Simple and Order-optimal Correlated Rounding Schemes for Multi-item E-commerce Order Fulfillment

(Authors' names blinded for peer review)

A fundamental problem faced in e-commerce is—how can we satisfy a multi-item order using a small number of fulfillment centers (FC's), while also respecting long-term constraints on how frequently each item should be drawing inventory from each FC? In a seminal paper, Jasin and Sinha (2015) identify and formalize this as a correlated rounding problem, and propose a scheme for *randomly* assigning an FC to each item according to the frequency constraints, so that the assignments are *positively correlated* and not many FC's end up used. Their scheme pays at most $\approx q/4$ times the optimal cost on a q-item order. In this paper we provide to our knowledge the first substantial improvement of their scheme, which pays only $1 + \ln(q)$ times the optimal cost. We provide another scheme that pays at most d times the optimal cost, when each item is stored in at most d FC's. Our schemes are fast and based on an intuitive new idea—items wait for FC's to "open" at random times, but observe them on "dilated" time scales. We also provide matching lower bounds of $\Omega(\log q)$ and d respectively for our schemes, by showing that the correlated rounding problem is a non-trivial generalization of Set Cover. Finally, we provide a new LP that solves the correlated rounding problem exactly in time exponential in the number of FC's (but not in q).

E-commerce has exploded in recent times, achieving unbelievable global scale, unimaginable delivery speed, and unfathomable system complexity. The short-term operations of a typical e-commerce giant involves pulling inventory from suppliers into its fulfillment centers (FC's), including retail stores that can also be used to fulfill online orders; waiting for customers to make purchases, which can be influenced by its powerful search/recommendation engine; and finally delivering the goods to the customer's doorstep, through a flexible transportation system that allows almost any FC in the network to be used for fulfilling demand from any particular region. This paper focuses on the final part of these operations, which is the problem of dynamically dispatching incoming customer orders to FC's, while treating the inventory replenishment schedule and search ranking/recommendation decisions as exogenous.

This dynamic fulfillment problem is challenging for several reasons. First, decisions must be made with consideration of the future orders to come, since depleting inventories at the wrong places can set off a chain reaction of long-distance and split shipments, as originally demonstrated by Xu et al. (2009). However, due to the uncertainty in future orders, forward-lookingness requires a high-dimensional stochastic dynamic program that is intractable to solve, as noted by Acimovic and Farias (2019). Meanwhile, even a myopic strategy like using the minimum number of FC's to satisfy each incoming order, without consideration of future orders, can be computationally hard. Finally, the mere scale and speed of the problem restricts us to fast and simple heuristics, with more elaborate optimizations exacerbating the issue of system complexity.

In light of these challenges, a prevailing approach to the dynamic fulfillment problem is LP-based, as pioneered by Jasin and Sinha (2015). In a nutshell, an LP which views the system as deterministic is written, describing inventory levels of every item at every FC, and expected demands at different regions which includes information about items frequently purchased together in the same order. The objective captures fixed shipping costs (mostly dependent on the number of distinct FC's used to fulfill an order), variable shipping costs (dependent on items and distances), and shortage costs (dependent on penalties paid for orders not fulfilled). The LP is then solved, providing a “master plan” of transporting supply to demand, which prescribes for different orders from different regions, how frequently each FC should be used to fulfill each item in that order. As orders come up in real-time, Jasin and Sinha (2015) randomly dispatch the items to FC’s, making sure to follow the fulfillment frequencies outlined in the LP’s plan.

Although seemingly uninformed, this *randomized fulfillment* algorithm is simple and fast, dispatching instantly and not requiring real-time inventory information across the network once the LP solution is given. Under large system scales, it also performs well, in terms of paying variable shipping and shortage costs similar to that of the LP benchmark. However, fixed costs remain a challenge—the problem of minimizing fixed costs for a single order was already difficult, and having to follow the LP’s fulfillment frequencies only introduces additional constraints. The seminal insight from Jasin and Sinha (2015) is that these frequencies are actually helpful—when using them to randomly assign an FC to each item, if *positive correlation* is induced in the assignments across items, then many items will end up assigned to the same FC, resulting in not many fixed costs being paid. Jasin and Sinha (2015) also derive an intricate method for inducing this correlation, which reduces the fixed costs from the naive independent method by a factor of 4.

Despite its significance and impact on subsequent work (e.g. Lei et al. 2018, 2021, Zhao et al. 2020), to our understanding, the correlation method of Jasin and Sinha (2015) has never been improved in a substantial way, until now. In this paper, we derive a new method (and extension)—based on observing Poisson processes under “dilated” time scales—that is intuitively simple, computationally faster, and achieves the *best-possible guarantee (in two different regimes).*

Correlated Rounding Problem of Jasin and Sinha (2015)

Consider a single order (from a particular region at a particular time), consisting of q items. For each item $i = 1, \ldots, q$ in the order, we are given the fraction of time u_{ki} it must be fulfilled from each eligible FC $k = 1, \ldots, K$, with $\sum_k u_{ki} = 1$. For each FC k, a fixed cost of c_k is paid if k is used to fulfill any item. The goal is to randomly “round’ each item’s probability vector $(u_{ki})_{k=1}^K$ to an actual FC for fulfilling item i, in a correlated fashion that minimizes the expected fixed costs paid.

Problem 1 (Jasin and Sinha (2015)). Given q marginal distributions $(u_{k1})_{k=1}^K, \ldots, (u_{kn})_{k=1}^K$ over a discrete set $\{1, \ldots, K\}$ and fixed costs $c_1, \ldots, c_K$, construct jointly-distributed random variables $Z_1, \ldots, Z_n$ satisfying $\Pr[Z_i = k] = u_{ki}$ for all k and i that minimizes

$$\sum_{k=1}^{K} c_k \cdot \Pr\left[\bigcup_{i=1,\ldots,q} (Z_i = k)\right]. \tag{1}$$

In Problem 1, $Z_i \in \{1, \ldots, K\}$ denotes the FC used to fulfill item i, and $\bigcup_i (Z_i = k)$ denotes the event that FC k is used to fulfill any item, in which case its fixed cost c_k must be paid. Although solving Problem 1 is hard in general, approximate solutions can be derived by observing that

$$\Pr\left[\bigcup_{i=1,\ldots,q} (Z_i = k)\right] \geq \max_{i=1,\ldots,q} \Pr[Z_i = k] = \max_i u_{ki}. \tag{2}$$

In words, $\max_i u_{ki}$ is a lower bound on the probability with which FC k must be used, and hence it suffices to ensure that no FC k is used too often in comparison to $\max_i u_{ki}$. Jasin and Sinha (2015) actually focus on deriving the following, which we will call α-competitive rounding schemes.

Definition 1 (α-competitive Rounding Scheme). For $\alpha \geq 1$, an *α-competitive (correlated) rounding scheme* is a method for constructing random variables $Z_1, \ldots, Z_n$ satisfying

$$\Pr[Z_i = k] = u_{ki} \qquad \forall i = 1, \ldots, q, k = 1, \ldots, K \tag{3}$$

$$\Pr\left[\bigcup_{i=1,\ldots,q} (Z_i = k)\right] \leq \alpha \cdot \max_i u_{ki} \qquad \forall k = 1, \ldots, K \tag{4}$$

given any q marginal distributions $(u_{k1})_{k=1}^K, \ldots, (u_{kn})_{k=1}^K$ over a discrete set $\{1, \ldots, K\}$.

An α-competitive rounding scheme provides a solution to Problem 1 that pays at most α times the optimal cost, due to the lower bound derived in (2). Jasin and Sinha (2015) derive a $B(q)$-competitive correlated rounding scheme, where $B(q) = \frac{(q+1)^2}{4n}$ if q is odd and $B(q) = \frac{q+2}{4}$ if q is even, with function $B(q)$ growing approximately as $q/4$. Meanwhile, it is easy to see that the naive independent rounding scheme is only q-competitive, which is worse than $B(q)$ by a factor of approximately 4. Our new result in this paper is a $(1+\ln(q))$-competitive rounding scheme, improving the order-dependence on q entirely, and matching an $\Omega(\log q)$ lower bound. Moreover, we use similar ideas to derive a d-competitive rounding scheme when d is an upper bound on the number of different FC's that can fulfill an item, which we also show is best-possible.

Summary of Contributions

We now list all our results related to Problem 1 and Definition 1, highlighting the virtues of our rounding schemes in relation to Jasin and Sinha (2015), which we refer to as "JS".

- We derive a $(1+\ln(q))$-competitive rounding scheme, which is a substantial improvement upon the $\approx q/4$-competitive rounding scheme of JS. We also derive a d-competitive rounding scheme, where $d := \max_i |\{k : u_{ki} > 0\}|$ is a sparsity parameter. These guarantees match respective lower bounds of $\Omega(\log q)$ and d, as will be shown below.
- Both of our rounding schemes have a runtime of $O(qK)$. By contrast, the rounding scheme of JS has a runtime of $O(q^2K)$, containing a loop that is quadratic in the number of items q.
- Our rounding schemes are intuitive—FC's have random opening times, and items are assigned to the first FC they see open on a dilated time scale. By contrast, the method of JS based on constructing *line partitions*, while clever and beautiful, is to our understanding not simple.
- We should acknowledge that our guarantee is $1+\ln 2 \approx 1.69$ when $q=2$. By contrast, JS is 1-competitive if $q=2$. We also note that if there are only two FC's, i.e. $K=2$, then a 1-competitive rounding scheme was recently discovered by Zhao et al. (2020). In this scenario, our second rounding scheme would only be 2-competitive, since $d=K=2$. However, we emphasize that parameter d represents the maximum number of distinct FC's that hold an item and can generally be much smaller than K, whereas their rounding scheme only works when $K=2$.

Next, we make further contributions by rigorously relating α-competitive rounding schemes to notions from the Set Cover problem, establishing the following.

- We show that an α-competitive rounding scheme implies a procedure for rounding a fractional Set Cover solution into a randomized cover, that is feasible w.p. 1, and having no set chosen with probability more than α times its fractional weight.
- Therefore, we can leverage hardness results from Set Cover to show that an α-competitive rounding scheme must have $\alpha = \Omega(\log q)$ and $\alpha \geq d$. The latter lower bound establishes our d-competitive rounding scheme to be exactly tight, not just order-optimal.
- Our $(1+\ln(q))$-competitive rounding scheme also improves existing guarantees in the aforementioned randomized rounding problem for Set Cover. Existing methods need to select sets with probability at least $2\ln(q)$ times their fractional weight. The key is that our method induces sets to be selected in a *negatively correlated* fashion, whereas existing methods select sets *independently* and only show that the solution is feasible with high probability.

We note that for the Set Cover problem itself, our rounding schemes do not improve existing guarantees—the Greedy algorithm already has a guarantee of $1+1/2+\cdots+1/q$ which is smaller (better) than our $1+\ln(q)$. A fractional Set Cover solution is also easily converted into an integral one while losing a factor of at most d. Nonetheless, we believe these connections highlight how the correlated rounding problem is a harder version of Set Cover—in which a randomized solution, that must satisfy constraints on how often each set is used to cover each element, is required. Furthermore, it is interesting to us that a modern problem from e-commerce practice, identified by Jasin and Sinha (2015), can lead us to improve randomized rounding schemes for the age-old Set Cover problem from CS theory.

References

Acimovic J, Farias VF (2019) The fulfillment-optimization problem. *Operations Research & Management Science in the age of analytics*, 218–237 (INFORMS).

Jasin S, Sinha A (2015) An lp-based correlated rounding scheme for multi-item ecommerce order fulfillment. *Operations Research* 63(6):1336–1351.

Lei Y, Jasin S, Sinha A (2018) Joint dynamic pricing and order fulfillment for e-commerce retailers. *Manufacturing & Service Operations Management* 20(2):269–284.

Lei Y, Jasin S, Uichanco J, Vakhutinsky A (2021) Joint product framing (display, ranking, pricing) and order fulfillment under the multinomial logit model for e-commerce retailers. *Manufacturing & Service Operations Management* .

Xu PJ, Allgor R, Graves SC (2009) Benefits of reevaluating real-time order fulfillment decisions. *Manufacturing & Service Operations Management* 11(2):340–355.

Zhao Y, Wang X, Xin L (2020) Multi-item online order fulfillment: A competitive analysis. *Chicago Booth Research Paper* (20-41).