixed value of $\underline{f}^{(\boldsymbol{v})}(1)$ to the resulting estimate.

[4] For $p = 0$ the estimate is $-\ln(x)$.

We conclude the proof of Theorem 4.1 using the following lemma that shows that if $\hat{f}^{(\boldsymbol{v})}(u)$ is square integrable, that is, (9) holds, then $\hat{f}^{(L)}(u,\boldsymbol{v})$ is also square integrable and the ratio between these integrals is at most 4.

LEMMA 4.2.

$$\forall \boldsymbol{v},\ \int_0^1 \hat{f}^{(\boldsymbol{v})}(u)^2 du < \infty \implies \frac{\int_0^1 \hat{f}^{(L)}(u,\boldsymbol{v})^2 du}{\int_0^1 \hat{f}^{(\boldsymbol{v})}(u)^2 du} \leq 4\ .$$

PROOF. Fixing $\boldsymbol{v}$, the function $\hat{f}^{(\boldsymbol{v})}$ only depends on the lower hull of the lower bound function $\underline{f}^{(\boldsymbol{v})}(u)$. The estimator $\hat{f}^{(L)}$ depends on the lower bound function $\underline{f}$ and can be different for different lower bound functions with the same lower hull. Fixing the lower hull, the variance of the L* estimator is maximized for f such that $\underline{f}^{(\boldsymbol{v})} \equiv H_f^{(\boldsymbol{v})}$. It therefore suffices to consider convex $\underline{f}^{(\boldsymbol{v})}(u)$, that is, $\frac{d^2\underline{f}^{(\boldsymbol{v})}(u)}{d^2u} > 0$ for which we have

$$\hat{f}^{(\boldsymbol{v})}(u) = -\frac{d\underline{f}^{(\boldsymbol{v})}(u)}{du}\ .$$

Recall that $\hat{f}^{(\boldsymbol{v})}(u)$ is monotone non-increasing. From (32), we have $\hat{f}^{(L)}(\rho,\boldsymbol{v}) = -\int_\rho^1 \frac{\frac{d\underline{f}^{(\boldsymbol{v})}(u)}{du}}{u}du$.

To establish our claim, it suffices to show that for all monotone, non increasing, square integrable functions $g : (0,1]$,

$$\frac{\int_0^1 (\int_x^1 \frac{g(u)}{u}du)^2 dx}{\int_0^1 g(x)^2 dx} \leq 4 \qquad (33)$$

Define $h(x) = \int_x^1 \frac{g(u)}{u}du$.

$$\int_0^1 h^2(x)dx = -\int_0^1 \int_x^1 2h(y)h'(y)dydx$$
$$= -\int_0^1 \int_0^y 2h(y)h'(y)dxdy = -2\int_0^1 h(y)h'(y)\int_0^y dxdy$$
$$= -2\int_0^1 h(y)h'(y)ydy = 2\int_0^1 h(y)\frac{g(y)}{y}ydy$$
$$= 2\int_0^1 h(y)g(y)dy \leq 2\sqrt{\int_0^1 h^2(y)dy}\sqrt{\int_0^1 g^2(y)dy}$$

The first equality uses $h(1) = 0$. The third uses $h'(x) = -g(x)/x$. The inequality uses Cauchy-Schwartz. Finally, to obtain (33), we divide both sides by $\sqrt{\int_0^1 h^2(y)dy}$.

□

THEOREM 4.2. *The estimator $\hat{f}^{(L)}$ is monotone. Moreover, it is the unique admissible monotone estimator and dominates all monotone estimators.*

PROOF. Recall that an estimator $\hat{f}$ is monotone if and only if, for any data $\boldsymbol{v}$, the estimate $\hat{f}(\rho,\boldsymbol{v})$ is non-increasing with ρ. To show monotonicity of the L* estimator, we rewrite (30) to obtain

$$\hat{f}^{(L)}(\rho,\boldsymbol{v}) = \underline{f}^{(\boldsymbol{v})}(\rho) + \int_\rho^1 \frac{\underline{f}^{(\boldsymbol{v})}(\rho) - \underline{f}^{(\boldsymbol{v})}(x)}{x^2}dx\ , \qquad (34)$$

which is clearly non-increasing with ρ.

We now show that $\hat{f}^{(L)}$ dominates all monotone estimators (and hence is the unique admissible monotone estimator). By definition, a monotone estimator $\hat{f}$ can not exceed λ_L on any outcome, that is, it must satisfy the inequalities $\forall \boldsymbol{v},\ \forall\rho \in [0,1]$:

$$\rho\hat{f}(\rho,\boldsymbol{v}) + \int_\rho^1 \hat{f}(u,\boldsymbol{v})du \leq \inf_{\boldsymbol{z}\in S^*(\rho,\boldsymbol{v})} \int_0^1 \hat{f}(u,\boldsymbol{z})du =$$
$$\inf_{\boldsymbol{z}\in S^*(\rho,\boldsymbol{v})} f(\boldsymbol{z}) = \underline{f}^{(\boldsymbol{v})}(\rho)\ . \qquad (35)$$

Estimator $\hat{f}^{(L)}$ satisfies (35) with equalities. If there is a monotone estimator $\hat{f}$ which is not equivalent to $\hat{f}^{(L)}$, that is, for some $\boldsymbol{v}$, the integral is strictly smaller than the integral of $\hat{f}^{(L)}$ on some interval $(\rho-\epsilon,\rho)$ ($\epsilon > 0$ may depend on $\boldsymbol{v}$), we can obtain a monotone estimator that strictly dominates $\hat{f}$ by decreasing the estimate for $u \leq \rho-\epsilon$ and increasing it for $u > \rho-\epsilon$. The variance decreases because we decrease the estimate on higher values and increase on lower values. □

Lastly, we show that $\hat{f}^{(L)}$ is order-optimal with respect to the order $\prec$ which prioritizes vectors with lower $f(\boldsymbol{v})$:

THEOREM 4.3. *A $\prec^+$-optimal estimator for f with respect to the partial order*

$$\boldsymbol{v} \prec \boldsymbol{v}' \iff f(\boldsymbol{v}) < f(\boldsymbol{v}')$$

must be equivalent to $\hat{f}^{(L)}$.

PROOF. We use our results of order-optimality (Section 5). We can check that we obtain (29) using (42) and $\prec$ as defined in the statement of the Theorem. Thus, a $\prec^+$-optimal solution must have this form. □

Example 4 contains an example derivation of the L* estimator. Note that it may not be bounded. Another estimator that is both bounded and competitive (but not necessarily in-range, not monotone, and has a large competive ratio) is the J estimator [15].

5. ORDER-OPTIMALITY

We identify conditions on f and $\prec$ under which a $\prec^+$-optimal estimator exists and specify this estimator as a solution of a set of equations. Our derivations of $\prec^+$-optimal estimators follow the intuition to require the estimate on an outcome S to be $\boldsymbol{v}$-optimal with respect to the $\prec$-minimal vector that is consistent with the outcome:

$$\forall S = S(\rho,\boldsymbol{v}),\ \hat{f}(S) = \lambda(\rho, \min_\prec(S^*))\ . \qquad (36)$$

When $\prec$ is a total order and V is finite, $\min_\prec(S^*)$ is unique and (36) is well defined. Moreover, as long as f has a nonnegative unbiased estimator, a solution (36) always exists and is $\prec^+$-optimal. We preview a simple construction of the solution: Process vectors in increasing $\prec$ order, iteratively building a partially defined nonnegative estimator. When processing $\boldsymbol{v}$, the estimator is already defined for $S(u,\boldsymbol{v})$ for $u \geq \rho_v$, for some $\rho_v \in (0,1]$. We extend it to the outcomes $S(u,\boldsymbol{v})$ for $u \leq \rho_v$ using the $\boldsymbol{v}$-optimal extension $\hat{f}^{(\boldsymbol{v},\rho_v,M)}(u)$, where $M = \int_{\rho_v}^1 \hat{f}(u,\boldsymbol{v})du$ (see Theorem 2.1).

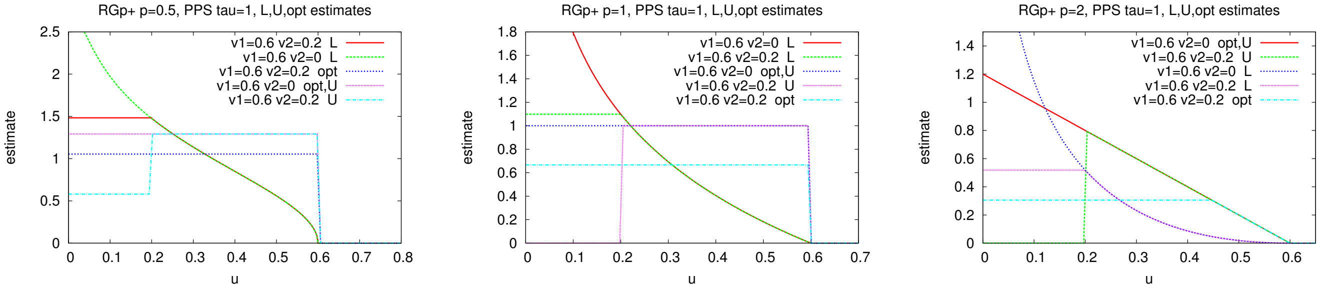
We now formulate conditions that will allow us to establish $\prec^+$-optimality of a solution of (36) in more general settings. These conditions always hold when $\prec$ is a total order and V is finite. Generally,

$$\min_\prec(S^*) = \{\boldsymbol{z} \in S^* | \neg\exists \boldsymbol{w} \in S^*,\ \boldsymbol{w} \prec \boldsymbol{z}\}$$

is a *set* and (36) is well defined when $\forall S$, this set is not empty and $\lambda(\rho, \min_\prec(S^*))$ is unique, that is, the value $\lambda(\rho,\boldsymbol{z})$ is the same for

Example 4 L* and U* estimates for Example 3

We compute the L* and U* estimators for $\textsc{RG}_{p+}$ for the sampling scheme and data in Example 3. For the two vectors $(0.6, 0.2)$ and $(0.6, 0)$, both the L* and U* estimates are 0 when $u \geq 0.6$, this is necessary from unbiasedness and nonnegativity because for these outcomes $\exists \boldsymbol{v} \in S^*, \textsc{RG}_{p+}(\boldsymbol{v}) = 0$. Otherwise, the L* estimate is $\hat{\textsc{RG}}^{(L)}_{p+}(S) = (v_1 - {v'}_2)^p / {v'}_2 - \int_{{v'}_2}^{v_1} \frac{(v_1-x)^p}{x^2} dx$, where ${v'}_2 = u$ when $S = \{1\}$ and ${v'}_2 = v_2$ when $S = \{1, 2\}$. When $p \geq 1$, the U* estimate is $\hat{\textsc{RG}}^{(U)}_{p+}(S) = p(v_1 - u)^{p-1}$ when $u \in (v_2, v_1]$ and 0 when $u \leq v_2 < v_1$. When $p \leq 1$ the U* estimate is v_1^{p-1} when $u \in (v_2, v_1]$ and $\frac{(v_1-v_2)^p - v_1^{p-1}(v_1-v_2)}{v_2}$ when $u \leq v_2 < v_1$.
The figure also include the $\boldsymbol{v}$-optimal estimates, discussed in Example 3. When $v_2 = 0$, the U* estimates are $\boldsymbol{v}$-optimal. The L* estimate is not bounded when $v_2 = 0$ (but has bounded variance and is competitive).



all $\prec$-minimal vectors $\boldsymbol{z} \in \min_{\prec}(S^*)$. A sufficient condition for this is that

$$\forall \rho\, \forall \boldsymbol{v}\, \forall x \in (0, \underline{f}(\rho, \boldsymbol{v})]\ \forall \boldsymbol{z}, \boldsymbol{w} \in \min_{\prec}(S^*(\rho, \boldsymbol{v})),$$
$$\inf_{\eta < \rho} \frac{\underline{f}(\eta, \boldsymbol{z}) - x}{\rho - \eta} = \inf_{\eta < \rho} \frac{\underline{f}(\eta, \boldsymbol{w}) - x}{\rho - \eta} \tag{37}$$

In this case, the respective Equation (36) on $u \in (0, \rho]$ are the same for all $\boldsymbol{z} \in \min_{\prec}(S^*)$ and thus so are the estimate values $\hat{f}(u, \boldsymbol{z})$.

We say that $Z \subset V$ is *$\prec$-bounded* if

$$\forall \boldsymbol{v} \in Z\ \exists \boldsymbol{z} \in \min_{\prec}(Z),\ \boldsymbol{z} \preceq \boldsymbol{v} \tag{38}$$

That is, for all $\boldsymbol{z} \in Z$, $\boldsymbol{z}$ is $\prec$-minimal or is preceded by some vector that is $\prec$-minimal in Z.

We say that an outcome S is $\prec$-bounded if S^* is $\prec$-bounded, that is,

$$\forall \boldsymbol{v} \in S^*\ \exists \boldsymbol{z} \in \min_{\prec}(S^*),\ \boldsymbol{z} \preceq \boldsymbol{v} \tag{39}$$

When all outcomes $S(u, \boldsymbol{v})$ are $\prec$-bounded, we say that a set of vectors R *represents* $\boldsymbol{v}$ if any outcome consistent with $\boldsymbol{v}$ has a $\prec$-minimal vector in R:

$$\forall u \in (0, 1], \exists \boldsymbol{z} \in R,\ \boldsymbol{z} \in \min_{\prec}(S^*(u, \boldsymbol{v}))\ .$$

We now show that we can obtain a $\prec^+$-optimal estimator if every vector $\boldsymbol{v}$ has a set of finite size that represents it. Example 5 walks through a derivation of $\prec^+$-optimal estimators.

LEMMA 5.1. *If f satisfies* (8), (37), (39) *and*

$$\forall \boldsymbol{v},\ \min\{|R|\ |\ \forall u \in (0, 1], \exists \boldsymbol{z} \in R,\ \boldsymbol{z} \in \min_{\prec} S^*(u, \boldsymbol{v})\} < \infty\ ,$$

then a $\prec^+$-optimal estimator exists and must be equivalent to a solution of (36).

PROOF. We provide an explicit construction of a $\prec^+$-optimal estimator for f.

Fixing $\boldsymbol{v}$, we select a finite set of representatives. We can map the representatives (or a subset of them) to distinct subintervals covering $(0, 1]$. The subintervals have the form $(a_i, a_{i-1}]$ where $0 = a_n < \cdots a_1 < a_0 = 1$ such that a representative $\boldsymbol{z}$ that is minimal for $(a_i, a_{i-1}]$ is not minimal for $u \leq a_i$. Such a mapping can always be obtained since from (2), each vector is consistent with an open interval of the form $(a, 1]$, and thus if $\prec$-minimum at $S^*(u, \boldsymbol{v})$ (we must have $u > a$) it must be $\prec$-minimum for $S^*(x, \boldsymbol{v})$ for $x \in (a, u]$. Thus, the region on which $\boldsymbol{z}$ is in $\min_{\prec} S^*(u, \boldsymbol{v})$ is open to the left. We can always choose a mapping such that the left boundary of this region corresponds to a_i.

Let $\boldsymbol{z}^{(i)}$ ($i \in [n]$) be the representative mapped to outcomes $S(u, \boldsymbol{v})$ where $u \in (a_i, a_{i-1}]$. Since $S^*(u, \boldsymbol{v})$ is monotone non-decreasing with u, $i < j$ implies that $\boldsymbol{z}^{(i)} \prec \boldsymbol{z}^{(j)}$ or that they are incomparable in the partial order.

We construct a partially specified nonnegative estimator in steps, by solving (36) iteratively for the vectors $\boldsymbol{z}^{(i)}$. Initially we invoke Theorem 2.1 to obtain estimate values for $S(u, \boldsymbol{z}^{(1)})$ $u \in (0, 1]$ that minimize the variance for $\boldsymbol{z}^{(1)}$. The result is a partially specified nonnegative estimator. In particular for $\boldsymbol{v}$, the estimator is now specified for outcomes $S(u, \boldsymbol{v})$ where $u \in (a_1, 1]$. Any modification of this estimator on a subinterval of $(a_1, 1]$ with positive measure will strictly increase the variance for $\boldsymbol{z}^{(1)}$ (or result in an estimator that can not be completed to a nonnegative unbiased one).

After step i, we have a partially specified nonnegative estimator that is specified for $S(u, \boldsymbol{v})$ for $u \in (a_i, 1]$. The estimator is fully specified for $\boldsymbol{z}^{(j)}$ $j \leq i$ and is $\prec^+$-optimal on these vectors in the sense that any other partially specified nonnegative estimator that is fully specified for $\boldsymbol{z}^{(j)}$ $j \leq i$ and has strictly lower variance on some $\boldsymbol{z}^{(j)}$ ($j \leq i$) must have strictly higher variance on some $\boldsymbol{z}^{(h)}$ such that $h < j$.

We now invoke Theorem 2.1 with respect to the vector $\boldsymbol{z}^{(i+1)}$. The estimator is partially specified for $S(u, \boldsymbol{z}^{(i+1)})$ on $u > a_i$ and we obtain estimate values for the outcomes $S(u, \boldsymbol{z}^{(i+1)})$ for $u \in (0, a_i]$ that constitute a partially specified nonnegative estimator with minimum variance for $\boldsymbol{z}^{(i+1)}$. Note again that this completion is unique (up to equivalence). This extension now defines $S(u, \boldsymbol{v})$ for $u \in (a_{i+1}, 1]$.

Lastly, we must have $f(\boldsymbol{z}^{(n)}) = f(\boldsymbol{v})$ because $f(\boldsymbol{z}^{(n)}) < f(\boldsymbol{v})$ implies that (8) is violated for $\boldsymbol{v}$ whereas the reverse inequality implies that (8) is violated for $\boldsymbol{z}^{(n)}$. Since at step n the estimator is specified for all outcomes $S(u, \boldsymbol{z}^{(n)})$ and unbiased, it is unbiased for $\boldsymbol{v}$.

The estimator is invariant to the choice of the representative sets $R_{\boldsymbol{v}}$ for $\boldsymbol{v} \in V$ and also remains the same if we restrict $\prec$ so that it includes only relations between $\boldsymbol{v}$ and $R_{\boldsymbol{v}}$.

We so far showed that there is a unique, up to equivalence, partially specified nonnegative estimator that is $\prec^+$ optimal with respect to a vector $\boldsymbol{v}$ and all vectors it depends on. Consider now all outcomes $S(u, \boldsymbol{v})$, for all u and $\boldsymbol{v}$, arranged according to the containment order on $S^*(u, \boldsymbol{v})$ according to decreasing u values with branching points when $S^*(u, \boldsymbol{v})$ changes. If for two vectors $\boldsymbol{v}$ and $\boldsymbol{z}$, the sets of outcomes $S(u, \boldsymbol{v}), u \in (0, 1]$ and $S(u, \boldsymbol{z}), u \in (0, 1]$ intersect, the intersection must be equal for $u > \rho$ for some $\rho < 1$. In this case the estimator values computed with respect to either $\boldsymbol{z}$ or $\boldsymbol{v}$ would be identical for $u \in (\rho, 1]$. Also note that partially specified nonnegative solutions on different branches are independent. Therefore, solutions with respect to different vectors $\boldsymbol{v}$ can be consistently combined to a fully specified estimator. □

5.1 Continuous domains

The assumptions of Lemma 5.1 may break on continuous domains. Firstly, outcomes may not be $\prec$-bounded and in particular, $\min_{\prec}(S^*)$ can be empty even when S^* is not, resulting in (36) not being well defined. Secondly, even if $\prec$ is a total order, minimum elements do not necessarily exist and thus (39) may not hold, and lastly, there may not be a finite set of representatives. To treat such domains, we utilize a notion of *convergence with respect to* $\prec$:

We define the $\prec$-lim of a function h on a set of vectors $Z \subset V$:

$$\begin{aligned} &\prec\text{-}\lim(h(\cdot), Z) = x \iff \\ &\forall \boldsymbol{v} \in Z\, \forall \epsilon > 0\, \exists \boldsymbol{w} \preceq \boldsymbol{v}, \forall \boldsymbol{z} \preceq \boldsymbol{w},\ |h(\boldsymbol{z}) - x| \le \epsilon \end{aligned} \tag{40}$$

The $\prec$-lim may not exist but is unique if it does. Note that when Z is finite or more generally, $\prec$-bounded, and $h(\boldsymbol{z})$ is unique for all $\boldsymbol{z} \in \min_{\prec} Z)$, then $\prec\text{-}\lim(h(\cdot), Z) = h(\min_{\prec} Z)$.

We define the $\prec$-closure of $\boldsymbol{z}$ as the set containing $\boldsymbol{z}$ and all preceding vectors $\mathrm{cl}_{\prec}(\boldsymbol{z}) = \{\boldsymbol{v} \in V | \boldsymbol{v} \preceq \boldsymbol{z}\}$.

We provide an alternative definition of the $\prec$-lim using the notion of $\prec$-closure.

$$\begin{aligned} &\prec\text{-}\lim(h(\cdot), Z) = x \\ \iff\quad &\inf_{\boldsymbol{v} \in Z} \sup_{\boldsymbol{z} \in \mathrm{cl}_{\prec}(\boldsymbol{v}) \cap Z} h(\boldsymbol{z}) = \sup_{\boldsymbol{v} \in Z} \inf_{\boldsymbol{z} \in \mathrm{cl}_{\prec}(\boldsymbol{v}) \cap Z} h(\boldsymbol{z}) = x \end{aligned} \tag{41}$$

We say that the lower bound function $\prec$*-converges* on outcome $S = S(\rho, \boldsymbol{v})$ if $\prec\text{-}\lim(\underline{f}(\eta, \cdot), S^*)$ exists for all $\eta \in (0, \rho)$. When this holds, the $\prec$ - lim of the optimal values (16) over consistent vectors S^* exists for all $M = \int_\rho^1 \hat{f}(u, \boldsymbol{v}) du \le \underline{f}(\rho, \boldsymbol{v})$. We use the notation

$$\begin{aligned} \lambda_{\prec}(S, M) &= \prec\text{-}\lim(\lambda(\rho, \cdot, M), S^*) \\ &= \inf_{0 \le \eta < \rho} \frac{\prec\text{-}\lim(\underline{f}(\eta, \cdot), S^*) - M}{\rho - \eta} . \end{aligned}$$

When the partially specified estimator $\hat{f}$ is clear from context, we omit the parameter M and use the notation

$$\begin{aligned} \lambda_{\prec}(S) &= \prec\text{-}\lim(\lambda(\rho, \cdot), S^*) \\ &= \inf_{0 \le \eta < \rho} \frac{\prec\text{-}\lim(\underline{f}(\eta, \cdot), S^*) - \int_\rho^1 \hat{f}(u, \boldsymbol{v}) du}{\rho - \eta} . \end{aligned}$$

We can finally propose a generalization of (36):

$$\forall S,\ \hat{f}(S) = \lambda_{\prec}(S) \tag{42}$$

which is well defined when the lower bound function $\prec$-converges for all S:

$$\forall S = S(\rho, \boldsymbol{v}), \forall \eta \le \rho,\ \prec\text{-}\lim(\underline{f}(\eta, \cdot), S^*) \text{ exists.} \tag{43}$$

Using the definition (41) of $\prec$-convergence and (3) we obtain that an estimator is equivalent to (42) if and only if

$$\forall \boldsymbol{v} \forall \rho \in (0, 1],\ \lim_{\eta \to \rho^-} \frac{\int_\eta^\rho \hat{f}(u, \boldsymbol{v}) du}{\rho - \eta} = \lambda_{\prec}(\rho, \boldsymbol{v}) \tag{44}$$

We show that equivalence to (42) is *necessary* for $\prec^+$-optimality. To facilitate the proof, we express $\prec^+$-optimality in terms of restricted admissiblity:

LEMMA 5.2. *An estimator is $\prec^+$-optimal if and only if, for all $\boldsymbol{v} \in V$, it is admissible with respect to $\mathrm{cl}_{\prec}(\boldsymbol{v})$.*

PROOF. If there is $\boldsymbol{v}$ such that $\hat{f}$ is not admissible on $\mathrm{cl}_{\prec}(\boldsymbol{v})$, there is an alternative estimator with strictly lower variance on some $\boldsymbol{z} \in \mathrm{cl}_{\prec}(\boldsymbol{v})$ and at most the variance on all $\mathrm{cl}_{\prec}(\boldsymbol{v}) \setminus \{\boldsymbol{z}\}$. Since $\mathrm{cl}_{\prec}(\boldsymbol{v})$ contains all vectors that precede $\boldsymbol{z}$, the estimator $\hat{f}$ can not be $\prec^+$-optimal. To establish the converse, assume an estimator $\hat{f}$ is admissible on $\mathrm{cl}_{\prec}(\boldsymbol{v})$ for all $\boldsymbol{v}$. Consider $\boldsymbol{z} \in V$. Since $\hat{f}$ is admissible on $\mathrm{cl}_{\prec}(\boldsymbol{z})$, there is no alternative estimator with strictly lower variance on $\boldsymbol{z}$ and at most the variance of $\hat{f}$ on all preceding vectors. Since this holds for all $\boldsymbol{z}$, we obtain that $\hat{f}$ is $\prec^+$-optimal. □

LEMMA 5.3. *If f satisfies* (8) *and* (43) *then $\hat{f}$ is $\prec^+$-optimal only if it satisfies* (44).

PROOF. Lemma 5.2 states that an estimator is $\prec^+$-optimal if and only if $\forall \boldsymbol{w} \in V$ it is admissible with respect to $\mathrm{cl}_{\prec}(\boldsymbol{w})$. Applying Lemma 3.1, the latter holds only if

$$\begin{aligned} &\forall \boldsymbol{v} \in V\, \forall \rho \in (0, 1] \\ \lim_{\eta \to \rho^-} \frac{\int_\eta^\rho \hat{f}(u, \boldsymbol{v}) du}{\rho - \eta} &\ge \inf_{\boldsymbol{z} \in \mathrm{cl}_{\prec}(\boldsymbol{v}) \cap S^*(\rho, \boldsymbol{v})} \lambda(\rho, \boldsymbol{z}) \\ &\le \sup_{\boldsymbol{z} \in \mathrm{cl}_{\prec}(\boldsymbol{v}) \cap S^*(\rho, \boldsymbol{v})} \lambda(\rho, \boldsymbol{z}) \end{aligned} \tag{45}$$

From definition, $S(\rho, \boldsymbol{z}) \equiv S(\rho, \boldsymbol{v})$ for all vectors $\boldsymbol{z} \in S^*(\rho, \boldsymbol{v})$. Moreover, for $\boldsymbol{z} \in S^*(\rho, \boldsymbol{v})$ there is a nonempty interval $(\eta_z, \rho]$ such that $\forall u \in (\eta_z, \rho]$, $S^*(u, \boldsymbol{z} \equiv S^*(u, \boldsymbol{v})$. Therefore, for all $\boldsymbol{z} \in S^*(\rho, \boldsymbol{v})$, the limits $\lim_{\eta \to \rho^-} \frac{\int_\eta^\rho \hat{f}(u, \boldsymbol{z}) du}{\rho - \eta}$ are the same. Therefore, (45) $\iff$

$$\begin{aligned} &\forall \boldsymbol{v} \in V\, \forall \rho \in (0, 1] \\ \lim_{\eta \to \rho^-} \frac{\int_\eta^\rho \hat{f}(u, \boldsymbol{v}) du}{\rho - \eta} &\ge \sup_{\boldsymbol{w} \in S^*(\rho, \boldsymbol{v})} \inf_{\boldsymbol{z} \in \mathrm{cl}_{\prec}(\boldsymbol{w}) \cap S^*(\rho, \boldsymbol{v})} \lambda(\rho, \boldsymbol{z}) \\ &\le \inf_{\boldsymbol{w} \in S^*(\rho, \boldsymbol{v})} \sup_{\boldsymbol{z} \in \mathrm{cl}_{\prec}(\boldsymbol{w}) \cap S^*(\rho, \boldsymbol{v})} \lambda(\rho, \boldsymbol{z}) \end{aligned} \tag{46}$$

□

We leave open the question of determining the most inclusive conditions on f and $\prec$ under which a $\prec^+$-optimum exists, and thus the solution of (42) is $\prec^+$-optimal. We show that any solution of (42) is unbiased and nonnegative when f has a nonnegative unbiased estimator.

LEMMA 5.4. *When f and $\prec$ satisfy* (8) *and* (43)*, a solution $\hat{f}^{(\prec+)}$ of* (42) *is unbiased and nonnegative.*

PROOF. From Lemma 3.1, since all values are in-range, the solution is unbiased and nonnegative. □

6. THE U* ESTIMATOR

The estimator $\hat{f}^{(U)}$ satisfies (20b) with equality.

$$\forall S(\rho,\boldsymbol{v}),\ \hat{f}(\rho,\boldsymbol{v}) = \sup_{\boldsymbol{z}\in S^*} \inf_{0\leq\eta<\rho} \frac{\underline{f}(\eta,\boldsymbol{z}) - \int_\rho^1 \hat{f}(u,\boldsymbol{v})du}{\rho-\eta} \qquad (47)$$

The U* estimator is not always admissible. We do show, however, that under a natural condition, it is order-optimal with respect to an order that prioritizes vectors with higher f values (and hence also admissible). The condition states that for all $S(\rho,\boldsymbol{v})$ and $\eta<\rho$, the supremum of the lower bound function $\underline{f}(\eta,\boldsymbol{z})$ over $\boldsymbol{z}\in S^*$ is attained (in the limiting sense) at vectors that maximize f on S^*. Formally:

$$\forall \eta<\rho,\ \lim_{x\to\overline{f}(S)} \sup_{\boldsymbol{z}\in S^*|f(\boldsymbol{z})\geq x} \underline{f}(\eta,\boldsymbol{z}) = \sup_{\boldsymbol{z}\in S^*} \underline{f}(\eta,\boldsymbol{z})\ , \qquad (48)$$

where $\overline{f}(S) = \sup_{\boldsymbol{z}\in S^*} f(\boldsymbol{z})$.

LEMMA 6.1. *If f satisfies* (48)*, then the U* estimator is $\prec^+$-optimal with respect to the order $\boldsymbol{z}\prec\boldsymbol{v} \iff f(\boldsymbol{z})>f(\boldsymbol{v})$.*

PROOF. We can show that when (48) holds then (47) is the same as (42). □

The condition (48) is satisfied by RG_p and RG_{p+}. In this case, the conditions of Lemma 5.1 are also satisfied and thus the U* estimator is $\prec^+$ optimal.

7. CONCLUSION

We define monotone sampling, and discuss its applications as a summarization tool of massive data. We propose general derivations of estimators with worst-case (competitiveness) or common-case (customization) guarantees.

Some interesting future research directions are (i) bounding the *universal ratio* for monotone sampling: the lowest ratio we can guarantee when an estimator with finite variances exists. Our L* estimator implies an upper bound of 4, and we have examples where the ratio is at least 1.4. (ii) Efficient constructions of estimators with *instance optimal* competitive ratio. (iii) Study monotone sampling with several independent seeds, which captures independent sampling [14].

Finally, we mention applications of our estimator constructions which include estimating L_p difference from sampled data [7] and sktech-based similarity estimation in social networks [9].

Acknowledgement

We thank Micha Sharir for his help in the proof of Lemma 4.2 and Haim Kaplan for many comments and helpful feedback.

8. REFERENCES

[1] K. S. Beyer, P. J. Haas, B. Reinwald, Y. Sismanis, and R. Gemulla. On synopses for distinct-value estimation under multiset operations. In *SIGMOD*, pages 199–210. ACM, 2007.

[2] K. R. W. Brewer, L. J. Early, and S. F. Joyce. Selecting several samples from a single population. *Australian Journal of Statistics*, 14(3):231–239, 1972.

[3] A. Z. Broder. On the resemblance and containment of documents. In *Proceedings of the Compression and Complexity of Sequences*, pages 21–29. IEEE, 1997.

[4] A. Z. Broder. Identifying and filtering near-duplicate documents. In *Proc.of the 11th Annual Symposium on Combinatorial Pattern Matching*, volume 1848 of *LNCS*, pages 1–10. Springer, 2000.

[5] J. W. Byers, J. Considine, M. Mitzenmacher, and S. Rost. Informed content delivery across adaptive overlay networks. *IEEE/ACM Trans. Netw.*, 12(5):767–780, October 2004.

[6] E. Cohen. Size-estimation framework with applications to transitive closure and reachability. *J. Comput. System Sci.*, 55:441–453, 1997.

[7] E. Cohen. Distance queries from sampled data: Accurate and efficient. Technical Report cs.DS/1203.4903, arXiv, 2012.

[8] E. Cohen. All-distances sketches, revisited: Hip estimators for massive graphs analysis. In *PODS*. ACM, 2014.

[9] E. Cohen, D. Delling, F. Fuchs, A. Goldberg, M. Goldszmidt, and R. Werneck. Scalable similarity estimation in social networks: Closeness, node labels, and random edge lengths. In *COSN*, 2013.

[10] E. Cohen and H. Kaplan. Spatially-decaying aggregation over a network: model and algorithms. *J. Comput. System Sci.*, 73:265–288, 2007. Full version of a SIGMOD 2004 paper.

[11] E. Cohen and H. Kaplan. Summarizing data using bottom-k sketches. In *ACM PODC*, 2007.

[12] E. Cohen and H. Kaplan. Tighter estimation using bottom-k sketches. In *Proceedings of the 34th VLDB Conference*, 2008.

[13] E. Cohen and H. Kaplan. Leveraging discarded samples for tighter estimation of multiple-set aggregates. In *ACM SIGMETRICS*, 2009.

[14] E. Cohen and H. Kaplan. Get the most out of your sample: Optimal unbiased estimators using partial information. In *Proc. of the 2011 ACM Symp. on Principles of Database Systems (PODS 2011)*. ACM, 2011. full version: `http://arxiv.org/abs/1203.4903`.

[15] E. Cohen and H. Kaplan. What you can do with coordinated samples. In *The 17th. International Workshop on Randomization and Computation (RANDOM)*, 2013. full version: `http://arxiv.org/abs/1206.5637`.

[16] E. Cohen, H. Kaplan, and S. Sen. Coordinated weighted sampling for estimating aggregates over multiple weight assignments. *Proceedings of the VLDB Endowment*, 2(1–2), 2009. full version: `http://arxiv.org/abs/0906.4560`.

[17] E. Cohen, Y.-M. Wang, and G. Suri. When piecewise determinism is almost true. In *Proc. Pacific Rim International Symposium on Fault-Tolerant Systems*, pages 66–71, December 1995.

[18] A. Das, M. Datar, A. Garg, and S. Rajaram. Google news personalization: scalable online collaborative filtering. In *WWW*, 2007.

[19] N. Duffield, M. Thorup, and C. Lund. Priority sampling for estimating arbitrary subset sums. *J. Assoc. Comput. Mach.*, 54(6), 2007.

[20] P. S. Efraimidis and P. G. Spirakis. Weighted random sampling with a reservoir. *Inf. Process. Lett.*, 97(5):181–185, 2006.

[21] P. Gibbons and S. Tirthapura. Estimating simple functions on the union of data streams. In *Proceedings of the 13th Annual ACM Symposium on Parallel Algorithms and Architectures*. ACM, 2001.

[22] P. B. Gibbons. Distinct sampling for highly-accurate answers to distinct values queries and event reports. In *International Conference on Very Large Databases (VLDB)*, pages 541–550, 2001.

[23] M. Hadjieleftheriou, X. Yu, N. Koudas, and D. Srivastava. Hashed samples: Selectivity estimators for set similarity selection queries. In *Proceedings of the 34th VLDB Conference*, 2008.

[24] J. Hájek. *Sampling from a finite population*. Marcel Dekker, New York, 1981.

[25] D. G. Horvitz and D. J. Thompson. A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, 47(260):663–685, 1952.

[26] D. E. Knuth. *The Art of Computer Programming, Vol 2, Seminumerical Algorithms*. Addison-Wesley, 1st edition, 1968.

[27] J. Lanke. On UMV-estimators in survey sampling. *Metrika*, 20(1):196–202, 1973.

[28] P. Li, , K. W. Church, and T. Hastie. One sketch for all: Theory and application of conditional random sampling. In *NIPS*, 2008.

[29] D. Mosk-Aoyama and D. Shah. Computing separable functions via gossip. In *ACM PODC*, 2006.

[30] P. Mukhopandhyay. *Theory and Methods of Survey Sampling*. PHI learning, New Delhi, 2 edition, 2008.

[31] E. Ohlsson. Sequential poisson sampling. *J. Official Statistics*, 14(2):149–162, 1998.

[32] E. Ohlsson. Coordination of pps samples over time. In *The 2nd International Conference on Establishment Surveys*, pages 255–264. American Statistical Association, 2000.

Example 5 Walk-through derivation of $\prec^+$-optimal estimators

We derive $\prec^+$-optimal RG_{1+} estimators over the finite domain $\mathbf{V} = \{0,1,2,3\}^2$. Assuming same sampling scheme on both entries, there are 3 threshold values of interest, where π_i $i \in [3]$ is such that entry of value i is sampled if and only if $u \le \pi_i$. We have $\pi_1 < \pi_2 < \pi_3$.
The lower bounds $\underline{\mathrm{RG}_{1+}}^{(\boldsymbol{v})}$ are step functions with steps at $u = \pi_i$. The table below shows $\underline{\mathrm{RG}_{1+}}^{(\boldsymbol{v})}(u)$ for all u and $\boldsymbol{v}$ such that $\mathrm{RG}_{1+}(\boldsymbol{v}) > 0$. When $\mathrm{RG}_{1+}(\boldsymbol{v}) = 0$, we have $\underline{\mathrm{RG}_{1+}}^{(\boldsymbol{v})}(u) \equiv 0$ and any unbiased nonnegative estimator must have 0 estimates on outcomes that are consistent with $\boldsymbol{v}$.

$\underline{\mathrm{RG}_{1+}}^{(\boldsymbol{v})}$	$(1,0)$	$(2,1)$	$(2,0)$	$(3,2)$	$(3,1)$	$(3,0)$
$(0,\pi_1]$	1	1	2	1	2	3
$(\pi_1,\pi_2]$	0	1	1	1	2	2
$(\pi_2,\pi_3]$	0	0	0	1	1	1
$(\pi_3,1]$	0	0	0	0	0	0

The $\boldsymbol{v}$-optimal estimate, $\hat{\mathrm{RG}}_{1+}^{(\boldsymbol{v})}(u)$ is the negated slope at u of the lower hull of $\underline{\mathrm{RG}_{1+}}^{(\boldsymbol{v})}$. The lower hull of each step function is piecewise linear with breakpoints at a subset of π_i, and thus, the $\boldsymbol{v}$-optimal estimates are constant on each segment $(\pi_{i-1},\pi_i]$. The table shows the estimates for all $\boldsymbol{v}$ and u. The notation $\downarrow$ refers to value in same column and one row below and $\Downarrow$ to value two rows below.

$\hat{\mathrm{RG}}_{1+}^{(\boldsymbol{v})}$	$(1,0)$	$(2,1)$	$(2,0)$	$(3,2)$	$(3,1)$	$(3,0)$
$(0,\pi_1]$	$\frac{1}{\pi_1}$	$\frac{1}{\pi_2}$	$\frac{2-(\pi_2-\pi_1)\downarrow}{\pi_1}$	$\frac{1}{\pi_3}$	$\frac{2-\Downarrow}{\pi_2}$	$\frac{3-\downarrow(\pi_3-\pi_2)-\Downarrow(\pi_2-\pi_1)}{\pi_1}$
$(\pi_1,\pi_2]$	0	$\frac{1}{\pi_2}$	$\min\{\frac{2}{\pi_2},\frac{1}{\pi_2-\pi_1}\}$	$\frac{1}{\pi_3}$	$\frac{2-\downarrow}{\pi_2}$	$\min\{\frac{3-\downarrow(\pi_3-\pi_2)}{\pi_2},\frac{2-\downarrow(\pi_3-\pi_2)}{\pi_2-\pi_1}$
$(\pi_2,\pi_3]$	0	0	0	$\frac{1}{\pi_3}$	$\min\{\frac{2}{\pi_3},\frac{1}{\pi_3-\pi_2}\}$	$\min\{\frac{3}{\pi_3},\frac{1}{\pi_3-\pi_2}\}$

The order $(2,1) \prec (2,0)$ and $(3,2) \prec (3,1) \prec (3,0)$ yields the L* estimator, which is $\boldsymbol{v}$-optimal for $(1,0)$, $(2,1)$, and $(3,2)$. The order $(2,0) \prec (2,1)$ and $(3,0) \prec (3,1) \prec (3,2)$ yields the U* estimator which is $\boldsymbol{v}$-optimal for $(1,0)$, $(2,0)$, and $(3,0)$. Observe that it suffices to only specify $\prec$ so that the order is defined between vectors consistent with the same outcome S when $\underline{\mathrm{RG}_{1+}}(S) > 0$. For RG_{1+}, this means specifying the order between vectors with the same v_1 value (and only consider those with strictly smaller v_2). In follows that any admissible estimator is $(1,0)$-optimal.
To specify an estimator, we need to specify it on all possible outcomes, where each distinct outcome is uniquely determined by a corresponding set of data vectors S^*. The 8 possible outcomes (we exclude those consistent with vectors with $\mathrm{RG}_{1+}(\boldsymbol{v}) = 0$ on which the estimate must be 0) are $(1,0)$, $(2,\le 1)$, $(2,1)$, $(3,\le 2)$, $(3,2)$, $(3,\le 1)$, $(3,1)$, and $(3,0)$, where an entry "$\le a$" specifies all vectors in $\mathbf{V}$ where the entry is at most a.
We show how we construct the $\prec^+$-optimal estimator for $\prec$ which prioritizes vectors with difference of 2: $(3,1) \prec (3,2) \prec (3,0)$ and $(2,0) \prec (2,1)$. The estimator is $\boldsymbol{v}$-optimal for $(3,1)$, $(2,0)$, and $(1,0)$. This determines the estimates $\hat{\mathrm{RG}}_{1+}^{(\prec)}$ on all outcomes consistent with these vectors: The value on outcome $(1,0)$ is $\hat{\mathrm{RG}}^{((1,0))}((0,\pi_1])$, the values on outcomes $(2,\le 1)$ and $(2,0)$ are according to $\hat{\mathrm{RG}}^{(2,0)}$ on $(\pi_1,\pi_2]$ and $(0,\pi_1]$, respectively, and value on the outcomes $(3,\le 2)$, $(3,\le 1)$ and $(3,1)$ is according to $\hat{\mathrm{RG}}^{(3,1)}$ on $(\pi_2,\pi_3]$ and $(\pi_1,\pi_2]$. These values are provided in the table above. The remaining outcomes are $(3,0)$, $(3,2)$, and $(2,1)$. We need to specify the estimator so that it is unbiased on these vectors, given the existing specification. We have

$$\begin{aligned}
\hat{\mathrm{RG}}_{1+}^{(\prec)}(2,1) &= \frac{1-(\pi_2-\pi_1)\hat{\mathrm{RG}}_{1+}^{(\prec)}(2,\le 1)}{\pi_1} \\
\hat{\mathrm{RG}}_{1+}^{(\prec)}(3,0) &= \frac{3-(\pi_3-\pi_2)\hat{\mathrm{RG}}_{1+}^{(\prec)}(3,\le 2)-(\pi_2-\pi_1)\hat{\mathrm{RG}}_{1+}^{(\prec)}(3,\le 1)}{\pi_1} \\
\hat{\mathrm{RG}}_{1+}^{(\prec)}(3,2) &= \frac{2-(\pi_3-\pi_2)\hat{\mathrm{RG}}_{1+}^{(\prec)}(3,\le 2)}{\pi_1} \ .
\end{aligned}$$

Observe that to apply these estimators, we do not have to precompute the estimator on all possible outcomes. An estimate only depends on values of the estimate on all less informative outcomes. In a discrete domain as in this example, it is the number of breakpoints larger than the seed u (which is at most the number of distinct values in the domain).

[33] B. Rosén. Asymptotic theory for successive sampling with varying probabilities without replacement, I. *The Annals of Mathematical Statistics*, 43(2):373–397, 1972.

[34] B. Rosén. Asymptotic theory for order sampling. *J. Statistical Planning and Inference*, 62(2):135–158, 1997.

[35] P. J. Saavedra. Fixed sample size pps approximations with a permanent random number. In *Proc. of the Section on Survey Research Methods, Alexandria VA*, pages 697–700. American Statistical Association, 1995.

[36] J.S. Vitter. Random sampling with a reservoir. *ACM Trans. Math. Softw.*, 11(1):37–57, 1985.