Bidder Selection Problem in Position Auctions via Poisson Approximation

Nick Gravin* Yixuan Even Xu† Renfei Zhou†

June 21, 2023

Abstract

We consider Bidder Selection Problem (BSP) in position auctions motivated by practical concerns of online advertising platforms. In this problem, the platform sells ad slots via an auction to a large pool of n potential buyers with independent values drawn from known prior distributions. The seller can only invite a fraction of $k < n$ advertisers to the auction due to communication and computation restrictions. She wishes to maximize either the social welfare or her revenue by selecting the set of invited bidders. Previous work has studied this problem under different names but only for the single-item auction formats. For the welfare objective [Chen, Hu, Li, Li, Liu, Lu 2016], [Mehta, Nadav, Psomas, Rubinstein 2020], [Segev and Singla 2021] gave different Polynomial Time Approximation Schemes, while for revenue objective [Mehta, Nadav, Psomas, Rubinstein 2020] showed that BSP is APX-hard for the second-price auction and [Bei, Gravin, Lu, Tang 2022] gave constant approximation algorithm for BSP of the second-price auction with anonymous reserve.

We study BSP in a classic multi-winner model of position auctions for welfare and revenue objectives using the optimal (respectively, VCG mechanism, or Myerson's auction) format for the selected set of bidders. We propose a novel Poisson-Chernoff relaxation of the problem with the following guarantees:

1. it is a continuous concave maximization problem that can be solved in time polynomial in n and k via standard concave maximization solvers;
2. its approximation guarantee (after standard rounding procedure) converges to the optimal solution at the rate $1 - O(k^{-1/4})$;
3. for the special case of single-item auction we get a better convergence rate of $1-O(\sqrt{\ln k/k})$.

These results show that BSP becomes polynomial time solvable up to a vanishingly small error as the problem size k grows. It also immediately implies a PTAS for more general environment of position auctions after combining our relaxation with the trivial brute force algorithm. The algorithm is in fact an Efficient PTAS (EPTAS) under a mild assumption $k \geq \log n$ with much better running time than previous EPTASes. Our approach yields simple and practically relevant algorithms unlike all previous complex PTAS algorithms, which had at least doubly exponential dependency of their running time on ε. In contrast, our algorithms are even faster than popular algorithms such as greedy for submodular maximization. Furthermore, we did extensive numerical experiments, which demonstrate high efficiency and practical applicability of our solution. Our experiments corroborate the experimental findings of [Mehta, Nadav, Psomas, Rubinstein 2020] that many simple heuristics perform surprisingly well, which indicates importance of using small ε for the BSP and practical irrelevance of all previous PTAS approaches.

*ITCS, Shanghai University of Finance and Economics. Email: `nikolai@mail.shufe.edu.cn`
†IIIS, Tsinghua University. Email: `{xuyx20,zhourf20}@mails.tsinghua.edu.cn`

1 Introduction

Online advertising is a big part of the modern e-commerce industry and a key to monetization of many online businesses. The majority of ad slots on a user web page are sold in *real time* via an automated auction to a group of candidate advertisers. The whole process from the time when an auction is initiated based on the impression about advertising opportunity up to the time when ads are displayed on the user's page usually has to be completed in a few milliseconds. This makes it imperative for the platform (Ad exchange, or Demand side) to keep communication with the relevant advertisers and the auction processing time under a strict limit, which is achieved in practice by placing a threshold k on the number of invited advertisers (e.g., $k = 20$, or $k = 50$). Some platforms already have or anticipate facing in the near future an excessive number of prospective advertisers, which results in a two-phase process: the platform first selects a subset of k bidders out of n candidates and then runs an auction only for that subset.

The above *Bidder Selection Problem* (BSP) is formalized (see [17, 3]) as follows. There are n prospective bidders with known independent prior distributions $(D_i)_{i\in[n]}$ for their values. The platform can only select and then run an auction for $k < n$ of them. It would like to choose this set with the objective of either maximizing its expected revenue or social welfare. The latter problem for second-price (single-item) auction is equivalent to the following fundamental algorithmic question: select k out of n independent random variables, with the objective of maximizing the expected maximum. It has received significant attention under different names: model driven optimization [9], k-MAX problem [6], team selection with test scores [13], subset selection for expected maximum [17], non-adaptive ProbeMax problem [22]. More recent work also looked at the revenue objective [17, 3] and other auction formats [3]. However, the main focus of all previous studies has been limited to single-item auctions.

In this work, we study the Bidder Selection Problem for a more general class of position auctions for welfare and revenue objectives[1]. Even more importantly, there is a large gap between theoretical results and practical algorithms for the BSP (see the overview of results and discussion below) and thus it is important to find theoretical approaches that lead to practically relevant algorithms.

Prior Results for the BSP. The basic BSP of welfare maximization (second-price auction) was shown to be NP-hard by Goel et al. [9] and later by Mehta et al. [17]. However, there are a few known Polynomial Time Approximation Schemes (PTAS) based on different ideas. First, Chen et al. [6] gave a dynamic programming based polynomial time approximation scheme (PTAS). Later, Mehta et al. [17] proposed an Efficient PTAS (EPTAS) with almost linear dependency on the number of bidders n. They discretize and then categorize all distributions into a constant ($C(\varepsilon)$ depending on the approximation guarantee $1-\varepsilon$) number of types and then do brute force search for k distributions. Segev and Singla [22] use a general framework based on a multi-dimensional extension of *Santa Claus* problem [2] with certain structural assumptions to get an EPTAS for *non-adaptive ProbMax* problem (equivalent to welfare-maximization of single-item auction).

Mehta et al. [17] also studied the revenue maximization BSP for the second-price auction, which is equivalent to maximizing the expectation of the second largest value among selected k random variables. They showed impossibility of any constant factor approximation polynomial time algorithm under either of the exponential time, or the planted clique hypothesises. Bei at al. [3] studied BSP with the revenue objective under multiple auction formats including Myerson's auction

[1]We are only aware of a single work that went beyond single-item format. Segev and Singla give a brief explanation in [22] how their results for single-item case can be extended to ℓ-unit auctions (non-adaptive Top-r ProbeMax problem) by separating the case when ℓ is large and when ℓ is small. However, their approach fails to generalize to position auctions as the argument hinges on the parameter ℓ, which is specific to ℓ-unit auctions.

and gave constant factor approximation for the second-price auction with anonymous reserve. They also introduced another optimization framework for the BSP under costs, which is significantly more challenging than the BSP under capacity constraint[2].

On the practical side, Klienberg and Raghu [13] and Mehta et al. [17] consider a few simple algorithms with constant approximation worst-case guarantees. Also, Mehta et al. [17] did extensive numerical experiments and observed that most of those heuristics perform much better than their worst-case performance guarantees. Interestingly, Mehta et al. [17] did not implement and test their theoretical EPTAS against those heuristics.

Practice vs. Theory. There are some useful theoretical algorithmic insights about BSP. However, they are far from forming a complete picture. Specifically, Chen et al. [6] were the first to observe that BSP is a special case of *monotone submodular maximization* and show that it is easier than the general problem under the cardinality constraint. Another insight from later EPTASes [17, 22] is that it is possible to reduce the large dependency on n, but at a rather hefty cost of at least doubly exponential dependency on ε (it is not clear that these EPTASes are better than the PTAS of Chen et al. [6] for mostly relevant in practice regimes with not too large $n \in [50, 500]$ and small $\varepsilon < 0.1$).

All the PTAS algorithms [6, 17, 22] are not easily adaptable to the problem's variations and are rather cumbersome. As such they are highly impractical and are less attractive than simpler approximation algorithms proposed in [13, 17]. Indeed, the $(1-\varepsilon)$ PTAS of [6] relies on a dynamic program of $n^{t(\varepsilon)}$ size and would be quite slow even for a relatively small $n = 50$ and large $\varepsilon = 0.2$; the doubly exponential dependency on ε of the EPTASes results in a huge running time even for a large $\varepsilon = 0.2$ ($2^{(1/\varepsilon)^{1/\varepsilon}} > 2^{3000}$). Furthermore, the experiments in [17] suggest that most of the heuristic algorithms may consistently outperform the previous PTAS algorithms even on synthetically generated data (which supposedly is more challenging than the real data). Indeed, all PTAS algorithms rely on multiplicative discretization of the input distributions, which leads to a likely loss of welfare up to the size of the discretization gap. At the same time, simple heuristic algorithms like Greedy for submodular optimization utilize all information about the distributions and usually produce much better results than their respective worst-case guarantees on almost all imaginable inputs of BSP. Hence, despite multiple attempts, it is still unclear whether there is a practically relevant algorithm with the worst-case performance that approaches optimum.

Position Auctions. A widely used position auction environment was formulated in [24, 8]. Position auctions generalize ℓ-unit auctions (in particular, single-item auction) and provide a simplified model of AdWords. There are m sorted positions that appear alongside the search results and n advertisers competing for these m slots. Each slot $j \in [m]$ has a click-through rate w_j (a click probability on j-th position), which translates into the value $v_i \cdot w_j$ for advertiser i if i's ad is displayed at j-th position[3]. The slots are ordered from highest click-through rate to lowest, $1 \geq w_1 \geq w_2 \geq \ldots \geq w_m \geq 0$. Every advertiser may be assigned to at most one position.

One of the most commonly used formats in advertising industry is the Generalized Second Price (GSP) auction. It was shown in [24, 8] that GSP of any position auction with any set of bidders has a Nash equilibrium equivalent to the welfare-maximizing outcome of the VCG. In some cases (e.g.,

[2]E.g., general (monotone) submodular maximization admits a $(1-1/e)$-approximation under capacity constraint, but is NP-hard to approximate within any constant factor under costs.

[3]This is a standard theoretical model of position auction as described, e.g., in a textbook [12]. For simplicity, it is assumed that bidders' click-through rates (CTR) depend only on their positions. It is straightforward to extend this model to the case where a click's probability on advertiser i in j-th position is the product $w_j \cdot \texttt{CTR}_i$. In this extension we can think of bidder's i value as $v_i \cdot \texttt{CTR}_i \sim D_i$.

Google's auction for selling contextual ads [25]) the platform's auction format is based directly on the VCG. Thus, in the context of Bidder Selection Problem with the welfare-maximization objective, it is most natural to study the VCG mechanism for the invited set of bidders. For the revenue objective, a natural auction format is the optimal Myerson's auction. Another option is to study BSP with the revenue objective for the GSP format (or by the revenue equivalence the VCG format). However, the latter variant is too hard even for the basic single-item auction environment [17].

1.1 Our Results

We propose a novel relaxation of the Bidder Selection Problem (BSP), which we call Poisson (or Poisson-Chernoff) relaxation. It has the following theoretical guarantees.

1. The relaxation is a continuous maximization problem with a concave objective that can be solved in time polynomial in n and k. In fact, the objective of this relaxation is a nicely structured algebraic function that lends itself to efficient convex minimization solvers.

2. With small adjustments (summarized in Algorithm 4), the relaxed objective *converges* at the rate $1 - O(k^{-1/4})$ to the actual social welfare of fractional BSP as the problem size k grows (Theorem 5.2). The standard rounding of this fractional solution suffers only a small loss of $O(k^{-1/2})$, yielding $\big(1 - O(k^{-1/4})\big)$-approximation (Theorem 5.4) for the integral BSP.

3. For the special case of the single-item auction, the Algorithm 3 achieves better convergence rate of $1 - O\big(\sqrt{\ln k/k}\big)$ (Theorem 4.9).

These results have immediate theoretical implications and can be applied in practice unlike all previous PTAS algorithms, as the above algorithms have far superior running time to the point where they outperform many simple heuristics. Furthermore, our results bring new theoretical insight into BSP for the more general environment of position auctions: the BSP converges to a polynomial-time solvable optimization problem as the problem size k grows.

Theoretical Implications. Our Algorithm 4 gives a $(1-\varepsilon)$ approximation to the BSP for any position auction with $\varepsilon = \Omega(k^{-1/4})$ and works in polynomial time (independent of ε). On the other hand, for small values of ε ($\varepsilon = O(k^{-1/4})$) a straightforward exhaustive search algorithm gives a perfect solution in $O(n^k) = n^{\mathrm{poly}(1/\varepsilon)}$ time. I.e., the combination of our relaxation with the brute force algorithm yields a PTAS:

Corollary 1.1. *Bidder Selection Problem for any position auction with either social welfare or revenue objective admits a $(1-\varepsilon)$ PTAS that runs in $n^{\mathrm{poly}(1/\varepsilon)}$ time.*

Mehta et al. [17] and Segev & Singla [22] gave two different EPTASes for single-item auction BSP, i.e., a PTAS that works in $O(\mathrm{poly}(n,k)) \cdot f(\varepsilon)$ time for some function of ε. It might appear to be better than the running time in Corollary 1.1; however, both papers [17, 22] have at least *doubly exponential dependency*[4] on ε in $f(\varepsilon)$, i.e., $f(\varepsilon) = \Omega\big(2^{(1/\varepsilon)^{1/\varepsilon}}\big)$. That running time is prohibitively large even for a fairly large ε, e.g., when $\varepsilon = 0.2$ there is $2^{(1/\varepsilon)^{1/\varepsilon}} > 2^{3000}$.

Moreover, it turns out that the same algorithm as in Corollary 1.1 is an EPTAS with a much better dependency on ε than previous PTAS algorithms under a mild assumption that $k \geq \log n$:

[4] Segev and Singla [22] do not provide an explicit algorithm, but rather give an existential result for EPTAS. However, at least doubly exponential dependency on ε seems unavoidable for their approach to work.

Corollary 1.2. *BSP for position auctions with either social welfare, or revenue objective admits a $(1-\varepsilon)$ EPTAS for any $k \geq \log n$ that runs in $O(\mathrm{poly}(n,k)) + 2^{O(\varepsilon^{-8})}$ time.*

Proof. The Algorithm 4 has $O(\mathrm{poly}(n,k))$ running time. We run brute force algorithm only when $\varepsilon = O(k^{-1/4})$, i.e., when $k^2 = O(\varepsilon^{-8})$. Then its running time is not more than $n^k \leq 2^{k^2} = 2^{O(\varepsilon^{-8})}$, since $k \geq \log n$. □

Corollaries 1.1 and 1.2 indicate that previous EPTASes may have merit over our relaxation approach (coupled with the basic brute force algorithm) only for extremely small values of $k = o(\log n)$. However, this regime is not relevant in the applications of BSP, where a typical instance has $k = \Theta(n)$: ad platform does not select $k = 10$ bidders out of $n = 1000$, but rather does $k = 20$ out of $n = 30$ or $n = 40$ candidates.

Applicability in Practice. Our relaxation is a white-box approach that can be easily adapted to different scenarios. E.g., if some bidders have to be included in the final solution (which is often the case in industry because of contract obligations), then our approach gives the same approximation with these additional constraints[5]. Also, if we expect the problem instance to satisfy certain reasonable criteria, then a few steps in our algorithm can be removed or simplified to fit the specific domain, which results in simpler and more efficient solutions (see Algorithms 1, 2). Furthermore, our approach works for any discretization of the input distributions unlike previous PTASes, which all used rather rigid multiplicative discretization structures.

We did extensive numerical experiments of our approach on several generated data sets summarized in Section 6. We slightly simplified our theoretical Algorithm 4 to avoid the hard-coded efficiency loss of $\left(1 - O(k^{-1/4})\right)$ in the approximation factor (the running time was not affected) and got the following experimental results:

- Our Algorithm 4 *outperforms* the staple Greedy algorithm for monotone submodular maximization: it has comparable or better approximate efficiency, and *significantly better running time* (e.g., on all large data points with $n = 3000$, $k = 300$ our algorithm worked under 11 seconds while Greedy took more than 19 minutes to complete). Interestingly, all previous PTASes regarded Greedy as a subroutine with negligible running time.

- We used Local Search heuristic as a benchmark and tested it against the optimum produced by the brute force algorithm on small data points ($n = 50$, $k = 5$). The Local Search produced optimal results except for one experimental setting, where it was off by less than 10^{-5}.

- The approximate efficiency losses compared to the Local Search were negligible (less than 1% in the worst case) on all data points. However, the Local Search (despite working in polynomial in n, k time with only $O(k)$ iterations) did poorly in terms of the running time: it took more than 20 hours on every large ($n = 3000$, $k = 300$) instance.

Our results further corroborate the experimental finding of Mehta et al. [17] that simple approximation algorithms such as Greedy have quite good performance for BSP. In fact, we are not aware of any instance on which Greedy would yield less than 0.9 of the optimum solution.

[5] In fact, the approximation guarantee holds not only for the global objective, but also for the surplus objective, i.e., the additional gain to welfare or revenue the platform gets from the non-fixed bidders.

1.2 Other Related Work

Mehta et al. [17] mention a few other scenarios besides bidder selection with similar mathematical formulations. The applications range from a two-tier solution for scoring documents in a search result [5], to filtering initial proposals in procurement auctions [21, 23], to voting theory [20].

Poisson approximation is a well-developed technique from probability theory and statistics. A survey [18] mentions at least twenty different results on the basic question of approximating the sum of independent Bernoulli random variables by the Poisson distribution. In statistics, Poisson approximation is commonly used in Extreme Value Theory (EVT) with applications to structural and geological engineering, traffic prediction, and finance (see, e.g., a book [19]). It has also been used in theoretical computer science, e.g., [15] used Le Cam's Poisson approximation theorem for stochastic bin packing and knapsack problems and also for EUM problem introduced in [14].

In fact, the expected utility maximization (EUM) is closely related to our objective. EUM is formulated as choosing a feasible subset S out of n random variables $X_1, \ldots, X_n$ to maximize $\mathbf{E}[u(\sum_{i \in S} X_i)]$, where u is a given utility function. The problem has been studied under capacity [4] or other combinatorial constraints [14, 15, 26] with a non-linear (typically concave) utility function. The BSP for single-item auction, i.e., the k-MAX problem, has a similar objective $\mathbf{E}[\max_{i \in S}\{X_i\}]$ to EUM but with max operator instead of the sum. When the distributions of X_i are unknown, EUM becomes an online learning problem. Chen et al. [6] gave the first PTAS for k-MAX problem, but their main focus is on Combinatorial Multi-Armed Bandits.

2 Preliminaries

A set of n bidders wish to receive some service and each bidder $i \in [n]$ has a private non-negative value $v_i \in \mathbb{R}_{\geq 0}$ indicating how much they are willing to pay for it. We denote the vector of bidder values as $\mathbf{v} = (v_i)_{i \in [n]}$. By the revelation principle, we can restrict our attention to incentive compatible and individually rational single-round auctions $\mathcal{A}$, where each bidder i submits a sealed bid b_i to the auctioneer. The auctioneer then decides on a feasible allocation vector $\mathbf{a}(\mathbf{b}) = (a_i(\mathbf{b}))_{i \in [n]}$ and payments $\mathbf{p}(\mathbf{b}) = (p_i(\mathbf{b}))_{i \in [n]}$. The incentive compatibility and individual rationality mean that by bidding truthfully $b_i = v_i$, each bidder $i \in [n]$ (a) maximizes her utility $u_i(b_i, \mathbf{b}_{\text{-}i}) \overset{\text{def}}{=\!=} v_i \cdot a_i(b_i, \mathbf{b}_{\text{-}i}) - p_i(b_i, \mathbf{b}_{\text{-}i})$ and (b) receives non-negative utility $u_i(v_i, \mathbf{b}_{\text{-}i}) \geq 0$. The seminal VCG mechanism (a second-price auction in the case of single-item auction) is an example of incentive compatible mechanism that also maximizes *social welfare* $\mathrm{SW}(\mathbf{a}, \mathbf{v}) = \sum_{i=1}^{n} a_i \cdot v_i$.

We study auctions in the Bayesian setting, where it is assumed that bidder values are drawn independently from known prior distributions $\mathbf{v} \sim \mathbf{D} = \prod_{i \in [n]} D_i$. We also use $D_i(\tau) = \mathbf{Pr}_{v_i \sim D_i}[v_i \leq \tau]$ to denote the cumulative distribution function. The auction designer is usually concerned about two objectives: the expected *social welfare* $\mathrm{SW} = \mathbf{E}_{\mathbf{v} \sim \mathbf{D}}[\mathrm{SW}(\mathbf{a}(\mathbf{v}), \mathbf{v})]$, and *revenue* $\textsc{Rev} = \mathbf{E}_{\mathbf{v} \sim \mathbf{D}}[\sum_{i \in [n]} p_i(\mathbf{v})]$. The VCG mechanism maximizes the welfare on every valuation profile $\mathbf{v}$, and thus maximizes SW in expectation for any prior $\mathbf{D}$. The well-known Myerson's auction maximizes $\textsc{Rev}$. This auction reduces the problem of revenue maximization to *virtual welfare* maximization by transforming values $(v_i)_{i \in [n]}$ to virtual values $\varphi_i(v_i)$ for regular distribution D_i and by doing ironing $\overline{\varphi}_i(v_i)$ for irregular distribution D_i. I.e., the expected revenue of the Myerson's auction can be written as $\textsc{Rev} = \mathbf{E}_{\mathbf{v}}[\mathrm{SW}(\mathbf{a}(\mathbf{v}), \varphi(\mathbf{v}))]$ for regular distributions and $\textsc{Rev} = \mathbf{E}_{\mathbf{v}}[\mathrm{SW}(\mathbf{a}(\mathbf{v}), \overline{\varphi}(\mathbf{v}))]$ for general distributions.

We consider the following auction environments with different feasible allocations.

Auction Environments. The single-item auction is an environment with the feasible allocations given by $\{\mathbf{a} : \sum_{i\in[n]} a_i \le 1\}$. A more general ℓ-unit auction environment for $\ell \in \mathbb{N}$ is given by the feasibility constraints: $\sum_{i\in[n]} a_i \le \ell$ and $a_i \in [0,1]$ for all $i \in [n]$. A position auction environment further generalizes ℓ-unit auctions. It is specified by a sorted weight vector $\mathbf{w} = (1 \ge w_1 \ge w_2 \ge \ldots \ge w_n \ge 0)$, which represents the click-through rate probabilities for n advert positions[6]. Every bidder may get at most one position and each advert position can be assigned to at most one advertiser. Formally, the feasibility allocations can be specified with an assignment function $\pi : [n] \to [n]$ of n advertisers to n sorted slots as follows: $\{\mathbf{a} : \exists\pi, \quad a_i \in [0, w_{\pi(i)}]$ for all $i \in [n]\}$.

Bidder Selection. In the Bidder Selection Problem (BSP) the seller first decides $x_i \in \{0,1\}$ which bidders $i \in [n]$ to invite to the auction. The selected set of bidders S may not exceed a certain capacity $k \ge |S|$, i.e., $\sum_{i\in[n]} x_i \le k$. Then the auctioneer runs an optimal auction $\mathcal{A}$ for the set S of invited bidders: VCG mechanism for the welfare objective, and Myerson for the revenue objective. Since revenue of the Myerson's auction can be rewritten as the expected virtual welfare with independent (ironed) virtual values $(\varphi_i(v_i) : v_i \sim D_i)_{i\in S}$, the BSP for the revenue maximization is equivalent to the BSP for the welfare maximization. Thus it suffices to only consider the welfare maximization problem. Specifically, we denote by $\mathbf{v} \sim \mathbf{x} \cdot \mathbf{D}$ the independent draws of $v_i \sim x_i \cdot D_i$ (i.e., $v_i \sim D_i$ when $x_i = 1$, and $v_i = 0$ when $x_i = 0$) for all $i \in [n]$, then the BSP of the VCG mechanism for any ℓ-unit/position auction with weights $\mathbf{w}$ can be written as follows:

$$\mathsf{OPT}(\ell) \stackrel{\text{def}}{=\!=} \max_{\substack{\mathbf{x}\in\{0,1\}^n \\ |\mathbf{x}|_1 \le k}} \mathop{\mathbf{E}}_{\mathbf{v}\sim\mathbf{x}\cdot\mathbf{D}} \left[\sum_{i=1}^{\ell} v_{(i)}\right], \qquad \mathsf{OPT}(\mathbf{w}) \stackrel{\text{def}}{=\!=} \max_{\substack{\mathbf{x}\in\{0,1\}^n \\ |\mathbf{x}|_1 \le k}} \mathop{\mathbf{E}}_{\mathbf{v}\sim\mathbf{x}\cdot\mathbf{D}} \left[\sum_{i=1}^{n} w_i \cdot v_{(i)}\right],$$

where $v_{(i)}$ is the i-th largest value among $\{v_i\}_{i\in[n]}$. We want to obtain good approximation algorithms for these BSPs. I.e., we would like to find in polynomial time $\mathbf{x} \in \{0,1\}^n$ with $|\mathbf{x}|_1 \le k$ such that $\mathbf{E}_{\mathbf{v}\sim\mathbf{x}\cdot\mathbf{D}}[\sum_{i=1}^{\ell} v_{(i)}] \ge (1-\varepsilon)\,\mathsf{OPT}(\ell)$ and $\mathbf{E}_{\mathbf{v}\sim\mathbf{x}\cdot\mathbf{D}}[\sum_{i=1}^{n} w_i \cdot v_{(i)}] \ge (1-\varepsilon)\,\mathsf{OPT}(\mathbf{w})$ for a small $\varepsilon > 0$. To simplify presentation, we assume that all distributions $(D_i)_{i\in[n]}$ have finite supports and are given explicitly as the algorithm's input[7].

It has been observed in the previous work on single-item auction that the BSP's objective is a monotone submodular function of the set $S = \{i : x_i = 1\}$ of invited bidders, i.e., $\mathrm{SW}(S) + \mathrm{SW}(T) \ge \mathrm{SW}(S \cap T) + \mathrm{SW}(S \cup T)$ for any $S, T \subseteq [n]$. The same property holds for the ℓ-unit and position auctions with an almost identical proof (see [6]).

We also consider the standard (in submodular optimization literature) multi-linear extension of the BSP objective. I.e., for a fractional $\mathbf{x} \in [0,1]^n$, we invite each bidder i to the auction independently with probability x_i. We employ the same notation $\mathbf{v} \sim \mathbf{x} \cdot \mathbf{D}$ as in the integral problem, where $v_i \sim x_i \cdot D_i$ means that we first decide whether to invite $\xi_i \in \{0,1\}$ bidder i according to a Bernoulli distribution $\xi_i \sim \mathrm{Ber}(x_i)$, then draw their value $v_i \sim \xi_i \cdot D_i$. The capacity constraint transforms into the bound on the expected number of invited bidders $\sum_{i=1}^{n} x_i \le k$. The next two sections 4 and 5 are primarily concerned with this fractional relaxation of BSP. We discuss in Section 5.2 how to obtain a good solution to the integral BSP from the fractional problem.

[6] The number of available positions m is usually smaller than the number of bidders n, in which case we simply let $w_{m+1} = \cdots = w_n = 0$.

[7] Previous works on single-item auction study continuous distributions and then discretize their supports in certain ways, so that auction's welfare only changes by a very small amount. Any such discretization also works for position auctions.

2.1 Mathematical Formulation

In this section, we give a fractional relaxation of BSP. As the welfare of any position auction can be written as a linear combination of ℓ-unit auctions, we begin with a fractional relaxation of BSP for the ℓ-unit auction. Specifically, the expected social welfare with a fractional set of bidders $\mathbf{x}$ is

$$\mathrm{SW}(\mathbf{x}, \ell) = \mathop{\mathbf{E}}_{\mathbf{v}\sim\mathbf{x}\cdot\mathbf{D}}\left[\sum_{i=1}^{\ell} v_{(i)}\right],$$

where $v_{(i)}$ denotes the i-th largest value among $\{v_i\}_{i\in\{1,2,\ldots,n\}}$. Hence, the following mathematical program represents BSP for ℓ-unit auction:

$$\begin{array}{ll} \text{Maximize} & \mathrm{SW}(\mathbf{x}, \ell) \\ \text{Subject To} & \sum_{i=1}^{n} x_i \leq k, \qquad x_i \in [0,1] \quad \forall i \in \{1,2,\ldots,n\}. \end{array} \tag{1}$$

As ℓ is fixed, whenever it is clear from the context, we will simply write $\mathrm{SW}(\mathbf{x})$.

Bernoulli Representation. We now derive an explicit formula for the social welfare $\mathrm{SW}(\mathbf{x})$. Fixing a threshold τ, the expected number of bidders with values exceeding τ among the highest ℓ bidders is $\mathbf{E}_{\mathbf{v}\sim\mathbf{x}\cdot\mathbf{D}}[\min(\sum_{i=1}^{n}\mathbb{I}[v_i \geq \tau], \ell)]$. Therefore, by integrating[8] over $\tau \in [0, +\infty)$, we get

$$\mathrm{SW}(\mathbf{x}) = \int_0^{+\infty} \mathop{\mathbf{E}}_{\mathbf{v}\sim\mathbf{x}\cdot\mathbf{D}}\left[\min\left(\sum_{i=1}^{n} \mathbb{I}\Big[v_i \geq \tau\Big],\, \ell\right)\right] \mathrm{d}\tau.$$

Note that $\mathbb{I}[v_i \geq \tau]$ for $v_i \sim x_i \cdot D_i$ is a Bernoulli random variable, and $\mathbf{E}_{\mathbf{v}\sim\mathbf{x}\cdot\mathbf{D}}[\min(\sum_{i=1}^{n}\mathbb{I}[v_i \geq \tau], \ell)]$ is the minimum of a sum of independent Bernoulli random variables and ℓ. To simplify notations, we explicitly define the probabilities of Bernoulli random variables $\mathbb{I}[v_i \geq \tau]$:

$$q_i(x_i, \tau) \overset{\text{def}}{=\!=} \mathop{\mathbf{Pr}}_{v_i\sim x_i\cdot D_i}[v_i \geq \tau] = x_i \cdot (1 - D_i(\tau)), \qquad \text{and let} \quad \mathbf{q}(\mathbf{x}, \tau) \overset{\text{def}}{=\!=} (q_i)_{i\in[n]}. \tag{2}$$

Definition 2.1 (Bernoulli Objective)**.** *For a real vector $\mathbf{q} \in [0,1]^n$ and ℓ, the **Bernoulli objective term** of $\mathbf{q}$ and ℓ is a function $\mathrm{H}_{\mathrm{ber}}(\mathbf{q}, \ell)$:*

$$\mathrm{H}_{\mathrm{ber}}(\mathbf{q}, \ell) \overset{\text{def}}{=\!=} \mathop{\mathbf{E}}_{\mathbf{z}\sim\mathrm{Ber}(\mathbf{q})}\left[\min\left(\sum_{i=1}^{n} z_i,\, \ell\right)\right], \quad \text{then} \quad \mathrm{SW}(\mathbf{x}) = \int_0^{+\infty} \mathrm{H}_{\mathrm{ber}}(\mathbf{q}(\mathbf{x}, \tau), \ell)\, \mathrm{d}\tau. \tag{3}$$

Position Auctions. A position auction is given by a vector of non-negative weights[9] $\mathbf{w} : (w_1 \geq w_2 \geq \cdots \geq w_n \geq 0)$. The highest social welfare we get from the set S of invited bidders is $\sum_{i=1}^{|S|} v_{(i)} \cdot w_i$, where $v_{(1)} \geq \cdots \geq v_{(|S|)}$ are ordered values of bidders in S. Thus the expected social welfare for a fractional set $\mathbf{x}$ is

$$\mathrm{SW}(\mathbf{x}, \mathbf{w}) = \mathop{\mathbf{E}}_{\mathbf{v}\sim\mathbf{x}\cdot\mathbf{D}}\left[\sum_{i=1}^{n} v_{(i)} \cdot w_i\right] = \sum_{\ell=1}^{n} (w_\ell - w_{\ell+1}) \cdot \mathrm{SW}(\mathbf{x}, \ell),$$

[8]The function inside the integral is piece-wise constant, i.e., it is constant between consecutive values of the threshold τ in the supports of $\{D_i\}_{i\in[n]}$.

[9]Usually weights are only for the first k slots, as we only select a set of k bidders. In this case, we simply assume that $w_{k+1} = \ldots = w_n = 0$.

where $w_{n+1} \stackrel{\text{def}}{=} 0$. Then the respective fractional BSP program for position auctions is as follows.

$$\begin{array}{ll} \text{Maximize} & \mathrm{SW}(\mathbf{x}, \mathbf{w}) \\ \text{Subject To} & \sum_{i=1}^{n} x_i \le k, \qquad x_i \in [0,1] \quad \forall i \in \{1, 2, \dots, n\}. \end{array} \tag{4}$$

We will often omit dependency on $\mathbf{w}$ in SW whenever it is clear from the context. We further consider the *Bernoulli representation* for position auctions:

$$\mathrm{SW}(\mathbf{x}, \mathbf{w}) = \int_0^{+\infty} \mathrm{H}_{\mathrm{ber}}(\mathbf{q}(\mathbf{x}, \tau), \mathbf{w})\, \mathrm{d}\tau, \quad \text{where} \quad \mathrm{H}_{\mathrm{ber}}(\mathbf{q}, \mathbf{w}) \stackrel{\text{def}}{=} \sum_{\ell=1}^{n} (w_\ell - w_{\ell+1})\, \mathrm{H}_{\mathrm{ber}}(\mathbf{q}, \ell). \tag{5}$$

As in (2), $\mathbf{q}(\mathbf{x}, \tau)$ represents the probabilities of each bidder's value exceeding τ. $\mathrm{H}_{\mathrm{ber}}(\mathbf{q}, \mathbf{w})$ is called the **Bernoulli objective term** for position auctions.

3 Technical Overview

The fractional relaxation (4) is still neither convex nor concave and thus is too unwieldy. The central idea of our paper is to use instead *Poisson approximation* to the Bernoulli objective terms $\mathrm{H}_{\mathrm{ber}}(\mathbf{q}(\mathbf{x}, \tau), \mathbf{w})$. Specifically, we substitute each Bernoulli random variable $z \sim \mathrm{Ber}(p)$ with $p = q_i(x_i, \tau)$ by the Poisson random variable $y \sim \mathrm{Pois}(p)$ with the same expectation as z. We use the following Poisson objective term $\mathrm{H}_{\mathrm{pois}}(\mathbf{q}, \ell)$ to approximate $\mathrm{H}_{\mathrm{ber}}(\mathbf{q}, \ell)$.

$$\mathrm{H}_{\mathrm{pois}}(\mathbf{q}, \ell) \stackrel{\text{def}}{=} \mathop{\mathbf{E}}_{\mathbf{y} \sim \mathrm{Pois}(\mathbf{q})} \left[\min\left(\sum_{i=1}^{n} y_i,\ \ell \right) \right] = \mathop{\mathbf{E}}_{Y \sim \mathrm{Pois}(\sum_{i=1}^{n} q_i)} [\min(Y,\ \ell)].$$

The advantage of the Poisson approximation $\mathrm{H}_{\mathrm{pois}}(\mathbf{q}, \ell)$ is that the resulting functions $\mathrm{H}_{\mathrm{pois}}(\mathbf{q}, \ell)$ and $\mathrm{H}_{\mathrm{pois}}(\mathbf{q}, \mathbf{w})$ are *concave* in $\mathbf{q}$. This in turn allows us to efficiently solve the analogous to (4), the optimization problem for the Poisson approximation.

This Poisson approximation works well[10] in the following situations.

1. When the probabilities $p_i = q_i(x_i, \tau)$ of $z_i \sim \mathrm{Ber}(p_i)$ are small. It allows us to handle the crucial contribution to the welfare comprised of small probability tail events for large thresholds τ.

2. When $\mathbf{E}[\sum_{i=1}^{n} z_i]$ is large (when thresholds τ are small). Indeed, by Chernoff bounds the sums of independent random variables (both Bernoulli and Poisson) $\sum_{i=1}^{n} z_i$ and $\sum_{i=1}^{n} y_i$ are close to their expectations. In fact, we simply use Chernoff objective term $\mathrm{H}_{\mathrm{cher}}(\mathbf{q}, \ell) \stackrel{\text{def}}{=} \min(\sum_{i=1}^{n} q_i,\ \ell)$ instead of Poisson approximation in this case.

3. When ℓ is large. In this case, either the concentration inequality gives a good approximation when $\mathbf{E}[\sum_{i=1}^{n} z_i] = \mathbf{E}[\sum_{i=1}^{n} y_i]$ is large, or when this expectation is much smaller than ℓ then the probability that either of $\sum_{i=1}^{n} z_i$ or $\sum_{i=1}^{n} y_i$ exceeds the threshold ℓ is small.

Then the algorithmic framework for the BSP is rather straightforward. We need to solve the following *concave* optimization problem of $\mathbf{x}$:

$$\begin{aligned} \max_{\mathbf{x} \in [0,1]^n} \quad & \widetilde{\mathrm{SW}}(\mathbf{x}, \mathbf{w}) \stackrel{\text{def}}{=} \sum_{\ell=1}^{n} (w_\ell - w_{\ell+1}) \int_0^{+\infty} \mathrm{H}_{\mathrm{pois}}(\mathbf{q}(\mathbf{x}, \tau), \ell)\, \mathrm{d}\tau \\ \text{Subject to} \quad & \sum_{i=1}^{n} x_i \le k, \end{aligned} \tag{6}$$

[10] The sum of Poisson random variables approaches the sum of Bernoulli random variables in TV and other statistical distances.

where $\mathrm{H}_{\text{pois}}(\mathbf{q}(\mathbf{x}, \tau), \ell) = \mathbf{E}_{Y \sim \text{Pois}(\sum q_i)}[\min(Y, \ell)]$ and all functions under the integral are piece-wise constant with the number of pieces bounded by the size of the union of the supports of D_i. The optimal fractional solution $\mathbf{x}$ can be computed rather efficiently using continuous convex optimization. Finally, we use a rounding procedure (similar to the multi-linear extension of submodular function) to the fractional solution of (6) to obtain an integral solution of BSP with only a small loss to the approximation guarantee.

This simple algorithm already achieves quite good approximation results on all our test cases with the rounding part either not needed or having very few fractional variables. Moreover, the function inside the integral has a nice algebraic form and, as our experiments indicate, the running time of standard continuous optimization methods is significantly better than the running time of the staple greedy algorithm for submodular optimization.

However, obtaining formal theoretical guarantees requires some extra work. We need to make sure that Poisson is a good approximation to all Bernoulli objective terms. We achieve this with a resource augmentation trick inspired by the EPTAS from Mehta et al. [17]. We fix a small set of bidders $|S_{\text{fix}}| = \varepsilon \cdot k$ such that all other bidders $i \notin S_{\text{fix}}$ have only a small chance to exceed a certain threshold $\eta > 0$, while the expected number of bidders in S_{fix} with values greater than or equal to any threshold $\tau \leq \eta$ is relatively large. We take set S_{fix} into the solution and use a linear combination of Chernoff and Poisson approximations for the remaining bidders. Specifically, we apply Chernoff approximation for the thresholds τ below η, and Poisson approximation for the thresholds $\tau > \eta$ and agents in $M = [n] \setminus S_{\text{fix}}$ (we also need to modify the Poisson objective term due to the fact that $x_i = 1$ for $\forall i \in S_{\text{fix}}$). Interestingly, it was fine in all of our experiments to run the more straightforward algorithm without fixing the set S_{fix}. The running times were the same and the simpler algorithm typically produced better approximation results.

4 Warm-up

In this section, we consider the Bidder Section Problem (BSP) for ℓ-unit auctions. It serves as a warm-up for more general position auctions presented in Section 5. Our main goal here will be to illustrate our approach and analysis ideas rather than to derive independent results for ℓ-unit auctions. We still obtain an independent result for single-item auction with better approximation guarantees than our more general result for position auctions.

In Section 2.1, we wrote the expected welfare $\mathrm{SW}(\mathbf{x}, \ell)$ as an integral of $\mathrm{H}_{\text{ber}}(\mathbf{q}, \ell)$. We will introduce several approximations of H_{ber} that make the optimization problem (1) easy to solve.

4.1 Chernoff Approximation

Definition 4.1 (Chernoff Objective)**.** *For a real vector $\mathbf{q} \in [0,1]^n$ and ℓ, the **Chernoff objective term** of $\mathbf{q}$ and ℓ is a function $\mathrm{H}_{\text{cher}}(\mathbf{q}, \ell)$:*

$$\mathrm{H}_{\text{cher}}(\mathbf{q}, \ell) \stackrel{\text{def}}{=\!=} \min\left(\mathop{\mathbf{E}}_{\mathbf{z} \sim \text{Ber}(\mathbf{q})}\left[\sum_{i=1}^{n} z_i\right], \ell\right) = \min\left(\sum_{i=1}^{n} q_i, \ell\right).$$

Formally, we have the following approximation guarantees of H_{ber} by H_{cher}:

Lemma 4.2. *For all* $\mathbf{q} \in [0,1]^n, \ell \in \mathbb{N}^+$*, let* $\lambda = \sum_{i=1}^n q_i$*, the following properties hold.*

(a) $$\mathrm{H}_{\mathrm{ber}}(\mathbf{q},\ell) \le \mathrm{H}_{\mathrm{cher}}(\mathbf{q},\ell) \le 7 \cdot \mathrm{H}_{\mathrm{ber}}(\mathbf{q},\ell).$$

(b) $$\mathrm{H}_{\mathrm{cher}}(\mathbf{q},\ell) - \mathrm{H}_{\mathrm{ber}}(\mathbf{q},\ell) \le \frac{3}{\sqrt{\lambda}} \cdot \mathrm{H}_{\mathrm{cher}}(\mathbf{q},\ell) \le \frac{21}{\sqrt{\lambda}} \cdot \mathrm{H}_{\mathrm{ber}}(\mathbf{q},\ell).$$

(c) $$\mathrm{H}_{\mathrm{cher}}(\mathbf{q},\ell) - \mathrm{H}_{\mathrm{ber}}(\mathbf{q},\ell) \le \frac{5}{\sqrt{\ell}} \cdot \mathrm{H}_{\mathrm{cher}}(\mathbf{q},\ell) \le \frac{35}{\sqrt{\ell}} \cdot \mathrm{H}_{\mathrm{ber}}(\mathbf{q},\ell).$$

The proof of Lemma 4.2 is deferred to Appendix A.1. It is a rather straightforward application of Chernoff bounds with careful handling of extreme cases. The merit of such an approximation is that $\mathrm{H}_{\mathrm{cher}}$ is a simple concave function which admits efficient maximization algorithms:

Claim 4.3. *For all input distributions* $\mathbf{D}$ *and* $\ell \in \mathbb{N}^+$*, the following properties hold.*

(a) *When* ℓ *is fixed,* $\mathrm{H}_{\mathrm{cher}}(\mathbf{q},\ell)$ *is a concave function of* $\mathbf{q}$*.*

(b) *When* $\mathbf{D}$ *and* ℓ *are fixed, the approximate social welfare*

$$\widetilde{\mathrm{SW}}(\mathbf{x}) \stackrel{\mathrm{def}}{=\!=} \int_0^{+\infty} \mathrm{H}_{\mathrm{cher}}(\mathbf{q}(\mathbf{x},\tau),\ell)\,\mathrm{d}\tau \quad \textit{is a concave function of } \mathbf{x}.$$

Proof. We first prove (a). Let $\lambda = \sum_{i=1}^n q_i$, then $\mathrm{H}_{\mathrm{cher}}(\mathbf{q},\ell) = \min(\lambda,\ell)$ is a concave function of λ. As λ is linear in $\mathbf{q}$, $\mathrm{H}_{\mathrm{cher}}(\mathbf{q},\ell)$ is a concave function of $\mathbf{q}$. To prove (b), recall that $q_i(x_i,\tau) = x_i \cdot (1 - D_i(\tau))$. Note that $\mathbf{q}(\mathbf{x},\tau)$ is linear in $\mathbf{x}$ for $\forall \tau \in [0,+\infty)$. Furthermore, $\mathrm{H}_{\mathrm{cher}}(\mathbf{q}(\mathbf{x},\tau),\ell)$ is a concave function in $\mathbf{q}$ by (a) and thus is concave in $\mathbf{x}$. Thus $\widetilde{\mathrm{SW}}$ is an integral (i.e., linear combination with positive coefficients) of concave functions in $\mathbf{x}$, which implies that $\widetilde{\mathrm{SW}}$ is concave in $\mathbf{x}$. $\square$

Algorithm for Inputs with Large ℓ. With Lemma 4.2 and Claim 4.3, we are ready to present our Algorithm 1 for the case when ℓ is large.

> **Algorithm 1**: Approximation to Fractional BSP for ℓ-unit Auctions with Large ℓ.
>
> Return the optimal solution $\widetilde{\mathbf{x}}^*$ to the continuous concave optimization program in $\mathbf{x}$:
>
> $$\begin{array}{ll} \text{Maximize} & \widetilde{\mathrm{SW}}(\mathbf{x},\ell) = \int_0^{+\infty} \min(\sum_{i=1}^n x_i \cdot (1 - D_i(\tau)),\ \ell)\,\mathrm{d}\tau \\ \text{Subject to} & \sum_{i=1}^n x_i \le k, \qquad x_i \in [0,1] \quad \forall i \in \{1,2,\ldots,n\}. \end{array}$$

To analyze Algorithm 1, first notice that according to Claim 4.3, the optimization program is concave. Thus, we can use standard concave optimization methods like gradient ascent or the ellipsoid method to solve the program in polynomial time[11]. Moreover, by Lemma 4.2 (c), we have for any $\mathbf{x} \in [0,1]^n$ and $\tau \ge 0$,

$$\mathrm{H}_{\mathrm{cher}}(\mathbf{q}(\mathbf{x},\tau),\ell) - \mathrm{H}_{\mathrm{ber}}(\mathbf{q}(\mathbf{x},\tau),\ell) \le \frac{5}{\sqrt{\ell}} \cdot \mathrm{H}_{\mathrm{cher}}(\mathbf{q}(\mathbf{x},\tau),\ell).$$

[11] The function inside the integral is piece-wise constant with finitely many points of discontinuity at $\tau \in \cup_{i\in[n]} \mathsf{supp}(D_i)$. Thus it can be precisely and efficiently computed in polynomial time.

We get $\widetilde{\mathrm{SW}}(\mathbf{x}) - \mathrm{SW}(\mathbf{x}) \leq \frac{5}{\sqrt{\ell}} \cdot \widetilde{\mathrm{SW}}(\mathbf{x})$ by integrating the above bound over $\tau \in [0, \infty)$ for each $\mathbf{x} \in [0,1]^n$. Similarly, by integrating the bound from Lemma 4.2 (a) over $\tau \in [0, \infty)$, we get $\widetilde{\mathrm{SW}}(\mathbf{x}) \geq \mathrm{SW}(\mathbf{x})$. Now, let $\mathbf{x}^*$ be the optimal solution to the fractional BSP. Then

$$\mathrm{SW}(\widetilde{\mathbf{x}}^*) \geq \left(1 - \frac{5}{\sqrt{\ell}}\right)\widetilde{\mathrm{SW}}(\widetilde{\mathbf{x}}^*) \geq \left(1 - \frac{5}{\sqrt{\ell}}\right)\widetilde{\mathrm{SW}}(\mathbf{x}^*) \geq \left(1 - \frac{5}{\sqrt{\ell}}\right)\mathrm{SW}(\mathbf{x}^*),$$

where to get the first inequality we use the upper bound on $\widetilde{\mathrm{SW}}$ for $\mathbf{x} = \widetilde{\mathbf{x}}^*$; the second inequality holds, as $\widetilde{\mathbf{x}}^*$ maximizes the $\widetilde{\mathrm{SW}}$ objective; the third inequality is a lower bound on $\widetilde{\mathrm{SW}}$ for $\mathbf{x} = \mathbf{x}^*$. This means that $\widetilde{\mathbf{x}}^*$ is an $\left(1 - \frac{5}{\sqrt{\ell}}\right)$-approximate solution.

Thus Algorithm 1 runs in polynomial time and its approximation ratio gets closer to 1 as ℓ gets larger. I.e., informally speaking, the problem gets easier as ℓ grows.

4.2 Poisson Approximation

The previous section 4.1 covers the case of large ℓ. The same approach also works for thresholds τ with large $\sum_i q_i(x_i, \tau)$. However, Chernoff bounds are not good for small probability tail events, which typically are the main contributors to the welfare for small ℓ. To handle these tail events we use *Poisson approximation*, which is the central idea of our paper.

Definition 4.4 (Poisson Objective)**.** *For a real vector* $\mathbf{q} \in [0,1]^n$ *and integer* ℓ*, the* ***Poisson objective term*** *is a function* $\mathrm{H}_{\mathrm{pois}}(\mathbf{q}, \ell)$ *given by*

$$\mathrm{H}_{\mathrm{pois}}(\mathbf{q}, \ell) \stackrel{\mathrm{def}}{=\!=} \mathop{\mathbf{E}}_{\mathbf{y} \sim \mathrm{Pois}(\mathbf{q})}\left[\min\left(\sum_{i=1}^{n} y_i,\ \ell\right)\right] = \mathop{\mathbf{E}}_{Y \sim \mathrm{Pois}(\sum_{i=1}^{n} q_i)}[\min(Y,\ \ell)].$$

Note that the latter equality is an important property of Poisson distribution: the sum of independent Poisson random variables $y_i \sim \mathrm{Pois}(q_i)$ follows the Poisson distribution $\mathrm{Pois}(\sum_{i=1}^{n} q_i)$. Another crucial property is concavity of the Poisson approximation.

Claim 4.5 (Concavity of Poisson)**.** $\mathrm{H}_{\mathrm{pois}}(\mathbf{q}, \ell)$ *is a concave function in* $\mathbf{q} \in [0,1]^n$ *for* $\forall \ell \in \mathbb{N}$.

Proof. Let $\lambda = \sum_{i=1}^{n} q_i$. Notice that $\mathrm{H}_{\mathrm{pois}}(\mathbf{q}, \ell)$ only depends on λ, which is linear in $\mathbf{q}$. We use $\mathrm{H}_{\mathrm{pois}}(\lambda)$ to represent this function. Then, we only need to prove that $\mathrm{H}_{\mathrm{pois}}(\lambda)$ is concave in λ. We first rewrite

$$\mathrm{H}_{\mathrm{pois}}(\lambda) = \ell - \sum_{j=0}^{\ell-1} \mathop{\mathbf{Pr}}_{Y \sim \mathrm{Pois}(\lambda)}[Y = j] \cdot (\ell - j) = \ell - \sum_{j=0}^{\ell-1} \frac{\lambda^j}{j!} e^{-\lambda} \cdot (\ell - j).$$

Then by a straightforward differentiation of the partial series we get

$$\frac{\mathrm{d}}{\mathrm{d}\lambda}\mathrm{H}_{\mathrm{pois}} = \sum_{j=0}^{\ell-1} \frac{\lambda^j}{j!} e^{-\lambda}, \qquad \frac{\mathrm{d}^2}{\mathrm{d}\lambda^2}\mathrm{H}_{\mathrm{pois}} = -e^{-\lambda}\frac{\lambda^{\ell-1}}{(\ell-1)!} < 0.$$

Therefore, $\mathrm{H}_{\mathrm{pois}}(\mathbf{q}, \ell)$ is concave in λ, and thus concave in $\mathbf{q}$. ∎

Approximation Guarantees. There are many known Poisson approximation results (see, e.g., a survey [18]) for the sum of independent Bernoulli random variables, e.g., in total variation, earth mover's, uniform (a.k.a. Kolmogorov) distances. These are typically absolute approximation guarantees, while we need relative approximations similar to Chernoff approximations from Lemma 4.2. As our goal is to handle small probability tail events, we assume that each bidder's value v_i has only a small probability δ to be greater than zero, i.e., $\forall i \in [n]$, $\mathbf{Pr}[v_i > 0] \le \delta$. The following Poisson absolute approximation result will be useful to us:

Lemma 4.6 ([7, Lemma 11.3.V, p. 162])**.** *Let $(z_i \sim \mathrm{Ber}(q_i))_{i=1}^n$ be n independent Bernoulli random variables with $q_i \le \delta$ for $\forall i \in [n]$. Let $Z \overset{\text{def}}{=\!=} \sum_{i=1}^n z_i$ and $Y \sim \mathrm{Pois}(\lambda)$, where $\lambda \overset{\text{def}}{=\!=} \sum_{i=1}^n q_i$. Then*

$$\text{(total variation distance)} \qquad \sum_{j=0}^{\infty} \left|\mathbf{Pr}[Z = j] - \mathbf{Pr}[Y = j]\right| \le \delta. \tag{7}$$

With the help of Lemma 4.6 we can derive relative approximations:

Lemma 4.7. *Suppose $\mathbf{q} \in [0, \delta]^n$ and $\ell \in \mathbb{N}$. Then*

$$\begin{aligned}
&\text{(a)} && |\mathrm{H}_{\mathrm{ber}}(\mathbf{q}, \ell) - \mathrm{H}_{\mathrm{pois}}(\mathbf{q}, \ell)| \le 17.5 \cdot \delta \cdot \mathrm{H}_{\mathrm{ber}}(\mathbf{q}, \ell) && (\forall \delta,\ \ell),\\
&\text{(b)} && 0 \le \mathrm{H}_{\mathrm{ber}}(\mathbf{q}, 1) - \mathrm{H}_{\mathrm{pois}}(\mathbf{q}, 1) \le \delta \cdot \mathrm{H}_{\mathrm{ber}}(\mathbf{q}, 1) && (\ell = 1).
\end{aligned}$$

The proof of Lemma 4.7 is deferred to Appendix A.2.

Algorithm for Small Tail Probabilities. Assume that each bidder's value v_i has at most δ probability to be greater than zero. Then Lemma 4.7 and Claim 4.5 suggest the following algorithm.

> **Algorithm 2**: Approximation to Fractional BSP for Tail Probabilities ℓ-unit Auctions.
>
> Return the optimal solution $\widetilde{\mathbf{x}}^*$ to the concave program in $\mathbf{x}$:
>
> $$\begin{array}{ll} \text{Maximize} & \widetilde{\mathrm{SW}}(\mathbf{x}, \ell) \overset{\text{def}}{=\!=} \int_0^{+\infty} \mathrm{H}_{\mathrm{pois}}(\mathbf{q}(\mathbf{x}, \tau), \ell)\, \mathrm{d}\tau \\ \text{Subject To} & \sum_{i=1}^n x_i \le k, \qquad x_i \in [0,1] \quad \forall i \in \{1, 2, \ldots, n\}. \end{array}$$

We can solve the above program efficiently via standard concave function maximization methods, as the objective $\widetilde{\mathrm{SW}}$ is concave in $\mathbf{x}$. Indeed, from Claim 4.5 we know that $\mathrm{H}_{\mathrm{pois}}(\mathbf{q}, \ell)$ is concave in $\mathbf{q}$ and, since $\mathbf{q}(\mathbf{x}, \tau)$ is linear in $\mathbf{x}$ for every fixed τ, $\mathrm{H}_{\mathrm{pois}}$ is also concave in $\mathbf{x}$. As $\widetilde{\mathrm{SW}}(\mathbf{x}, \ell)$ is an integral (non-negative linear combination) of $\mathrm{H}_{\mathrm{pois}}(\mathbf{q}(\mathbf{x}, \tau), \ell)$, $\widetilde{\mathrm{SW}}$ is concave in $\mathbf{x}$.

From Lemma 4.7, we obtain the following approximation guarantees of $\mathrm{SW}(\mathbf{x}, \ell)$ by $\widetilde{\mathrm{SW}}(\mathbf{x}, \ell)$.

$$\left|\mathrm{SW}(\mathbf{x}, \ell) - \widetilde{\mathrm{SW}}(\mathbf{x}, \ell)\right| \le \int_0^{+\infty} |\mathrm{H}_{\mathrm{ber}}(\mathbf{q}(\mathbf{x}, \tau), \ell) - \mathrm{H}_{\mathrm{pois}}(\mathbf{q}(\mathbf{x}, \tau), \ell)|\, \mathrm{d}\tau \le 17.5\, \delta \cdot \mathrm{SW}(\mathbf{x}, \ell). \tag{8}$$

Let $\mathbf{x}^*$ be the best solution of the original problem (1). We have

$$\mathrm{SW}(\widetilde{\mathbf{x}}^*, \ell) \ge \widetilde{\mathrm{SW}}(\widetilde{\mathbf{x}}^*, \ell) - 17.5\, \delta \cdot \mathrm{SW}(\widetilde{\mathbf{x}}^*, \ell) \ge \widetilde{\mathrm{SW}}(\mathbf{x}^*, \ell) - 17.5\, \delta \cdot \mathrm{SW}(\mathbf{x}^*, \ell) \ge (1 - 35\, \delta)\mathrm{SW}(\mathbf{x}^*, \ell).$$

Hence, Algorithm 2 is $(1 - 35\, \delta)$-approximation.

4.3 Single-Item Auction

We saw in Section 4.1 that BSP gets easier as ℓ grows. We also applied a Poisson approximation framework in the previous section 4.2 to handle BSP for small probability tail events. Here we illustrate how to apply Poisson approximation to the case of small ℓ without any assumption on the distributions $\mathbf{D}$. We focus on the case of $\ell = 1$, i.e., single-item auction, as it was extensively studied in previous work.

The idea to avoid the extra assumption $\mathbf{Pr}[v_i > 0] \le \delta$ is a simple resource augmentation argument: reserve a small fraction $\varepsilon \cdot k$ of bidders, which affects the final approximation by at most $(1 - \varepsilon)$ factor due to the submodularity of BSP as a set function of the invited bidders. We adopt the following variant for our Poisson approximation, which is inspired by the so-called Core-Tail decomposition from [17].

Recall that for the single-item auction $\mathrm{SW}(\mathbf{x}) = \int_0^{+\infty} \mathrm{H}_{\mathrm{ber}}(\mathbf{q}(\mathbf{x}, \tau), 1)\, \mathrm{d}\tau = \int_0^{+\infty} \mathbf{Pr}_{\mathbf{v} \sim \mathbf{x} \cdot \mathbf{D}}[\exists v_i \ge \tau]\, \mathrm{d}\tau$. We fix a small set S_{fix} of $\varepsilon \cdot k$ bidders, such that $\mathbf{Pr}_{\mathbf{v} \sim \mathbf{D}}[\exists i \in S_{\text{fix}} : v_i \ge \eta] \ge 1 - \varepsilon$ for a threshold η. This allows us to take care of thresholds τ in a low range $\tau \in [0, \eta]$ by including S_{fix} in the solution (i.e., make $x_i = 1$ for all $i \in S_{\text{fix}}$). Naturally, we want to pick bidders with higher probabilities $\mathbf{Pr}[v_i > \eta]$ into S_{fix}, which means that for the high range thresholds $\tau > \eta$ we get the small probability property for each bidder $i \notin S_{\text{fix}}$. This allows us to reduce BSP to the case of small probabilities tail events for $\tau > \eta$ and bidders $i \in [n] \setminus S_{\text{fix}}$, which can be effectively solved by the Poisson approximation. Formally, we can get the following guarantees for S_{fix}.

Claim 4.8 (Small Bidder Set)**.** *Let $\varepsilon \ge \sqrt{\frac{\ln k}{k}}$ be a multiple of $1/k$. We can find in polynomial time a threshold $\eta \ge 0$ and a set $S_{\text{fix}} \subset [n]$ of size $|S_{\text{fix}}| = \varepsilon \cdot k$, such that*

$$\text{(a)} \ \mathop{\mathbf{Pr}}_{\mathbf{v} \sim \mathbf{D}}[\exists i \in S_{\text{fix}} : v_i \ge \eta] \ge 1 - \frac{1}{k}; \qquad \text{(b)} \ \forall i \notin S_{\text{fix}}, \ \mathop{\mathbf{Pr}}_{v_i \sim D_i}[v_i > \eta] < \varepsilon.$$

Proof. Recall that all distributions $\{D_i\}_{i \in [n]}$ have finite support. Thus we can search through all thresholds τ in polynomial time. There must be two consecutive threshold values η and $\eta_+ > \eta$ such that the number of bidders $|\{i : \mathbf{Pr}[v_i \ge \eta] \ge \varepsilon\}| \ge \varepsilon \cdot k$ with large tail probabilities $\mathbf{Pr}[v_i \ge \eta] \ge \varepsilon$ is at least $\varepsilon \cdot k$, but a similar number of bidders $|\{i : \mathbf{Pr}[v_i > \eta] = \mathbf{Pr}[v_i \ge \eta_+] \ge \varepsilon\}| < \varepsilon \cdot k$ for the next threshold value η_+ is strictly less than $\varepsilon \cdot k$.

Let us place each bidder i with $\mathbf{Pr}[v_i > \eta] \ge \varepsilon$ into S_{fix} and fill the remaining positions in S_{fix} up to size $\varepsilon \cdot k$ (so that $|S_{\text{fix}}| = \varepsilon \cdot k$) with bidders from $\{i : \mathbf{Pr}[v_i \ge \eta] \ge \varepsilon > \mathbf{Pr}[v_i \ge \eta_+]\}$. Then, every bidder $i \notin S_{\text{fix}}$ has $\mathbf{Pr}[v_i > \eta] = \mathbf{Pr}[v_i \ge \eta_+] < \varepsilon$ as required by condition (b). On the other hand, $|S_{\text{fix}}| = \varepsilon \cdot k$ and $\mathbf{Pr}[v_i \ge \eta] \ge \varepsilon$ for every $i \in S_{\text{fix}}$, i.e., condition (a) is satisfied since

$$\mathop{\mathbf{Pr}}_{\mathbf{v} \sim \mathbf{D}}[\exists i \in S_{\text{fix}} : v_i \ge \eta] \ge 1 - (1 - \varepsilon)^{\varepsilon \cdot k} \ge 1 - e^{-\varepsilon^2 \cdot k} \ge 1 - \frac{1}{k},$$

where to get the second inequality, we used the fact that $(1 - \frac{1}{x})^x < e^{-1}$ for any $x \ge 1$. □

After selecting such set S_{fix} and threshold η, we are ready to give the complete description of Algorithm 3.

Algorithm 3: Fractional BSP for Single-Item Auction.

Fix $\varepsilon = \sqrt{\frac{\ln k}{k}}$ rounded up to a multiple of $1/k$, then do the following steps:

1. Find (η, S_{fix}) as in Claim 4.8. Set $x_i = 1$ for $\forall i \in S_{\text{fix}}$.
2. For the remaining bidders $M \stackrel{\text{def}}{=\!=} [n] \setminus S_{\text{fix}}$ let the Poisson approximation $\widetilde{\mathrm{SW}}(\mathbf{x}_{_M})$ be
$$\widetilde{\mathrm{SW}}(\mathbf{x}_{_M}) \stackrel{\text{def}}{=\!=} \left(1 - \frac{1}{k}\right)\eta + \int_{\eta}^{+\infty} \widetilde{\mathrm{H}}_{\text{pois}}(\mathbf{q}(\mathbf{x}_{_M}, \tau))\, \mathrm{d}\tau, \qquad \text{where} \tag{9}$$
$$\widetilde{\mathrm{H}}_{\text{pois}}(\mathbf{q}_{_M}) \stackrel{\text{def}}{=\!=} r_\tau + (1 - r_\tau) \cdot \mathrm{H}_{\text{pois}}(\mathbf{q}_{_M}, \ell = 1), \quad \text{and} \quad r_\tau \stackrel{\text{def}}{=\!=} \Pr_{\mathbf{v}\sim\mathbf{D}}[\exists i \in S_{\text{fix}} : v_i \geq \tau]. \tag{10}$$
3. Return $\widetilde{\mathbf{x}}^* = (\mathbf{1}_{S_{\text{fix}}}, \widetilde{\mathbf{x}}^*_{_M})$, where $\widetilde{\mathbf{x}}^*_{_M}$ is the solution to the concave program in $\mathbf{x}_{_M}$:
$$\begin{array}{ll} \text{Maximize} & \widetilde{\mathrm{SW}}(\mathbf{x}_{_M}) \\ \text{Subject To} & \sum_{i\in M} x_i \leq k - \varepsilon \cdot k, \quad x_i \in [0,1] \quad \forall i \in M. \end{array} \tag{11}$$

In the algorithm, we ignore thresholds $\tau \in [0, \eta]$, as by taking S_{fix} we have already achieved high success probability of at least $1 - \frac{1}{k}$ by Claim 4.8. For the high range thresholds $\tau > \eta$, we first observe that as the result of fixing set S_{fix}, the probability that there is a bidder with value greater than the threshold τ becomes

$$\mathrm{H}_{\text{ber}}(\mathbf{q}, 1) = \Pr[\exists i \in [n] : v_i \geq \tau] = r_\tau + (1 - r_\tau) \cdot \Pr[\exists i \in M : v_i \geq \tau],$$

where r_τ is a constant that we can easily compute. Hence, we respectively adjust the Poisson approximation term $\widetilde{\mathrm{H}}_{\text{pois}}(\mathbf{q}_{_M})$ in the algorithm according to (10) (we slightly abuse notations by writing $\mathrm{H}_{\text{pois}}(\mathbf{q}_{_M})$ instead of $\mathrm{H}_{\text{pois}}(\mathbf{q})$: for the coordinates $i \notin M$ we have $\{q_i = 0,\, i \in S_{\text{fix}}\}$).

Theorem 4.9. *Algorithm 3 for single-item auction works in polynomial time and is a $(1 - 2\varepsilon)$-approximation, i.e., a $\left(1 - O\left(\sqrt{\ln k/k}\right)\right)$-approximation to the fractional BSP.*

Proof. We first verify that Algorithm 3 is polynomial. Note that the step (1) works in polynomial time by Claim 4.8. For each τ in the support of D_i we calculate in polynomial time the constants $r_\tau \in [0,1]$. Both $\widetilde{\mathrm{H}}_{\text{pois}}(\mathbf{q}_{_M})$ for each $\mathbf{x}_{_M}$ and τ in the support and $\widetilde{\mathrm{SW}}(\mathbf{x}_{_M})$ for each $\mathbf{x}_{_M}$ can be computed in polynomial time. Moreover, all first and second order partial derivatives of $\widetilde{\mathrm{SW}}(\mathbf{x}_{_M})$ can be computed in the same way as the integral of respective derivatives of $\widetilde{\mathrm{H}}_{\text{pois}}(\mathbf{x}_{_M})$. Furthermore, it is easy to see that $\widetilde{\mathrm{SW}}(\mathbf{x}_{_M})$ is a concave function in $\mathbf{x}_{_M}$, since it is a positive linear combination of constant terms (such as $(1-1/k)\eta$ and r_τ) and concave functions $\mathrm{H}_{\text{pois}}(\mathbf{x}_{_M})$ according to Claim 4.5. Hence, we can find the optimal solution $\widetilde{\mathbf{x}}^*_{_M}$ in polynomial time using standard concave (first or second order) maximization methods.

To prove an approximation guarantee of $1 - 2\varepsilon$ for Algorithm 3, we first derive the following approximations of $\mathrm{SW}(\mathbf{x}_{_M}, \mathbf{1}_{S_{\text{fix}}})$ by $\widetilde{\mathrm{SW}}(\mathbf{x}_{_M})$ similar to (8) (but in a special case of $\ell = 1$).

Lemma 4.10. $\forall \mathbf{x}_{_M} \in [0,1]^{|M|}, \qquad 0 \leq \mathrm{SW}(\mathbf{x}_{_M}, \mathbf{1}_{S_{\text{fix}}}) - \widetilde{\mathrm{SW}}(\mathbf{x}_{_M}) \leq \varepsilon \cdot \mathrm{SW}(\mathbf{x}_{_M}, \mathbf{1}_{S_{\text{fix}}}).$

Proof. Recall that by (3) the Social Welfare $\mathrm{SW}(\mathbf{x})$ for $\mathbf{x} = (\mathbf{x}_{_M}, \mathbf{1}_{S_{\text{fix}}})$ and $\ell = 1$ is

$$\mathrm{SW}(\mathbf{x}) = \int_0^{\eta} \mathrm{H}_{\text{ber}}(\mathbf{q}(\mathbf{x}, \tau))\, \mathrm{d}\tau + \int_{\eta}^{+\infty} \mathrm{H}_{\text{ber}}(\mathbf{q}(\mathbf{x}, \tau))\, \mathrm{d}\tau.$$

For the low range $\tau \in [0, \eta]$ we have $\mathrm{H}_{\mathrm{ber}}(\mathbf{q}(\mathbf{x}, \tau)) \in [1 - \frac{1}{k}, 1]$ due to the choice of S_{fix} in Claim 4.8. It is well approximated by the respective term $\left(1 - \frac{1}{k}\right)\eta$ in (9).

For the high range $\tau > \eta$, we apply Lemma 4.7 (b) with $\delta = \varepsilon$ and get the following bound:

$$\mathrm{H}_{\mathrm{ber}}(\mathbf{q}) - \widetilde{\mathrm{H}}_{\mathrm{pois}}(\mathbf{q}) \;=\; (1 - r_\tau)(\mathrm{H}_{\mathrm{ber}}(\mathbf{q}_{M}) - \mathrm{H}_{\mathrm{pois}}(\mathbf{q}_{M})) \;\leq\; (1 - r_\tau)\varepsilon \cdot \mathrm{H}_{\mathrm{ber}}(\mathbf{q}_{M}) \;\leq\; \varepsilon \cdot \mathrm{H}_{\mathrm{ber}}(\mathbf{q}).$$

On the other hand, $\mathrm{H}_{\mathrm{ber}}(\mathbf{q}) - \widetilde{\mathrm{H}}_{\mathrm{pois}}(\mathbf{q}) \geq 0$, as $\mathrm{H}_{\mathrm{ber}}(\mathbf{q}_{M}) \geq \mathrm{H}_{\mathrm{pois}}(\mathbf{q}_{M})$ by Lemma 4.7 (b). Thus

$$0 \leq \mathrm{SW}(\mathbf{x}) - \widetilde{\mathrm{SW}}(\mathbf{x}_{M}) \leq \frac{\eta}{k} + \varepsilon \cdot \int_{\eta}^{+\infty} \mathrm{H}_{\mathrm{ber}}(\mathbf{q}(\mathbf{x}, \tau))\, \mathrm{d}\tau \leq \max\left(\frac{1}{k-1}, \varepsilon\right)\mathrm{SW}(\mathbf{x}) = \varepsilon \cdot \mathrm{SW}(\mathbf{x}),$$

where the last equality holds since $\varepsilon \geq \sqrt{\frac{\ln k}{k}}$. □

Now we are ready to complete the proof of Theorem 4.9. Let $\mathbf{x}^*$ be the optimal solution to fractional BSP. We consider $\mathbf{x}^*_+ \in \mathbb{R}^n_{\geq 0}$ defined as $\mathbf{x}^*_+ \stackrel{\text{def}}{=\!=} (\mathbf{x}^*_{M}, \mathbf{1}_{S_{\mathrm{fix}}})$, so that $\mathbf{x}^*_+ \succeq \mathbf{x}^*$. Then, by Lemma 4.10 for $\mathbf{x} = \mathbf{x}^*_+$ we have

$$\widetilde{\mathrm{SW}}(\mathbf{x}^*_{M}) \geq (1 - \varepsilon) \cdot \mathrm{SW}(\mathbf{x}^*_+) \geq (1 - \varepsilon) \cdot \mathrm{SW}(\mathbf{x}^*).$$

On the other hand, by Lemma 4.10 for $\mathbf{x} = \widetilde{\mathbf{x}}^*$ we have

$$\mathrm{SW}(\widetilde{\mathbf{x}}^*) \geq \widetilde{\mathrm{SW}}(\widetilde{\mathbf{x}}^*_{M}) \geq \widetilde{\mathrm{SW}}\left(\frac{k - \varepsilon \cdot k}{k} \cdot \mathbf{x}^*_{M}\right) \geq (1 - \varepsilon) \cdot \widetilde{\mathrm{SW}}(\mathbf{x}^*_{M}) \geq (1 - 2\varepsilon) \cdot \mathrm{SW}(\mathbf{x}^*),$$

where the second inequality holds, as $\widetilde{\mathbf{x}}^*_{M}$ is the optimal solution to (11) and $\frac{k - \varepsilon \cdot k}{k} \cdot \mathbf{x}^*_{M}$ is a feasible solution; the third inequality holds, as $\widetilde{\mathrm{SW}}(\mathbf{x})$ is a concave function in $\mathbf{x}$ by Claim 4.5; the last inequality holds, as $\widetilde{\mathrm{SW}}(\mathbf{x}^*_{M}) \geq (1-\varepsilon) \cdot \mathrm{SW}(\mathbf{x}^*)$ and $(1-\varepsilon)^2 \geq 1 - 2\varepsilon$. This concludes the proof. □

5 Algorithm for Position Auctions

In this section, we consider Bidder Section Problem for position auctions. We first analyze the fractional BSP, and then explain in Section 5.2 how to do integral rounding of the fractional solution with only a small loss to the approximation guarantee. For the fractional BSP, we give an efficient polynomial time $(1 - O(\varepsilon))$-approximation algorithm under similar conditions on k, and ε to Theorem 4.9 for the single-item auction (see Theorem 5.2 for the exact statement).

Recall that the Bernoulli objective term for position auctions $\mathrm{H}_{\mathrm{ber}}(\mathbf{q}, \mathbf{w})$ is defined in Section 2.1. The expected welfare $\mathrm{SW}(\mathbf{x}) = \mathrm{SW}(\mathbf{x}, \mathbf{w})$ is written as an integral of $\mathrm{H}_{\mathrm{ber}}$. Below, we will also need the **Chernoff objective term** for position auctions defined as a non-negative linear combination of respective ℓ-unit auction terms:

$$\mathrm{H}_{\mathrm{cher}}(\mathbf{q}, \mathbf{w}) \stackrel{\text{def}}{=\!=} \sum_{\ell=1}^{n} (w_\ell - w_{\ell+1})\, \mathrm{H}_{\mathrm{cher}}(\mathbf{q}, \ell) = \sum_{\ell=1}^{n} (w_\ell - w_{\ell+1}) \min\left(\sum_{i=1}^{n} q_i, \ell\right).$$

We sometimes slightly abuse notations and write $\mathrm{H}_{\mathrm{ber}}(\mathbf{x}, \tau)$ and $\mathrm{H}_{\mathrm{cher}}(\mathbf{x}, \tau)$ instead of $\mathrm{H}_{\mathrm{ber}}(\mathbf{q}, \mathbf{w})$ and $\mathrm{H}_{\mathrm{cher}}(\mathbf{q}, \mathbf{w})$. Similar to Section 4, for high range thresholds we will use Poisson approximation.

In order to use effectively Poisson approximation we need the small probability assumption $q_i \leq \delta$, which was achieved by fixing a small bidder set S_{fix} in Section 4.3. However, unlike the case of single-item auction, for large ℓ and large $\mathbf{q}$, neither the Poisson approximation, nor a small set of bidders S_{fix} approximates well the welfare of ℓ-unit auction. Thus we have to modify the selection of the small set of bidders S_{fix} and incorporate Chernoff approximation in this process.

Fixing Small Bidder Set. The idea is again to fix a small set of bidders S_{fix} (set $x_i = 1$ for $i \in S_{\text{fix}}$) with $|S_{\text{fix}}| = \varepsilon \cdot k$ and make sure that all other bidders $i \notin S_{\text{fix}}$ have only a small probability $\mathbf{Pr}[v_i \geq \tau] \leq \delta$ to exceed any of the thresholds $\tau > \eta$ for certain $\eta > 0$. This allows us to use Poisson approximation for the high range thresholds $\tau > \eta$ and bidders $i \in M \stackrel{\text{def}}{=\!=} [n] \setminus S_{\text{fix}}$ in the same way we did it in Section 4.3, i.e., by recalculating the expected contribution of bidders in M given that $\mathbf{x}_{S_{\text{fix}}} = \mathbf{1}$. On the other hand, for the low range thresholds $\tau \leq \eta$, we would like to see a certain number ℓ^* of bidders $i \in S_{\text{fix}}$ to exceed the threshold $v_i \geq \tau$. To this end, we choose S_{fix} so that the expected number of bidders $i \in S_{\text{fix}}$ with $v_i \geq \tau$ is at least ℓ^*. We can achieve the following guarantees for ε, δ, and ℓ^*.

Claim 5.1 (Small Bidder Set). *Let $\varepsilon \in [\frac{\ell^*}{\delta \cdot k}, 1)$ be a multiple of $1/k$ for $\ell^* \in \mathbb{R}_{\geq 0} : \ell^* < k$ and $\delta \in (0,1)$. We can find in polynomial time a threshold $\eta \geq 0$ and a set $S_{\text{fix}} \subseteq [n]$ of size $|S_{\text{fix}}| = \varepsilon \cdot k$:*

$$\text{(a)}\ \ \forall\, 0 \leq \tau \leq \eta, \quad \sum_{i \in S_{\text{fix}}} \mathop{\mathbf{Pr}}_{v_i \sim D_i}[v_i \geq \tau] \geq \ell^*; \qquad \text{(b)}\ \ \forall\, i \notin S_{\text{fix}}, \quad \mathop{\mathbf{Pr}}_{v_i \sim D_i}[v_i > \eta] < \delta.$$

Proof. We search through all thresholds τ in the supports of $\{D_i\}_{i \geq 1}$ and find two consecutive threshold values η and $\eta_+ > \eta$ such that $|\{i : \mathbf{Pr}[v_i \geq \eta] \geq \delta\}| \geq \varepsilon \cdot k$, but a similar number of bidders $|\{i : \mathbf{Pr}[v_i > \eta] = \mathbf{Pr}[v_i \geq \eta_+] \geq \delta\}| < \varepsilon \cdot k$ for the next value η_+. As in Claim 4.8 we place each bidder i with $\mathbf{Pr}[v_i > \eta] \geq \delta$ into S_{fix} and fill the remaining positions in S_{fix} up to $\varepsilon \cdot k$ with bidders from $\{i : \mathbf{Pr}[v_i \geq \eta] \geq \delta > \mathbf{Pr}[v_i \geq \eta_+]\}$.

Thus, every bidder $i \notin S_{\text{fix}}$ has $\mathbf{Pr}[v_i > \eta] = \mathbf{Pr}[v_i \geq \eta_+] < \delta$ as required by condition (b). On the other hand, $|S_{\text{fix}}| = \varepsilon \cdot k$ and $\mathbf{Pr}[v_i \geq \eta] \geq \delta$ for every $i \in S_{\text{fix}}$, which implies (a), since

$$\forall\, \tau \leq \eta, \quad \sum_{i \in S_{\text{fix}}} \mathop{\mathbf{Pr}}_{v_i \sim D_i}[v_i \geq \tau] \geq \sum_{i \in S_{\text{fix}}} \mathop{\mathbf{Pr}}_{v_i \sim D_i}[v_i \geq \eta] \geq \delta \cdot |S_{\text{fix}}| = \delta \cdot \varepsilon \cdot k \geq \ell^*.$$

Thus we constructed in polynomial time the desired threshold η and set S_{fix}. $\square$

We need to balance three parameters ℓ^*, δ, and ε, which must satisfy conditions of Claim 5.1. Specifically, we choose $\ell^* = k^{1/2}$, $\varepsilon = k^{-1/4}$ rounded up to a multiple of $1/k$, and $\delta = \varepsilon \geq k^{-1/4}$. The Claim 5.1 lends itself to the following Algorithm 4 (which we will present shortly). We first outline some differences with Algorithm 3 for single-item auction.

Similar to the adjustments we did in Section 4.3 to $\widetilde{\mathrm{H}}_{\text{pois}}(\mathbf{x}_M)$ for high thresholds $\tau > \eta$, we also adjust respective Poisson approximation part for position auctions. We proceed in a similar way to Algorithm 3 and define the Poisson objective for $\mathbf{x}_M$ as a conditional expectation depending on $Z_{\text{fix}}(\tau) \stackrel{\text{def}}{=\!=} \sum_{i \in S_{\text{fix}}} \mathbb{I}[v_i \geq \tau]$ (as $\mathbf{x}_{S_{\text{fix}}} = \mathbf{1}_{S_{\text{fix}}}$, random variable $Z_{\text{fix}}(\tau)$ has Poisson binomial distribution). Indeed, we may calculate all probabilities $\mathbf{Pr}[Z_{\text{fix}}(\tau) = j]$ for each $j \in [0, \varepsilon \cdot k]$ in polynomial time and define the **adjusted Poisson objective term** as

$$\mathrm{G}_{\text{pois}}(\mathbf{x}_M, \tau) \stackrel{\text{def}}{=\!=} \sum_{j=0}^{|S_{\text{fix}}|} \mathbf{Pr}[Z_{\text{fix}}(\tau) = j] \cdot \left(\sum_{\ell=1}^{j} w_\ell + \sum_{\ell=j+1}^{n} (w_\ell - w_{\ell+1}) \cdot \mathrm{H}_{\text{pois}}(\mathbf{x}_M,\, \ell - j,\, \tau) \right) \tag{12}$$
$$\text{where} \qquad \mathrm{H}_{\text{pois}}(\mathbf{x}_M, \ell - j, \tau) = \mathop{\mathbf{E}}_{Y \sim \text{Pois}(\lambda_M)}[\min\{Y,\, \ell - j\}], \qquad \lambda_M = \sum_{i \in M} q_i(x_i, \tau).$$

The main difference with the single-item auction, however, is for the low range thresholds $\tau \leq \eta$. We only expect a relatively small number of bidders ℓ^* from S_{fix} to have values above τ, which may not be enough to cover contribution from the k bidders for low thresholds. To this end, we use Chernoff objective $\mathrm{H}_{\text{cher}}(\mathbf{x}, \tau) = \mathrm{H}_{\text{cher}}(\mathbf{q}(\mathbf{x}, \tau), \mathbf{w})$ to approximate $\mathrm{H}_{\text{ber}}(\mathbf{x}, \tau) = \mathrm{H}_{\text{ber}}(\mathbf{q}(\mathbf{x}, \tau), \mathbf{w})$

for $\mathbf{x} = (\mathbf{x}_M, \mathbf{1}_{S_{\text{fix}}})$. Importantly, unlike the case of high range thresholds, we do not recalculate H_{cher} as a function of $\mathbf{x}_M$, but use Chernoff approximation for the entire n-dimensional vector $\mathbf{x} = (\mathbf{x}_M, \mathbf{x}_{S_{\text{fix}}})$ with $\mathbf{x}_{S_{\text{fix}}} = \mathbf{1}_{S_{\text{fix}}}$. Thus, as $\ell^* = k^{1/2}$ grows with k, $\mathrm{H}_{\text{cher}}(\mathbf{x}, \tau) \to \mathrm{H}_{\text{ber}}(\mathbf{x}, \tau)$ by Lemma 4.2 (c).

5.1 Algorithm

Our main Algorithm 4 is summarized below. It essentially follows the same scheme as Algorithm 3 for single-item auction with two modifications we have discussed above.

Algorithm 4: Fractional BSP for Position Auctions

Let $\ell^* = k^{1/2}$; ε be $k^{-1/4}$ rounded up to a multiple of $1/k$ $\left(\varepsilon = \frac{\lceil k \cdot k^{-1/4} \rceil}{k}\right)$; $\delta = \varepsilon$.

1. Find η and S_{fix} according to Claim 5.1. Set $x_i = 1$ for $\forall i \in S_{\text{fix}}$. Let $M \stackrel{\text{def}}{=\!=} [n] \setminus S_{\text{fix}}$.
2. Define the approximate welfare using adjusted Poisson objective (12):

$$\widetilde{\mathrm{SW}}(\mathbf{x}_M) \stackrel{\text{def}}{=\!=} \int_0^{\eta} \mathrm{H}_{\text{cher}}((\mathbf{x}_M, \mathbf{1}_{S_{\text{fix}}}), \tau)\, \mathrm{d}\tau + \int_{\eta}^{+\infty} \mathrm{G}_{\text{pois}}(\mathbf{x}_M, \tau)\, \mathrm{d}\tau, \tag{13}$$

3. Return $\widetilde{\mathbf{x}}^* = (\widetilde{\mathbf{x}}^*_M, \mathbf{1}_{S_{\text{fix}}})$, where $\widetilde{\mathbf{x}}^*_M$ is the solution to the concave program in $\mathbf{x}_M$:

$$\begin{array}{ll} \text{Maximize} & \widetilde{\mathrm{SW}}(\mathbf{x}_M) \\ \text{Subject To} & \sum_{i \in M} x_i \le k - \varepsilon \cdot k, \quad x_i \in [0,1] \quad \forall i \in M. \end{array} \tag{14}$$

Theorem 5.2. *Algorithm 4 works in polynomial time and is a $(1 - O(\varepsilon))$-approximation, i.e., $(1 - 43\, k^{-1/4})$-approximation to the fractional BSP for any position auction.*

Proof. We first show that Algorithm 4 is polynomial. Indeed, step (1) works in polynomial time by Claim 5.1. For each τ in the support of $\{D_i\}_{i \in [n]}$ and $\mathbf{x}_M$ we can efficiently calculate $\mathrm{H}_{\text{cher}}((\mathbf{x}_M, \mathbf{1}_{S_{\text{fix}}}), \tau)$ and $\mathrm{G}_{\text{pois}}(\mathbf{x}_M, \tau)$, which allows us to compute $\widetilde{\mathrm{SW}}(\mathbf{x}_M)$ in polynomial time similarly to Algorithm 3. It is easy to see that $\widetilde{\mathrm{SW}}(\mathbf{x}_M)$ is a concave function in $\mathbf{x}_M$ by Claim 4.5 and Claim 4.3, as G_{pois} is a non-negative linear combination of constant terms and concave functions $\mathrm{H}_{\text{pois}}(\mathbf{x}_M, \ell - j, \tau)$, and H_{cher} is a non-negative linear combination of concave functions $\mathrm{H}_{\text{cher}}(\mathbf{q}(\mathbf{x}, \tau), \ell)$ in $\mathbf{x}$. Furthermore, given the representation of $\widetilde{\mathrm{SW}}(\mathbf{x}_M)$ as an integral of nice algebraic functions H_{cher} and G_{pois} we can also compute all first and second order derivatives of $\widetilde{\mathrm{SW}}(\mathbf{x}_M)$ in polynomial time. This allows us to find the optimal solution $\widetilde{\mathbf{x}}^*_M$ to (14) in polynomial time using standard concave (first or second order) maximization methods.

To get the stated approximation guarantee we first compare the original objective $\mathrm{SW}(\mathbf{x}_M, \mathbf{1}_{S_{\text{fix}}})$ in (4) with $\widetilde{\mathrm{SW}}(\mathbf{x}_M)$ and get the following Lemma.

Lemma 5.3. *For any $\mathbf{x} = (\mathbf{x}_M, \mathbf{1}_{S_{\text{fix}}})$ with $\mathbf{x}_M \in [0,1]^{|M|}$ and any weights vector $\mathbf{w}$,*

$$\left|\widetilde{\mathrm{SW}}(\mathbf{x}_M) - \mathrm{SW}(\mathbf{x})\right| \le 21\, \varepsilon \cdot \mathrm{SW}(\mathbf{x}).$$

Proof. We rewrite $\mathrm{SW}(\mathbf{x}, \mathbf{w})$ for $\mathbf{x} = (\mathbf{x}_M, \mathbf{1}_{S_{\text{fix}}})$ in the same form as (13).

$$\mathrm{SW}(\mathbf{x}) = \int_0^{\eta} \mathrm{H}_{\text{ber}}(\mathbf{x}, \tau)\, \mathrm{d}\tau + \int_{\eta}^{+\infty} \mathrm{G}_{\text{ber}}(\mathbf{x}_M, \tau)\, \mathrm{d}\tau, \qquad \text{where}$$

$$\mathrm{G}_{\mathrm{ber}}(\mathbf{x}_M,\tau) \stackrel{\text{def}}{=\!=} \sum_{j=0}^{|S_{\mathrm{fix}}|} \mathbf{Pr}[Z_{\mathrm{fix}}(\tau)=j]\cdot\left(\sum_{\ell=1}^{j} w_\ell + \sum_{\ell=j+1}^{n}(w_\ell - w_{\ell+1})\cdot \mathrm{H}_{\mathrm{ber}}(\mathbf{x}_M,\,\ell-j,\,\tau)\right).$$

We first compare corresponding adjusted Bernoulli and Poisson terms. Observe that $q_i(\mathbf{x},\tau)\le\delta=\varepsilon$ for each bidder $i\in M$ and threshold $\tau>\eta$ by condition (b) in Claim 5.1. Thus by Lemma 4.7 (a),

$$\begin{aligned}
&|\mathrm{G}_{\mathrm{pois}}(\mathbf{x}_M,\tau)-\mathrm{G}_{\mathrm{ber}}(\mathbf{x}_M,\tau)|\\
&\le \sum_{j=0}^{|S_{\mathrm{fix}}|}\mathbf{Pr}[Z_{\mathrm{fix}}(\tau)=j]\cdot\sum_{\ell=j+1}^{n}(w_\ell-w_{\ell+1})\cdot|\mathrm{H}_{\mathrm{pois}}(\mathbf{x}_M,\,\ell-j,\,\tau)-\mathrm{H}_{\mathrm{ber}}(\mathbf{x}_M,\,\ell-j,\,\tau)|\\
&\le \sum_{j=0}^{|S_{\mathrm{fix}}|}\mathbf{Pr}[Z_{\mathrm{fix}}(\tau)=j]\cdot\sum_{\ell=j+1}^{n}(w_\ell-w_{\ell+1})\cdot 17.5\,\delta\cdot\mathrm{H}_{\mathrm{ber}}(\mathbf{x}_M,\,\ell-j,\,\tau)\ \le\ 17.5\,\varepsilon\cdot\mathrm{G}_{\mathrm{ber}}(\mathbf{x}_M,\tau).
\end{aligned}$$

Next, we compare Chernoff and Bernoulli objectives ($\mathrm{H}_{\mathrm{cher}}$ and $\mathrm{H}_{\mathrm{ber}}$) for low range thresholds $\tau\le\eta$. As both $\mathrm{H}_{\mathrm{cher}}(\mathbf{q},\mathbf{w})$ and $\mathrm{H}_{\mathrm{ber}}(\mathbf{q},\mathbf{w})$ are non-negative linear combinations of respective terms for ℓ-unit auctions with $\sum_{i\in[n]}q_i(x_i,\tau)\ge\ell^*$ for $\tau\le\eta$, we get by Lemma 4.2 (b) that

$$|\mathrm{H}_{\mathrm{cher}}(\mathbf{q},\mathbf{w})-\mathrm{H}_{\mathrm{ber}}(\mathbf{q},\mathbf{w})|\le 21\cdot\frac{\mathrm{H}_{\mathrm{ber}}(\mathbf{q},\mathbf{w})}{\sqrt{\ell^*}}\le 21\,\varepsilon\cdot\mathrm{H}_{\mathrm{ber}}(\mathbf{q},\mathbf{w}).$$

Thus after combining the two bounds for high and low ranges of thresholds τ we get

$$\left|\widetilde{\mathrm{SW}}(\mathbf{x}_M)-\mathrm{SW}(\mathbf{x})\right|\le 21\,\varepsilon\cdot\int_0^{\eta}\mathrm{H}_{\mathrm{ber}}(\mathbf{x},\tau)\,\mathrm{d}\tau+17.5\,\varepsilon\cdot\int_{\eta}^{+\infty}\mathrm{G}_{\mathrm{ber}}(\mathbf{x}_M,\tau)\,\mathrm{d}\tau\le 21\,\varepsilon\cdot\mathrm{SW}(\mathbf{x}),$$

which concludes the proof of Lemma 5.3. $\square$

We complete the proof of the Theorem 5.2 in a similar way to Theorem 4.9. As before let $\mathbf{x}^*$ be the optimal solution to fractional BSP in (4). We consider $\mathbf{x}^*_+\in\mathbb{R}^n_{\ge 0}$ defined as $\mathbf{x}^*_+ \stackrel{\text{def}}{=\!=} (\mathbf{x}^*_M,\mathbf{1}_{S_{\mathrm{fix}}})$, so that $\mathbf{x}^*_+\succeq\mathbf{x}^*$. Then, by Lemma 5.3 for $\mathbf{x}=\mathbf{x}^*_+$ we have

$$\widetilde{\mathrm{SW}}(\mathbf{x}^*_M)\ge(1-21\,\varepsilon)\cdot\mathrm{SW}(\mathbf{x}^*_+)\ge(1-21\,\varepsilon)\cdot\mathrm{SW}(\mathbf{x}^*).$$

On the other hand, by Lemma 5.3 for $\mathbf{x}=\widetilde{\mathbf{x}}^*$ we have

$$\begin{aligned}
(1+21\,\varepsilon)\cdot\mathrm{SW}(\widetilde{\mathbf{x}}^*)\ge\widetilde{\mathrm{SW}}(\widetilde{\mathbf{x}}^*_M)\ge\widetilde{\mathrm{SW}}\left(\frac{k-\varepsilon\cdot k}{k}\cdot\mathbf{x}^*_M\right)&\ge(1-\varepsilon)\cdot\widetilde{\mathrm{SW}}(\mathbf{x}^*_M)\\
&\ge(1-\varepsilon)\cdot(1-21\,\varepsilon)\cdot\mathrm{SW}(\mathbf{x}^*),
\end{aligned}$$

where the second inequality holds, as $\widetilde{\mathbf{x}}^*_M$ is the optimal solution to (14) and $\frac{k-\varepsilon\cdot k}{k}\cdot\mathbf{x}^*_M$ is a feasible solution; the third inequality holds, as $\widetilde{\mathrm{SW}}(\mathbf{x})$ is a concave function in $\mathbf{x}$; the last inequality holds by the last lower bound on $\widetilde{\mathrm{SW}}(\mathbf{x}^*_M)$. Finally, as $\frac{(1-\varepsilon)(1-21\,\varepsilon)}{1+21\,\varepsilon}\ge 1-(1+2\cdot 21)\cdot\varepsilon$,

$$\mathrm{SW}(\widetilde{\mathbf{x}}^*)\ge(1-43\,\varepsilon)\cdot\mathrm{SW}(\mathbf{x}^*),$$

which concludes the proof of the theorem. $\square$

5.2 Rounding

We conclude Section 5 by presenting the rounding algorithm, which takes our solution $\widetilde{\mathbf{x}}^*$ to the fractional BSP produced by Algorithm 4 and returns a solution to the integral BSP.

Our fractional relaxation works as the standard multi-linear extension of submodular functions, which corresponds to sampling a random set of bidders $S \sim \prod_{i=1}^{n} \mathrm{Ber}(x_i)$ in the integral BSP. To align the notations for fractional and integral BSPs, we shall use vectors $\mathbf{y}, \mathbf{z} \in \{0,1\}^n$ for the respective sets of selected bidders. Specifically, we use $\mathbf{y} \sim \mathrm{Ber}(\mathbf{x})$ to represent the random set S in the multi-linear extension. Our rounding procedure is quite standard and proceeds as follows.

Algorithm 5: Rounding: algorithm for Integral BSP

1. Run Algorithm 4 to obtain a fractional solution $\mathbf{x}$.
2. Sample an integral solution $\mathbf{y} \sim \mathrm{Ber}(\mathbf{x})$, with $\mathbf{y} \in \{0,1\}^n$.
3. - If $|\mathbf{y}|_1 \leq k$, return $\mathbf{z} = \mathbf{y}$,
   - Else ($|\mathbf{y}|_1 > k$), return k bidders $\mathbf{z} \sim \binom{\mathbf{y}}{k}$ chosen uniformly at random from $\mathbf{y}$.

Theorem 5.4. *Algorithm 5 works in polynomial time and in expectation is a* $\big(1 - 43\, k^{-1/4} - O(k^{-1/2})\big)$-*approximation to the integral BSP for any position auction.*

The proof of Theorem 5.4 is deferred to Appendix B. Since $\mathrm{SW}(\mathbf{z}) \leq \mathsf{OPT}$, by running the rounding algorithm a few times and taking the best produced solution, we can get a slightly worse approximation guarantee of $(1 - O(k^{-1/4}))\mathsf{OPT}$ with high probability.

6 Numerical Experiments

We describe in this section numerical experiments designed to test our approach for the optimization component of BSP. We deliberately avoid the data collection and interpretation by focusing on synthetically generated prior distributions. The emphasis of this section is on the practical utility of our algorithm. None of the previous PTASes for BSP [6, 17, 22] have been implemented, as all of them (a) are fairly involved algorithms especially one in [22]; (b) have at least a doubly exponential $\Omega(2^{(1/\varepsilon)^{1/\varepsilon}})$ dependency on ε, which is not feasible for practically relevant ε (e.g., $\varepsilon = 0.2$) (c) quite likely have worse performance than much simpler and faster heuristic algorithms (e.g., [17] did extensive numerical experiments of various simple algorithms, but did not implement their EPTAS). In contrast, our algorithms are straightforward to implement, have better running times than most of the simple heuristics (e.g., faster than greedy algorithm for submodular optimization), and our experiments suggest that they have quite good approximations to the optimum[12].

Implemented algorithms. We implemented two versions of our algorithm for BSP. The first one is the original Algorithms 4 and 5 stated in Section 5.1. The second is Algorithms 4 and 5 with the following modifications: we did not fix $x_i = 1$ in the first step of Algorithm 4 for $i \in S_{\mathrm{fix}}$;

[12] For larger instances, it is impossible to find the exact optimum. We used instead local search algorithm as a surrogate substitute for the exact optimum.

consequently, we used a slightly different objective $\widetilde{\mathrm{SW}}(\mathbf{x}) \stackrel{\text{def}}{=\!=} \int_0^\eta \mathrm{H}_{\text{cher}}(\mathbf{x},\tau)\,\mathrm{d}\tau + \int_\eta^{+\infty} \mathrm{G}_{\text{pois}}(\mathbf{x},\tau)\,\mathrm{d}\tau$ than (13) and optimized (14) with respect to $\mathbf{x}$ instead of $\mathbf{x}_M$ using the full budget $\sum_{i\in[n]} x_i \le k$; the rounding step done in Algorithm 5 remained the same. These changes did not affect the runtime of the algorithm much, but allowed us to avoid hard-coded approximation loss of $O(k^{-1/4})$ due to potentially suboptimal decision of fixing a small bidder set S_{fix}.

The original Algorithm 4 is designed with the theoretical worst-case approximation guarantees in mind, while the modified version is intended as a more practical version. It is likely how our algorithm is going to be used in practice. We will refer to the first version as **theoretical algorithm** and to the second one as **practical algorithm**.

Test setup. We implemented our algorithms in MATLAB [16] on position auction with the help of convex optimization toolboxes CVX [10, 11] and Mosek [1]. The data sets were generated as follows. All distributions D_i for $i \in [n]$ have finite support on $\{0, 1, 2, \ldots, 50\}$. Every D_i belongs to one of the three families of distributions and whenever necessary is discretized to the specified support. The description and random parameters of the distribution are given below.

(1) 3 point: D is supported on 3 different values $a, b, c \sim \mathcal{U}[50]$, each value with $\frac{1}{3}$ probability;

(2) normal: $D = \mathcal{N}(\mu, \sigma^2)$, with $\mu, \sigma \in \mathbb{R}$ respectively chosen from $\mathcal{U}[0, 20]$ and $\mathcal{U}[0, 30]$;

(3) shifted student's t-distribution: D is a student's t-distribution with parameter $\nu \in \mathbb{R}$ drawn from $\mathcal{U}[0, 1]$, shifted by $\mu \sim \mathcal{U}[0, 20]$.[13]

The weights $\mathbf{w}$ of the position auction on each (n, k) instance were set as:

$$w_i = 1.0 \quad \text{for} \quad i \in [1, 0.3k], \qquad w_i = 0.2 \quad \text{for} \quad i \in (0.3k, 0.6k], \qquad w_i = 0.0 \quad \text{for} \quad i \in (0.6k, n].$$

We constructed 4 types of instances: all distributions $\{D_i\}_{i\in[n]}$ of the same type (1), (2), or (3), with their parameters generated as described above; mixed instances with all 3 types combined together with some adjustments to the random parameters in order to avoid dominance of one distribution type. We made 3 different tests of various scale for each instance type $(n, k) = (50, 5), (500, 50), (3000, 300)$.

The overall idea in this setup was to avoid too obvious domination of some distributions by others, and in this way produce challenging instances of the BSP. We picked the instance sizes $(n, k) \in \{(50, 5), (500, 50), (3000, 300)\}$ with a constant ratio $n/k = 10$, which is a bit larger than the instances we expect in practice where n/k is likely to be smaller than 2. One reason for that was to make larger variation among the input distributions (note that for a small ratio of n/k, by selecting k agents at random we get in expectation at least k/n approximation to the optimum) and thus get harder problems. We also wanted to get reasonable running time for the brute force algorithm on the smallest instances to test how our benchmark compares to the real optimum. We believe that the real life instances would be noticeably less challenging than the ones above.

6.1 Benchmarks

Ideally, we would like to compare our solutions to the optimum. That is not feasible in general, since BSP is an NP-hard problem even for the case of single-item auction [9]. And as mentioned above, all existing PTAS algorithms [6, 17, 22] are purely theoretical, as they are difficult to implement,

[13] The notations $\mathcal{U}[a, b]$ refers to continuous uniform distribution on the interval $[a, b]$, while $\mathcal{U}[m]$ for $m \in \mathbb{N}$ denotes the uniform distribution on the set $\{0, 1, \ldots, m\}$.

and have bad dependency of the running time on the approximation parameter ε (e.g., the doubly exponential dependency on ε for $\varepsilon = 0.2$ gives $2^{(1/\varepsilon)^{1/\varepsilon}} > 2^{3000}$ many operations).

We consider instead two well-known heuristic algorithms as our benchmarks: **Greedy** algorithm for submodular maximization following numerical experiments in [6, 17], and **Local Search** mentioned in [3]. These algorithms are easy to implement and run in time polynomial[14] in n and k on our datasets. Note that if Local Search starts with the solution produced by Greedy, it can only improve upon it. Thus we use Local Search as the main reference point for measuring quality of approximation. To further justify this choice, we compare Local Search to the exact optimum computed by Brute Force algorithm on small instances ($n = 50, k = 5$) (see Table 1). Greedy, Local Search, and Brute Force algorithms were implemented in C++ and compiled with O3 optimization.

Table 1: Local Search vs. Brute Force (averaged over 100 randomly generated input sets). The table shows the relative quality of the solution produced by Local Search to that of Brute Force is constantly close to 1.0, indicating that Local Search is a good surrogate for the optimal solution.

Setting			Local Search vs. Brute Force
Distribution	n	k	
3 Point	50	5	1.000000
Normal	50	5	1.000000
Student-T	50	5	1.000000
Mixed	50	5	0.999992

6.2 Experimental Results

The numerical experiments are given in Tables 2 and 3. We run 4 algorithms: Local Search, Greedy, and both our theoretical and practical algorithms. There are 4 types of instances, respectively with 3-point, normal, student-T, and mixed distributions, each type having 3 different scale sizes. We look at the approximate efficiency ("solution" column) and the running time of each of the algorithms. We measure efficiency as the relative quality of the produced solution with respect to the solution of the Local Search. For small scale ($n = 50, k = 5$) we generated 100 instances, for medium scale ($n = 500, k = 50$) $-$ 10 instances, and for large scale ($n = 3000, k = 300$) $-$ 1 instance of each type and present in the table averaged results respectively over 100, 10, and 1 generated instances of each type and scale. The reason for the fewer number of instances is because of decreasing variance for larger data sets, and enormous running time of the benchmark on large instances. (Although taking $\leq k$ iterations on our datasets, the complexity of Local Search per iteration is already $O(nk^3 \cdot |\text{Support}|)$, and the complexity of Greedy is also $O(nk^3 \cdot |\text{Support}|)$.)

The performance of Greedy. As shown in Table 2, Greedy performs surprisingly well. On all of our generated datasets, Greedy produces solutions that are within 5% of the solution of Local Search in contrast with the $(1-1/e)$-approximation guarantee for general submodular maximization. Similar observations have been made in [17]. This indicates that for BSP, Greedy probably has a better performance than for general submodular maximization. In fact, we do not know a single instance where greedy would yield a solution worse than 0.9 of the optimum.

[14] In theory, local search may take exponentially many steps to stop, but in all our experiments it was always $\leq k$.

Table 2: Experimental results of Local Search, Greedy, and our practical algorithm. The "solution" column for each algorithm denotes the relative quality of the produced solution to that of Local Search. The table shows as the problem scales, the performance of our algorithm improves while that of Greedy worsens. Moreover, the running time of our algorithm scales much slower than Greedy and Local Search.

Setting			Local Search		Greedy		Our Algorithm	
Distribution	n	k	Solution	Time	Solution	Time	Solution	Time
3 Point	50	5	1.000000	0.01s	0.999477	0.01s	0.998857	2.96s
	500	50	1.000000	7.66s	0.998573	0.51s	0.999914	3.01s
	3000	300	1.000000	28.34h	0.997344	19.49m	0.999983	6.51s
Normal	50	5	1.000000	0.01s	0.989101	0.01s	0.999748	2.86s
	500	50	1.000000	16.86s	0.968220	0.51s	0.999997	3.54s
	3000	300	1.000000	55.18h	0.965290	19.85m	0.999996	10.50s
Student-T	50	5	1.000000	0.01s	0.999112	0.01s	0.997935	2.04s
	500	50	1.000000	7.31s	0.998858	0.52s	0.999958	2.40s
	3000	300	1.000000	22.26h	0.999017	19.69m	0.999996	9.28s
Mixed	50	5	1.000000	0.01s	0.998393	0.01s	0.999006	2.73s
	500	50	1.000000	12.32s	0.994949	0.51s	0.999944	3.73s
	3000	300	1.000000	34.53h	0.994841	21.53m	0.999986	10.32s

The performance of our practical algorithm. In Table 2, our practical algorithm always produces solutions that are within 1% of the solution of Local Search, which demonstrates its effectiveness. As the problem size scales up, our algorithm's performance improves ($\geq 99.99\%$ on $(n, k) = (3000, 300)$), which concurs with our theoretical analysis. On the other hand, the performance of Greedy worsens. Moreover, the running time of our algorithm scales much slower than Greedy and Local Search. For small instances, our algorithm pays an overhead of $2 \sim 3$ seconds due to the convex optimization libraries[15], but for large instances, it is $500 \sim 10^4$ times faster than Greedy and Local Search respectively. Note that all previous PTASes [6, 17, 22] for BSP use Greedy as a subroutine.

Comparing our theoretical and practical algorithm. In Table 3, we compare the performance of our theoretical and practical algorithms. The table shows that the running time of both algorithms are similar, but the practical algorithm performs steadily better than theoretical one, which sometimes produces clearly suboptimal results (e.g., on normal distributions). This is because to achieve a worst-case guarantee, the theoretical algorithm needs to fix a small set of bidders S_{fix}, which may be far from the optimum. In fact, one can easily construct large instances such that theoretical algorithm loses almost all value from S_{fix} and thus has even larger gaps compared to the optimum. Therefore, we remove the fixing step in the practical algorithm and recommend to use the modified algorithm in practice.

7 Concluding Remarks

In this paper we described a theoretically sound and practically applicable Poisson (Poisson-Chernoff) approximation approach to BSP for position auctions under capacity constraint. Our

[15]This overhead could be avoided by using product-level toolboxes.

Table 3: Experimental results of our theoretical and practical algorithm. The "solution" column for each algorithm denotes the relative quality of the produced solution to that of Local Search. The table shows that the running time of both algorithms are similar, but the practical algorithm performs steadily better than the theoretical algorithm.

Setting			Theoretical		Practical	
Distribution	n	k	Solution	Time	Solution	Time
3 Point	50	5	0.988412	1.63s	0.998857	2.96s
	500	50	0.999760	1.51s	0.999914	3.01s
	3000	300	0.999985	5.65s	0.999983	6.51s
Normal	50	5	**0.882275**	1.48s	0.999748	2.86s
	500	50	0.997930	2.32s	0.999997	3.54s
	3000	300	0.999996	9.76s	0.999996	10.50s
Student-T	50	5	0.997081	1.62s	0.997935	2.04s
	500	50	0.999883	4.39s	0.999958	2.40s
	3000	300	0.999996	10.03s	0.999996	9.28s
Mixed	50	5	0.990852	1.99s	0.999006	2.73s
	500	50	0.999561	2.64s	0.999944	3.73s
	3000	300	0.997915	12.31s	0.999986	10.32s

theoretical bounds are quite loose (our goal was to optimize presentation rather than get the best possible convergence rates). First, our approximation constants are not optimized, e.g., it is possible that the constant 43 could be significantly improved, perhaps to a number smaller than 1. Second, the exponent 1/4 in the convergence rate of $O(k^{-1/4})$ is suboptimal. We can improve it to $\widetilde{O}(k^{-1/3})$ but at the cost of significantly more involved analysis. Interestingly, our main Algorithm 4 can be modified by changing the Chernoff approximation part with another Poisson approximation. Indeed, the idea behind Chernoff approximation was to change the sum of Bernoulli random variables with its mean when it is large enough. Poisson approximation concentrates around the same mean and thus would also work well in this case.

A natural next step would be to consider BSP of various auction environments under richer sets of feasibility constraints such as matroid, matching, and intersection of matroids. Another interesting direction is to identify conditions under which it is possible to efficiently optimize the revenue objective of BSP for VCG/GSP auction formats.

References

[1] MOSEK ApS. *The MOSEK optimization toolbox for MATLAB manual. Version 10.0.* MOSEK, 2022.

[2] Nikhil Bansal and Maxim Sviridenko. The santa claus problem. In *Proceedings of the 38th Annual ACM Symposium on Theory of Computing, Seattle, WA, USA, May 21-23, 2006*, STOC '06, page 31–40, New York, NY, USA, 2006. Association for Computing Machinery.

[3] Xiaohui Bei, Nick Gravin, Pinyan Lu, and Zhihao Gavin Tang. Bidder subset selection problem in auction design. In Nikhil Bansal and Viswanath Nagarajan, editors, *Proceedings of the 2023 ACM-SIAM Symposium on Discrete Algorithms, SODA 2023, Florence, Italy, January 22-25, 2023*, pages 3788–3801. SIAM, 2023.

[4] Anand Bhalgat and Sanjeev Khanna. A utility equivalence theorem for concave functions. In Jon Lee and Jens Vygen, editors, *Integer Programming and Combinatorial Optimization*, pages 126–137, Cham, 2014. Springer International Publishing.

[5] Andrei Z. Broder, David Carmel, Michael Herscovici, Aya Soffer, and Jason Zien. Efficient query evaluation using a two-level retrieval process. In *Proceedings of the Twelfth International Conference on Information and Knowledge Management*, CIKM '03, page 426–434, New York, NY, USA, 2003. Association for Computing Machinery.

[6] Wei Chen, Wei Hu, Fu Li, Jian Li, Yu Liu, and Pinyan Lu. Combinatorial multi-armed bandit with general reward functions. In Daniel D. Lee, Masashi Sugiyama, Ulrike von Luxburg, Isabelle Guyon, and Roman Garnett, editors, *Advances in Neural Information Processing Systems*, volume 29, pages 1651–1659, Red Hook, NY, USA, 2016. Curran Associates Inc.

[7] D. J. Daley and D. Vere-Jones. *An introduction to the theory of point processes. Vol. II: General Theory and Structure*. Probability and its Applications (New York). Springer, New York, second edition, 2008.

[8] Benjamin Edelman, Michael Ostrovsky, and Michael Schwarz. Internet advertising and the generalized second-price auction: Selling billions of dollars worth of keywords. *American Economic Review*, 97(1):242–259, March 2007.

[9] Ashish Goel, Sudipto Guha, and Kamesh Munagala. How to probe for an extreme value. *ACM Trans. Algorithms*, 7(1):12:1–12:20, 2010.

[10] Michael Grant and Stephen Boyd. CVX: Matlab software for disciplined convex programming, version 2.1. http://cvxr.com/cvx, March 2014.

[11] Michael C. Grant and Stephen P. Boyd. Graph implementations for nonsmooth convex programs. In Vincent D. Blondel, Stephen P. Boyd, and Hidenori Kimura, editors, *Recent Advances in Learning and Control*, pages 95–110, London, 2008. Springer London.

[12] Jason Hartline. Mechanism design and approximation. http://jasonhartline.com/MDnA/. Accessed: 2022-01-10.

[13] Jon Kleinberg and Maithra Raghu. Team performance with test scores. *ACM Trans. Econ. Comput.*, 6(3–4), oct 2018.

[14] Jian Li and Amol Deshpande. Maximizing Expected Utility for Stochastic Combinatorial Optimization Problems. *Mathematics of Operations Research*, 44(1):354–375, February 2019.

[15] Jian Li and Wen Yuan. Stochastic combinatorial optimization via poisson approximation. In *Proceedings of the Forty-Fifth Annual ACM Symposium on Theory of Computing*, STOC '13, page 971–980, New York, NY, USA, 2013. Association for Computing Machinery.

[16] MATLAB. *9.12.0.1884302 (R2022a)*. The MathWorks Inc., Natick, Massachusetts, 2022.

[17] Aranyak Mehta, Uri Nadav, Alexandros Psomas, and Aviad Rubinstein. Hitting the high notes: Subset selection for maximizing expected order statistics. *Advances in Neural Information Processing Systems*, 33:15800–15810, 2020.

[18] S. Y. Novak. Poisson approximation. *Probability Surveys*, 16:228 – 276, 2019.

[19] S.Y. Novak. *Extreme Value Methods with Applications to Finance.* Chapman & Hall/CRC Monographs on Statistics & Applied Probability. CRC Press, Boca Raton, 2011.

[20] Ariel D. Procaccia, Sashank Jakkam Reddi, and Nisarg Shah. A maximum likelihood approach for selecting sets of alternatives. In Nando de Freitas and Kevin P. Murphy, editors, *Proceedings of the Twenty-Eighth Conference on Uncertainty in Artificial Intelligence, Catalina Island, CA, USA, August 14-18, 2012*, pages 695–704, Corvallis, Oregon, 2012. AUAI Press.

[21] Zhou Rong and Zhixue Lu. Two-stage procurement auction with bidders of asymmetric capacity. *Systems Engineering-Theory & Practice*, 27(12):36–41, 2007.

[22] Danny Segev and Sahil Singla. Efficient approximation schemes for stochastic probing and prophet problems. In Péter Biró, Shuchi Chawla, and Federico Echenique, editors, *EC '21: The 22nd ACM Conference on Economics and Computation, Budapest, Hungary, July 18-23, 2021*, pages 793–794, New York, NY, USA, 2021. ACM.

[23] Guofu Tan. Entry and r & d in procurement contracting. *Journal of Economic Theory*, 58(1):41–60, 1992.

[24] Hal Varian. Position auctions. *International Journal of Industrial Organization*, 25(6):1163–1178, 2007.

[25] Hal R. Varian and Christopher Harris. The vcg auction in theory and practice. *American Economic Review*, 104(5):442–45, May 2014.

[26] Jiajin Yu and Shabbir Ahmed. Maximizing expected utility over a knapsack constraint. *Operations Research Letters*, 44(2):180–185, 2016.

A Missing Proofs in Section 4

A.1 Proof of Lemma 4.2

We recall the following definitions

$$\mathrm{H_{ber}}(\mathbf{q},\ell) = \mathop{\mathbf{E}}_{\mathbf{z}\sim\mathrm{Ber}(\mathbf{q})}\left[\min\left(\sum_{i=1}^{n} z_i,\, \ell\right)\right], \qquad \mathrm{H_{cher}}(\mathbf{q},\ell) = \min\left(\sum_{i=1}^{n} q_i,\, \ell\right).$$

In order to prove Lemma 4.2, we first prove the following two auxiliary lemmas.

Lemma A.1. *For all $\mathbf{q} \in [0,1]^n$, $\ell \in \mathbb{N}^+$, let $\lambda = \sum_{i=1}^n q_i$, the following properties hold.*

(a) $\mathrm{H_{ber}}(\mathbf{q},\ell) \geq \lambda(1 - \frac{1}{2}\lambda)$.

(b) *If* $\lambda \geq 1$, $\quad \mathrm{H_{ber}}(\mathbf{q},\ell) \geq \frac{1}{2}$.

(c) *If* $\ell \geq \lambda$, *then* $\quad \mathrm{H_{cher}}(\mathbf{q},\ell) - \mathrm{H_{ber}}(\mathbf{q},\ell) \leq \sum_{i=\ell+1}^{n} e^{-\frac{\delta^2(i)}{2+\delta(i)}\cdot\lambda}$, $\quad$ *where* $\delta(i) = \frac{i-\lambda}{\lambda}$.

(d) *If* $\ell \geq \lambda$ *and* $\alpha \in (0,1)$, *then* $\quad \mathrm{H_{cher}}(\mathbf{q},\ell) - \mathrm{H_{ber}}(\mathbf{q},\ell) \leq \alpha\lambda + \frac{6}{\alpha}e^{-\alpha^2\cdot\frac{\lambda}{3}}$.

(e) *If* $\lambda \geq \ell$, *then* $\quad \mathrm{H_{cher}}(\mathbf{q},\ell) - \mathrm{H_{ber}}(\mathbf{q},\ell) \leq \sum_{i=0}^{\ell-1} e^{-\frac{(\lambda-i)^2}{2\lambda}}$.

(f) *If $\lambda \geq \ell$ and $\alpha \in (0,1)$, then* $\mathrm{H}_{\text{cher}}(\mathbf{q},\ell) - \mathrm{H}_{\text{ber}}(\mathbf{q},\ell) \leq \alpha\lambda + \frac{4}{\alpha}e^{-\alpha^2\cdot\frac{\lambda}{2}}$.

Proof of Lemma A.1. (a). Since $\ell \geq 1$, We have the following lower bound on $\mathrm{H}_{\text{ber}}(\mathbf{q},\ell)$

$$\mathrm{H}_{\text{ber}}(\mathbf{q},\ell) \geq \Pr_{\substack{\mathbf{z}\sim\mathrm{Ber}(\mathbf{q})\\ Z=\sum_{i=1}^n z_i}}[Z \geq 1] = \sum_{i=1}^n \Pr_{z_i\sim\mathrm{Ber}(q_i)}[z_i = 1]\cdot \Pr_{\mathbf{z}\sim\mathrm{Ber}(\mathbf{q})}[\forall j < i,\ z_j = 0]$$

$$\geq \sum_{i=1}^n q_i \cdot \left(1 - \sum_{j=1}^{i-1} q_j\right) = \sum_{i=1}^n q_i - \sum_{j<i} q_i q_j \geq \lambda\Big(1 - \frac{1}{2}\lambda\Big).$$

(b). If $\lambda \geq 1$, there exists $\mathbf{q}' \in [0,1]^n$ such that $q_i' \leq q_i, \forall i \in \{1,2,\ldots,n\}$ and $\sum_{i=1}^n q_i' = 1$. Using Lemma A.1 (a), we can see that $\mathrm{H}_{\text{ber}}(\mathbf{q}',\ell) \geq 0.5$. Then $\mathrm{H}_{\text{ber}}(\mathbf{q},\ell) \geq \mathrm{H}_{\text{ber}}(\mathbf{q}',\ell) \geq 0.5$.

(c). As $\ell \geq \lambda$, we have $\mathrm{H}_{\text{cher}}(\mathbf{q},\ell) = \lambda$. Then

$$\mathrm{H}_{\text{cher}}(\mathbf{q},\ell) - \mathrm{H}_{\text{ber}}(\mathbf{q},\ell) = \mathop{\mathbf{E}}_{\substack{\mathbf{z}\sim\mathrm{Ber}(\mathbf{q})\\ Z=\sum_{i=1}^n z_i}}[Z - \min\{Z,\ell\}] = \sum_{i=\ell+1}^n \Pr_{\substack{\mathbf{z}\sim\mathrm{Ber}(\mathbf{q})\\ Z=\sum_{i=1}^n z_i}}[Z \geq i].$$

We apply Chernoff bound to each tail probability $Z \geq i$ under the sum. Thus

$$\mathrm{H}_{\text{cher}}(\mathbf{q},\ell) - \mathrm{H}_{\text{ber}}(\mathbf{q},\ell) \leq \sum_{i=\ell+1}^n e^{-\frac{\delta^2(i)}{2+\delta(i)}\cdot\lambda}, \quad \text{where } \delta(i) = \frac{i-\lambda}{\lambda}$$

(d). By Lemma A.1 (c)

$$\mathrm{H}_{\text{cher}}(\mathbf{q},\ell) - \mathrm{H}_{\text{ber}}(\mathbf{q},\ell) \leq \sum_{i=\lceil\lambda+1\rceil}^{\infty} e^{-\frac{\delta^2(i)}{2+\delta(i)}\cdot\lambda} \leq \alpha\lambda + \sum_{i=\lceil(1+\alpha)\lambda\rceil}^{\infty} e^{-\frac{\delta^2(i)}{2+\delta(i)}\cdot\lambda}.$$

Note that inside the last summation, $\delta(i) \geq \alpha$ and $\frac{\delta(i)}{2+\delta(i)} \geq \frac{\alpha}{2+\alpha} \geq \frac{\alpha}{3}$. Therefore,

$$\mathrm{H}_{\text{cher}}(\mathbf{q},\ell) - \mathrm{H}_{\text{ber}}(\mathbf{q},\ell) \leq \alpha\lambda + \sum_{i=\lceil(1+\alpha)\lambda\rceil}^{\infty} e^{-\frac{\alpha}{2+\alpha}\cdot\delta(i)\lambda} \leq \alpha\lambda + \sum_{i=\lceil(1+\alpha)\lambda\rceil}^{\infty} e^{-(i-\lambda)\cdot\frac{\alpha}{3}}.$$

As the summation in the right hand side becomes a geometric series, we get

$$\mathrm{H}_{\text{cher}}(\mathbf{q},\ell) - \mathrm{H}_{\text{ber}}(\mathbf{q},\ell) \leq \alpha\lambda + e^{-\alpha^2\cdot\frac{\lambda}{3}}\cdot(1 - e^{-\frac{\alpha}{3}})^{-1} \leq \alpha\lambda + \frac{6}{\alpha}e^{-\alpha^2\cdot\frac{\lambda}{3}}.$$

The last inequality holds as $1 - e^{-x} \geq x/2$ for $x \in [0,1]$. Hence, Lemma A.1 (d) holds.

(e). If $\ell \leq \lambda$, then $\mathrm{H}_{\text{cher}}(\mathbf{q},\ell) = \ell$. We have

$$\mathrm{H}_{\text{cher}}(\mathbf{q},\ell) - \mathrm{H}_{\text{ber}}(\mathbf{q},\ell) = \mathop{\mathbf{E}}_{\substack{\mathbf{z}\sim\mathrm{Ber}(\mathbf{q})\\ Z=\sum_{i=1}^n z_i}}[\ell - \min\{Z,\ell\}] = \sum_{i=0}^{\ell-1} \Pr_{\substack{\mathbf{z}\sim\mathrm{Ber}(\mathbf{q})\\ Z=\sum_{i=1}^n z_i}}[Z \leq i].$$

We again apply Chernoff bound for each tail event $Z \leq i$ under summation and get

$$\mathrm{H}_{\text{cher}}(\mathbf{q},\ell) - \mathrm{H}_{\text{ber}}(\mathbf{q},\ell) \leq \sum_{i=0}^{\ell-1} e^{-\frac{(\lambda-i)^2}{2\lambda}}.$$

(f). By Lemma A.1 (e), we have

$$\mathrm{H}_{\mathrm{cher}}(\mathbf{q},\ell)-\mathrm{H}_{\mathrm{ber}}(\mathbf{q},\ell)\le\sum_{i=0}^{\ell-1}e^{-\frac{(\lambda-i)^2}{2\lambda}}\le\sum_{i=1}^{\infty}e^{-\frac{i^2}{2\lambda}}\le\alpha\lambda+\sum_{i=\lceil\alpha\lambda\rceil}^{\infty}e^{-\frac{i^2}{2\lambda}}\le\alpha\lambda+\sum_{i=\lceil\alpha\lambda\rceil}^{\infty}e^{-i\cdot\frac{\alpha}{2}}.$$

The summation in the right hand side is again a geometric series, which allows us to get

$$\mathrm{H}_{\mathrm{cher}}(\mathbf{q},\ell)-\mathrm{H}_{\mathrm{ber}}(\mathbf{q},\ell)\le\alpha\lambda+e^{-\alpha^2\cdot\frac{\lambda}{2}}\cdot(1-e^{-\frac{\alpha}{2}})^{-1}\le\alpha\lambda+\frac{4}{\alpha}e^{-\alpha^2\cdot\frac{\lambda}{2}}.$$

Hence Lemma A.1 (f) holds. $\square$

Claim A.2. *For any* $\alpha\ge 3.4$, $\sum_{i=1}^{\infty}e^{-\frac{i^2}{2\alpha+i}}\le 0.85\alpha$.

Proof. When $\alpha,x>0$, the function $-\frac{x^2}{2\alpha+x}$ is decreasing in x. Therefore,

$$\sum_{x=1}^{\infty}e^{-\frac{x^2}{2\alpha+x}}\le\int_{y=0}^{+\infty}e^{-\frac{y^2}{2\alpha+y}}\mathrm{d}y\overset{[y=\alpha\cdot z]}{=}\alpha\cdot\int_{z=0}^{+\infty}e^{-\frac{\alpha z^2}{2+z}}\mathrm{d}z\le\alpha\cdot\int_{z=0}^{+\infty}e^{-\frac{3.4z^2}{2+z}}\mathrm{d}z\le\alpha\cdot 0.85.\qquad\square$$

Now we are ready to prove Lemma 4.2.

Lemma 4.2. *For all* $\mathbf{q}\in[0,1]^n,\ell\in\mathbb{N}^+$, *let* $\lambda=\sum_{i=1}^n q_i$, *the following properties hold.*

(a) $$\mathrm{H}_{\mathrm{ber}}(\mathbf{q},\ell)\le\mathrm{H}_{\mathrm{cher}}(\mathbf{q},\ell)\le 7\cdot\mathrm{H}_{\mathrm{ber}}(\mathbf{q},\ell).$$

(b) $$\mathrm{H}_{\mathrm{cher}}(\mathbf{q},\ell)-\mathrm{H}_{\mathrm{ber}}(\mathbf{q},\ell)\le\frac{3}{\sqrt{\lambda}}\cdot\mathrm{H}_{\mathrm{cher}}(\mathbf{q},\ell)\le\frac{21}{\sqrt{\lambda}}\cdot\mathrm{H}_{\mathrm{ber}}(\mathbf{q},\ell).$$

(c) $$\mathrm{H}_{\mathrm{cher}}(\mathbf{q},\ell)-\mathrm{H}_{\mathrm{ber}}(\mathbf{q},\ell)\le\frac{5}{\sqrt{\ell}}\cdot\mathrm{H}_{\mathrm{cher}}(\mathbf{q},\ell)\le\frac{35}{\sqrt{\ell}}\cdot\mathrm{H}_{\mathrm{ber}}(\mathbf{q},\ell).$$

Proof of Lemma 4.2. (a). As $\min\{t,\ell\}$ is a concave function in t, the Jensen inequality gives us

$$\mathrm{H}_{\mathrm{cher}}(\mathbf{q},\ell)=\min\left\{\mathop{\mathbf{E}}_{\substack{\mathbf{z}\sim\mathrm{Ber}(\mathbf{q})\\ Z=\sum_{i=1}^n z_i}}[Z],\ \ell\right\}\ge\mathop{\mathbf{E}}_{\substack{\mathbf{z}\sim\mathrm{Ber}(\mathbf{q})\\ Z=\sum_{i=1}^n z_i}}[\min\{Z,\ \ell\}]=\mathrm{H}_{\mathrm{ber}}(\mathbf{q},\ell),$$

which gives the first inequality. To prove the second inequality we consider the following 4 cases:

- If $\lambda\le 1$, we use Lemma A.1 (a) to get $\mathrm{H}_{\mathrm{ber}}(\mathbf{q},\ell)\ge\lambda\big(1-\frac{1}{2}\lambda\big)\ge\frac{1}{2}\lambda\ge\frac{1}{2}\cdot\mathrm{H}_{\mathrm{cher}}(\mathbf{q},\ell)$.
- If $\lambda>1$ and $\min(\lambda,\ell)\le 3.5$, then Lemma A.1 (b) gives us $\mathrm{H}_{\mathrm{ber}}(\mathbf{q},\ell)\ge\frac{1}{2}\ge\frac{1}{7}\cdot\min(\lambda,\ell)=\frac{1}{7}\cdot\mathrm{H}_{\mathrm{cher}}(\mathbf{q},\ell)$.
- If $\min(\lambda,\ell)>3.5$ and $\ell\ge\lambda$, then Lemma A.1 (c) gives us

  $$\mathrm{H}_{\mathrm{cher}}(\mathbf{q},\ell)-\mathrm{H}_{\mathrm{ber}}(\mathbf{q},\ell)\le\sum_{i=\ell+1}^{n}e^{-\frac{\delta^2(i)}{2+\delta(i)}\cdot\lambda}\le\sum_{i=\lceil\lambda+1\rceil}^{\infty}e^{-\frac{(i-\lambda)^2}{2\lambda+(i-\lambda)}}\le\sum_{i=1}^{\infty}e^{-\frac{i^2}{2\lambda+i}}.$$

  We apply Claim A.2 for $\alpha=\lambda$ and get $\mathrm{H}_{\mathrm{cher}}(\mathbf{q},\ell)-\mathrm{H}_{\mathrm{ber}}(\mathbf{q},\ell)\le 0.85\lambda$, which implies that $\mathrm{H}_{\mathrm{ber}}(\mathbf{q},\ell)\ge 0.15\cdot\mathrm{H}_{\mathrm{cher}}(\mathbf{q},\ell)$ for $\mathrm{H}_{\mathrm{cher}}(\mathbf{q},\ell)=\lambda\le\ell$.

- If $\min(\lambda,\ell) > 3.5$ and $\lambda \geq \ell$, then Lemma A.1 (e) gives us $\mathrm{H}_{\text{cher}}(\mathbf{q},\ell) - \mathrm{H}_{\text{ber}}(\mathbf{q},\ell) \leq \sum_{i=0}^{\ell-1} e^{-\frac{1}{2\lambda}(\lambda-i)^2}$. Note that $\frac{1}{2\lambda}(\lambda-i)^2 \geq \frac{1}{2\ell}(\ell-i)^2$, when $\lambda \geq \ell \geq i$. Thus

$$\mathrm{H}_{\text{cher}}(\mathbf{q},\ell) - \mathrm{H}_{\text{ber}}(\mathbf{q},\ell) \leq \sum_{i=0}^{\ell-1} e^{-\frac{1}{2\ell}(\ell-i)^2} \leq \sum_{i=1}^{\infty} e^{-\frac{i^2}{2\ell}} \leq \sum_{i=1}^{\infty} e^{-\frac{i^2}{2\ell+i}}.$$

  By applying Claim A.2 with $\alpha = \ell$, we get $\mathrm{H}_{\text{cher}}(\mathbf{q},\ell) - \mathrm{H}_{\text{ber}}(\mathbf{q},\ell) \leq 0.85\ell$, which implies that $\mathrm{H}_{\text{ber}}(\mathbf{q},\ell) \geq 0.15 \cdot \mathrm{H}_{\text{cher}}(\mathbf{q},\ell)$, as $\mathrm{H}_{\text{cher}}(\mathbf{q},\ell) = \ell \leq \lambda$.

In all 4 cases, $7\mathrm{H}_{\text{ber}}(\mathbf{q},\ell) \geq \mathrm{H}_{\text{cher}}(\mathbf{q},\ell)$.

(b). Let $\varepsilon \stackrel{\text{def}}{=\!=} \frac{1}{\sqrt{\lambda}}$. We consider three cases to show that $\mathrm{H}_{\text{cher}}(\mathbf{q},\ell) - \mathrm{H}_{\text{ber}}(\mathbf{q},\ell) \leq 3\varepsilon \cdot \mathrm{H}_{\text{cher}}(\mathbf{q},\ell)$, then combine it with the inequality $7\mathrm{H}_{\text{ber}}(\mathbf{q},\ell) \geq \mathrm{H}_{\text{cher}}(\mathbf{q},\ell)$ from Lemma 4.2 (a) to conclude the proof. Without loss of generality, we may assume that $\varepsilon < 1/3$.

- If $\ell \geq \lambda$, then Lemma A.1 (d) for $\alpha = 1.8\varepsilon$, gives us $\mathrm{H}_{\text{cher}}(\mathbf{q},\ell) - \mathrm{H}_{\text{ber}}(\mathbf{q},\ell) \leq 2.94\varepsilon\lambda = 2.94\varepsilon \cdot \mathrm{H}_{\text{cher}}(\mathbf{q},\ell)$.
- If $\ell \leq \frac{3}{4}\lambda$, then by Lemma A.1 (e) we have

$$\mathrm{H}_{\text{cher}}(\mathbf{q},\ell) - \mathrm{H}_{\text{ber}}(\mathbf{q},\ell) \leq \sum_{i=0}^{\ell-1} e^{-\frac{(\lambda-i)^2}{2\lambda}} \leq \ell \cdot e^{-\frac{\lambda}{32}} \leq \frac{4e^{-1/2}}{\sqrt{\lambda}} \cdot \mathrm{H}_{\text{cher}}(\mathbf{q},\ell) \leq 2.5 \cdot \varepsilon \cdot \mathrm{H}_{\text{cher}}(\mathbf{q},\ell),$$

  where the first inequality holds because $\lambda - i \geq \lambda/4$; the second inequality holds, as $\mathrm{H}_{\text{cher}}(\mathbf{q},\ell) = \ell$ and $e^{-x/c}\sqrt{x} \leq e^{-1/2}\sqrt{c/2}$ for all $x, c > 0$.
- If $\ell \in \left(\frac{3}{4}\lambda, \lambda\right)$, by Lemma A.1 (f) for $\alpha = 1.8\varepsilon$ we have

$$\mathrm{H}_{\text{cher}}(\mathbf{q},\ell) - \mathrm{H}_{\text{ber}}(\mathbf{q},\ell) \leq 2.24\varepsilon\lambda \leq 2.99\varepsilon\ell = 2.99\varepsilon \cdot \mathrm{H}_{\text{cher}}(\mathbf{q},\ell).$$

(c). Let $\varepsilon \stackrel{\text{def}}{=\!=} \frac{1}{\sqrt{\ell}}$. We first prove $\mathrm{H}_{\text{cher}}(\mathbf{q},\ell) - \mathrm{H}_{\text{ber}}(\mathbf{q},\ell) \leq 5\varepsilon \cdot \mathrm{H}_{\text{cher}}(\mathbf{q},\ell)$ by considering the following four cases. It implies the second bound when combined with inequality $7\mathrm{H}_{\text{ber}}(\mathbf{q},\ell) \geq \mathrm{H}_{\text{cher}}(\mathbf{q},\ell)$ from Lemma 4.2 (a). We may assume without loss of generality that $\varepsilon < 1/5$.

- If $\lambda < \frac{1}{\sqrt{\ell}}$, then $\mathrm{H}_{\text{cher}}(\mathbf{q},\ell) = \lambda$ and Lemma A.1 (a) gives us

$$\mathrm{H}_{\text{cher}}(\mathbf{q},\ell) - \mathrm{H}_{\text{ber}}(\mathbf{q},\ell) \leq \lambda - \lambda\left(1 - \frac{1}{2}\lambda\right) = \frac{\lambda^2}{2} < \frac{\varepsilon\lambda}{2} = \frac{\varepsilon}{2} \cdot \mathrm{H}_{\text{cher}}(\mathbf{q},\ell).$$

- If $\lambda \in [\frac{1}{\sqrt{\ell}}, \frac{\ell}{3})$, then $\mathrm{H}_{\text{cher}}(\mathbf{q},\ell) = \lambda$ and by Lemma A.1 (c) we have

$$\mathrm{H}_{\text{cher}}(\mathbf{q},\ell) - \mathrm{H}_{\text{ber}}(\mathbf{q},\ell) \leq \sum_{i=\ell+1}^{n} e^{-\frac{\delta^2(i)}{2+\delta(i)}\cdot\lambda} \leq \sum_{i=\ell+1}^{n} e^{-\delta(i)\cdot\frac{\lambda}{2}} \leq \sum_{i=\ell+1}^{+\infty} e^{-\frac{i-\lambda}{2}},$$

  where the second inequality holds, as $\delta(i) = \frac{i-\lambda}{\lambda} \geq 2$ for all $i \geq \ell+1$ and $\lambda < \ell/3$. Furthermore,

$$\mathrm{H}_{\text{cher}}(\mathbf{q},\ell) - \mathrm{H}_{\text{ber}}(\mathbf{q},\ell) \leq e^{-\frac{\ell}{3}} \cdot (1 - e^{-\frac{1}{2}})^{-1} \leq \frac{3e^{-1}/\ell}{1 - e^{-1/2}} \leq 2.81 \cdot \varepsilon \cdot \mathrm{H}_{\text{cher}}(\mathbf{q},\ell),$$

  where to get the first inequality we simply use the formula for the sum of geometric progression and estimate $e^{-\frac{\ell+1-\lambda}{2}} \leq e^{-\ell/3}$; the second inequality holds, because $e^{-x/c} \leq c \cdot e^{-1}/x$ for any $x, c > 0$; the last inequality holds, because $\mathrm{H}_{\text{cher}}(\mathbf{q},\ell) = \lambda \geq \varepsilon = 1/\sqrt{\ell}$.

- If $\lambda \in [\frac{\ell}{3}, \ell]$, then $\mathrm{H}_{\mathrm{cher}}(\mathbf{q}, \ell) = \lambda$ and by Lemma A.1 (d) for $\alpha = 4\varepsilon$ ($\alpha < 1$, since $\varepsilon \leq 1/5$) we get $\mathrm{H}_{\mathrm{cher}}(\mathbf{q}, \ell) - \mathrm{H}_{\mathrm{ber}}(\mathbf{q}, \ell) \leq 4\varepsilon\lambda + \frac{6}{4\varepsilon} e^{-16\varepsilon^2 \cdot \frac{\lambda}{3}} \leq 4.77\varepsilon\lambda = 4.77\varepsilon \cdot \mathrm{H}_{\mathrm{cher}}(\mathbf{q}, \ell)$, where the second inequality holds, since $\ell \leq 3\lambda$ and $\frac{6}{4\varepsilon} e^{-16\varepsilon^2 \cdot \frac{\lambda}{3}} = \frac{3\varepsilon\ell}{2} e^{-16 \cdot \frac{\lambda}{3\ell}} \leq \frac{9\varepsilon\lambda}{2} e^{-\frac{16\ell}{9\ell}} \leq 0.77 \cdot \varepsilon\lambda$.
- If $\lambda > \ell$, then by considering $\mathbf{q}'$ with $q'_i \leq q_i$ and $\sum_{i=1}^n q'_i = \ell$, we get $\mathrm{H}_{\mathrm{ber}}(\mathbf{q}) \geq \mathrm{H}_{\mathrm{ber}}(\mathbf{q}')$ and $\mathrm{H}_{\mathrm{cher}}(\mathbf{q}) = \mathrm{H}_{\mathrm{cher}}(\mathbf{q}') = \ell$. Then according to the previous case, $\mathrm{H}_{\mathrm{cher}}(\mathbf{q}, \ell) - \mathrm{H}_{\mathrm{ber}}(\mathbf{q}, \ell) \leq \mathrm{H}_{\mathrm{cher}}(\mathbf{q}', \ell) - \mathrm{H}_{\mathrm{ber}}(\mathbf{q}', \ell) \leq 4.77\varepsilon \cdot \mathrm{H}_{\mathrm{cher}}(\mathbf{q}', \ell) = 4.77\varepsilon \cdot \mathrm{H}_{\mathrm{cher}}(\mathbf{q}, \ell)$.

These bounds combined with inequality $7\mathrm{H}_{\mathrm{ber}}(\mathbf{q}, \ell) \geq \mathrm{H}_{\mathrm{cher}}(\mathbf{q}, \ell)$ conclude the proof of (c). ☐

A.2 Proof of Lemma 4.7

Lemma 4.7. *Suppose* $\mathbf{q} \in [0, \delta]^n$ *and* $\ell \in \mathbb{N}$. *Then*

$$
\begin{array}{llr}
\text{(a)} & |\mathrm{H}_{\mathrm{ber}}(\mathbf{q}, \ell) - \mathrm{H}_{\mathrm{pois}}(\mathbf{q}, \ell)| \leq 17.5 \cdot \delta \cdot \mathrm{H}_{\mathrm{ber}}(\mathbf{q}, \ell) & (\forall \delta,\ \ell), \\
\text{(b)} & 0 \leq \mathrm{H}_{\mathrm{ber}}(\mathbf{q}, 1) - \mathrm{H}_{\mathrm{pois}}(\mathbf{q}, 1) \leq \delta \cdot \mathrm{H}_{\mathrm{ber}}(\mathbf{q}, 1) & (\ell = 1).
\end{array}
$$

Proof of Lemma 4.7. (a). Let $z_i \sim \mathrm{Ber}(q_i)$ be Bernoulli random variables and $Z = \sum_{i=1}^n z_i$ be their sum. Then we can rewrite $\mathrm{H}_{\mathrm{ber}}(\mathbf{q}, \ell) = \mathbf{E}[\min(Z, \ell)] = \sum_{j=1}^{\ell} \mathbf{Pr}[Z \geq j]$. Also let $y_i \sim \mathrm{Pois}(q_i)$ and define $Y = \sum_{i=1}^n y_i$. Clearly, $Y \sim \mathrm{Pois}(\lambda)$ for $\lambda = \sum_{i=1}^n q_i$. Then $\mathrm{H}_{\mathrm{pois}}(\mathbf{q}, \ell) = \mathbf{E}[\min(Y, \ell)] = \sum_{j=1}^{\ell} \mathbf{Pr}[Y \geq j]$. Hence, we have

$$
|\mathrm{H}_{\mathrm{ber}}(\mathbf{q}, \ell) - \mathrm{H}_{\mathrm{pois}}(\mathbf{q}, \ell)| \leq \sum_{j=1}^{\ell} |\mathbf{Pr}[Z \geq j] - \mathbf{Pr}[Y \geq j]|.
$$

The RHS is similar to the earth mover's distance between the sum of Bernoulli and Poisson random variables (the difference is that the summation instead of $+\infty$ goes only up to ℓ). Next, we shall prove the following inequality:

$$
\sum_{j=1}^{\ell} |\mathbf{Pr}[Z \geq j] - \mathbf{Pr}[Y \geq j]| \leq 2.5\,\delta \cdot \min(\lambda, \ell),
$$

which together with Lemma 4.2 (a) immediately implies the desired result:

$$
|\mathrm{H}_{\mathrm{ber}}(\mathbf{q}, \ell) - \mathrm{H}_{\mathrm{pois}}(\mathbf{q}, \ell)| \leq 2.5\,\delta \cdot \min(\lambda, \ell) \leq 2.5\,\delta \cdot 7\,\mathrm{H}_{\mathrm{ber}}(\mathbf{q}, \ell) = 17.5\,\delta \cdot \mathrm{H}_{\mathrm{ber}}(\mathbf{q}, \ell).
$$

We consider two cases. First, when $\lambda \geq \ell$. Then for any $j \in \mathbb{N}^+$ we have by (7)

$$
\sum_{j=1}^{\ell} |\mathbf{Pr}[Z \geq j] - \mathbf{Pr}[Y \geq j]| \leq \ell \cdot \sum_{j'=0}^{\infty} \left|\mathbf{Pr}[Z = j'] - \mathbf{Pr}[Y = j']\right| \leq \ell \cdot \delta = \delta \cdot \min(\lambda, \ell).
$$

Second, when $\min(\lambda, \ell) = \lambda < \ell$, we use instead the earth mover's distance $d_{\mathrm{G}}(Z, Y) = \sum_{j=0}^{+\infty} |\mathbf{Pr}[Z \geq j] - \mathbf{Pr}[Y \geq j]|$. We do not calculate the cumulative density functions of $Z = \sum_i^n z_i$ and $Y = \sum_i^n y_i$, but get an upper bound by coupling individual z_i and y_i. Specifically, we couple z_i and y_i so that $z_i = 0$ implies $y_i = 0$ (note that $\mathbf{Pr}[z_i = 0] = 1 - q_i \leq \mathbf{Pr}[y_i = 0] = e^{-q_i}$). Conversely, if $z_i = 1$ it is matched with all $y_i = 1, 2, \ldots$ and the remaining probability for $y_i = 0$. Then we have

$$
d_{\mathrm{G}}(Z, Y) \leq \mathop{\mathbf{E}}_{(z_i, y_i)_{i=1}^n} [|Z - Y|] \leq \sum_{i=1}^n \mathop{\mathbf{E}}_{(z_i, y_i)} [|z_i - y_i|].
$$

Since $z_i = 0$ is matched to $y_i = 0$, we get the following expression for the term $\mathbf{E}[|z_i - y_i|]$,

$$\mathop{\mathbf{E}}_{(z_i,y_i)}[|z_i - y_i|] = \sum_{j=0}^{+\infty} \mathop{\mathbf{Pr}}_{(z_i,y_i)}[z_i = 1 \wedge y_i = j] \cdot |j-1| = \mathop{\mathbf{Pr}}_{(z_i,y_i)}[z_i = 1 \wedge y_i = 0]$$
$$+ \sum_{j=2}^{+\infty} \mathop{\mathbf{Pr}}_{y_i}[y_i = j] \cdot (j-1) \le \frac{q_i^2}{2} + q_i^2 \sum_{j=2}^{+\infty} 2^{-(j-1)}(j-1) \le 2.5\, q_i^2,$$

where the first inequality holds, since $\mathbf{Pr}[z_i = 1 \wedge y_i = 0] = (\mathbf{Pr}[y_i = 0] - \mathbf{Pr}[z_i = 0]) = e^{-q_i} - 1 + q_i \le q_i^2/2$ and $\mathbf{Pr}[y_i = j] = e^{-q_i} q_i^j / j! \le q_i^2 / 2^{j-1}$ for $j \ge 2$. Therefore,

$$\sum_{j=1}^{\ell} |\mathbf{Pr}[Z \ge j] - \mathbf{Pr}[Y \ge j]| \le d_{\mathrm{G}}(Z, Y) \le \sum_{i=1}^{n} 2.5\, q_i^2 \le 2.5 \cdot \delta \cdot \sum_{i=1}^{n} q_i = 2.5 \cdot \delta \cdot \min(\lambda,\, \ell),$$

which concludes the proof.

(b). As $\ell = 1$, $\mathrm{H}_{\mathrm{ber}}(\mathbf{q}, 1) = 1 - \prod_{i=1}^{n}(1 - q_i) \overset{\text{def}}{=\!=} 1 - e^{-s}$, where $s = -\sum_{i=1}^{n} \ln(1 - q_i)$. Let $\lambda = \sum_{i=1}^{n} q_i$, then $\mathrm{H}_{\mathrm{pois}}(\mathbf{q}, 1) = 1 - e^{-\lambda}$. Observe that $q_i \le -\ln(1 - q_i) \le \frac{q_i}{1 - q_i} \le \frac{q_i}{1 - \delta}$. Thus $\lambda \le s \le \frac{\lambda}{1 - \delta}$. The former inequality implies the desired lower bound $\mathrm{H}_{\mathrm{ber}}(\mathbf{q}, 1) - \mathrm{H}_{\mathrm{pois}}(\mathbf{q}, 1) \ge 0$.

To prove the required upper bound, observe that $\mathrm{H}_{\mathrm{ber}}(\mathbf{q}, 1) = f(s)$, $\mathrm{H}_{\mathrm{pois}}(\mathbf{q}, 1) = f(\lambda)$ for $f(t) = 1 - e^{-t}$. As $f(t)$ is a concave function with $f(0) = 0$, we have $f(\lambda) \ge f(s) \cdot \frac{\lambda}{s} \ge f(s) \cdot (1 - \delta)$. Thus $\mathrm{H}_{\mathrm{ber}}(\mathbf{q}, 1) - \mathrm{H}_{\mathrm{pois}}(\mathbf{q}, 1) = f(s) - f(\lambda) \le \delta \cdot f(s) = \delta \cdot \mathrm{H}_{\mathrm{ber}}(\mathbf{q}, 1)$, which concludes the proof. ☐

B Missing Proofs in Section 5

Theorem 5.4. *Algorithm 5 works in polynomial time and in expectation is a* $\left(1 - 43\, k^{-1/4} - O(k^{-1/2})\right)$*-approximation to the integral BSP for any position auction.*

Proof of Theorem 5.4. Recall that the social welfare $\mathrm{SW}(S)$ of a ℓ-unit or position auction is submodular as a function of the invited set of bidders S.

Claim B.1. *The expected social welfare* $\mathrm{SW}(S)$ *for a set of bidders* $S \subseteq [n]$ *in any* ℓ*-unit or position auction is a submodular function of* S*.*

Analysis of Algorithm 5. Clearly, $\mathbf{z}$ is a feasible solution to the integral BSP. We will prove below that Algorithm 5 has almost the same approximation guarantee as Algorithm 4. Let us denote the optimal social welfare for integral BSP as OPT. Then the best solution $\mathbf{x}^*$ to the fractional BSP (4) may have only higher welfare $\mathrm{SW}(\mathbf{x}^*, \mathbf{w}) \ge \mathsf{OPT}$. Furthermore, since (4) is a multi-linear extension of the integral BSP, we have

$$\mathop{\mathbf{E}}_{\mathbf{y} \sim \mathrm{Ber}(\mathbf{x})}[\mathrm{SW}(\mathbf{y}, \mathbf{w})] = \mathrm{SW}(\mathbf{x}, \mathbf{w}) \ge \left(1 - 43\, k^{-1/4}\right) \mathrm{SW}(\mathbf{x}^*, \mathbf{w}) \ge \left(1 - 43\, k^{-1/4}\right) \mathsf{OPT}.$$

Our solution $\mathbf{z}$ suffers an additional loss when the sample vector $\mathbf{y}$ has more than k elements $|\mathbf{y}|_1 > k$. On the other hand, for each given $\mathbf{y}$ with $|\mathbf{y}|_1 > k$ we have $\mathbf{E}_{\mathbf{z} \sim \binom{\mathbf{y}}{k}}[\mathrm{SW}(\mathbf{z}, \mathbf{w})] \ge \frac{k}{|\mathbf{y}|_1} \mathrm{SW}(\mathbf{y}, \mathbf{w})$, due to submodularity of $\mathrm{SW}(\mathbf{y}, \mathbf{w})$ (here we use a standard fact about monotone non-negative submodular function f: a uniformly sampled subset $T \subset S$ of given size $|T| = k$ has $\mathbf{E}_T[f(T)] \ge \frac{k}{|S|} f(S)$). Thus

$$\mathop{\mathbf{E}}_{\mathbf{y} \sim \mathrm{Ber}(\mathbf{x})}[\mathrm{SW}(\mathbf{y}, \mathbf{w}) - \mathrm{SW}(\mathbf{z}, \mathbf{w})] \le \mathop{\mathbf{E}}_{\mathbf{y} \sim \mathrm{Ber}(\mathbf{x})}\left[\mathbb{I}\Big[|\mathbf{y}|_1 > k\Big] \cdot \frac{|\mathbf{y}|_1 - k}{|\mathbf{y}|_1} \cdot \mathrm{SW}(\mathbf{y}, \mathbf{w})\right].$$

Furthermore, as $\mathsf{OPT} = \max_{S\subseteq[n]:|S|=k} \mathrm{SW}(S, \mathbf{w})$ we have $\mathsf{OPT} \geq \mathbf{E}_{\mathbf{z}\sim\binom{\mathbf{y}}{k}}[\mathrm{SW}(\mathbf{z}, \mathbf{w})] \geq \frac{k}{|\mathbf{y}|_1}\mathrm{SW}(\mathbf{y}, \mathbf{w})$ for each $\mathbf{y}$ with $|\mathbf{y}|_1 > k$. Therefore,

$$\mathop{\mathbf{E}}_{\substack{\mathbf{y}\sim\mathrm{Ber}(\mathbf{x})\\ \mathbf{z}\sim\binom{\mathbf{y}}{k}}}[\mathrm{SW}(\mathbf{y}, \mathbf{w}) - \mathrm{SW}(\mathbf{z}, \mathbf{w})] \leq \mathop{\mathbf{E}}_{\mathbf{y}\sim\mathrm{Ber}(\mathbf{x})}\left[\mathbb{I}\Big[|\mathbf{y}|_1 > k\Big] \cdot \frac{|\mathbf{y}|_1 - k}{k} \cdot \mathsf{OPT}\right]$$
$$= \frac{\mathsf{OPT}}{k} \cdot \sum_{i=1}^{n-k} i \cdot \mathop{\mathbf{Pr}}_{\mathbf{y}\sim\mathrm{Ber}(\mathbf{x})}[|\mathbf{y}|_1 = k + i] = \frac{\mathsf{OPT}}{k} \cdot \sum_{i=1}^{n-k} \mathop{\mathbf{Pr}}_{\mathbf{y}\sim\mathrm{Ber}(\mathbf{x})}[|\mathbf{y}|_1 \geq k + i].$$

Claim B.2. *Let* $\mathbf{x} \in [0, 1]^n$ *with* $|\mathbf{x}|_1 = k$. *Then* $\sum_{i\geq 1} \mathbf{Pr}_{\mathbf{y}\sim\mathrm{Ber}(\mathbf{x})}[|\mathbf{y}| \geq k + i] = O(\sqrt{k})$.

Proof of Claim B.2. By Chernoff bound, we have

$$\sum_{i=1}^{n-k} \mathop{\mathbf{Pr}}_{\mathbf{y}\sim\mathrm{Ber}(\mathbf{x})}[|\mathbf{y}| \geq k + i] \leq \sum_{i=1}^{n-k} e^{-\frac{i^2}{i+2k}} \leq \sum_{i=1}^{\infty} e^{-\frac{i^2}{i+2k}} \leq \int_0^\infty e^{-\frac{x^2}{x+2k}}\,\mathrm{d}x.$$

We further estimate this integral as

$$\int_0^\infty e^{-\frac{x^2}{x+2k}}\,\mathrm{d}x \leq \int_0^{\sqrt{k}} e^{-\frac{x^2}{x+2k}}\,\mathrm{d}x + \int_{\sqrt{k}}^\infty e^{-\frac{x}{1+2\sqrt{k}}}\,\mathrm{d}x \leq 1 \cdot \sqrt{k} - (1 + 2\sqrt{k})\,e^{-\frac{x}{1+2\sqrt{k}}}\Big|_{\sqrt{k}}^{+\infty} = O(\sqrt{k}).$$

This concludes the proof of Claim B.2. $\square$

Claim B.2 allows us to conclude that $\mathbf{E}_{\mathbf{y}\sim\mathrm{Ber}(\mathbf{x})}[\mathrm{SW}(\mathbf{y}, \mathbf{w}) - \mathrm{SW}(\mathbf{z}, \mathbf{w})] = O\big(\frac{1}{\sqrt{k}}\big) \cdot \mathsf{OPT}$, i.e.,

$$\mathop{\mathbf{E}}_{\mathbf{y}\sim\mathrm{Ber}(\mathbf{x})}[\mathrm{SW}(\mathbf{z}, \mathbf{w})] = \mathrm{SW}(\mathbf{x}, \mathbf{w}) - O\left(\frac{1}{\sqrt{k}}\right) \cdot \mathsf{OPT} = \Big(1 - 43\,k^{-1/4} - O(k^{-1/2})\Big)\mathsf{OPT}. \quad \square$$