Operations Research Letters 42 (2014) 70–75

Contents lists available at ScienceDirect

Operations Research Letters

journal homepage: www.elsevier.com/locate/orl

Sequential resource allocation with constraints: Two-customer case

Xin Geng *, Woonghee Tim Huh, Mahesh Nagarajan

Sauder School of Business, University of British Columbia, 2052 Main Mall, Canada V6T 1Z2

ARTICLE INFO

Article history:
Received 3 May 2013
Received in revised form
23 November 2013
Accepted 6 December 2013
Available online 12 December 2013

Keywords:
Sequential resource allocation
Equity
Structural property

ABSTRACT

We study a two-customer sequential resource allocation problem with equity constraint, which is reflected by a max–min objective. For finite discrete demand distribution, we give a sufficient and necessary condition under which the optimal solution has monotonicity property. However, this property never holds with unbounded discrete distribution.

1. Introduction

The sequential resource allocation (SRA) problem has received much attention in the literature. In this problem, a supplier has a limited quantity of resource available for allocation. Independent random demands arrive sequentially from a number of customers (or agencies), and the supplier needs to sequentially allocate the resource for each customer at a time. When allocating resource to a customer, the supplier sees the realization of the customer's demand, but does not know the remaining demands except for their distributions. Hence, the trade-off is whether to allocate the limited resource to the current customer or save it for future demands.

Two types of objectives are commonly studied in this area. The first type involves maximizing profit/revenue. The single resource capacity allocation problem in revenue management is a good example; see the first chapter of [18] for the detailed study on the theoretical properties and useful heuristics.

The second type of objective in SRA problem does not explicitly model monetary pay-off. When some non-profit organizations act as the supplier, they often aim at enhancing the satisfaction level of the overall society rather than making profit. In fact, the SRA problem arises naturally in the contexts such as healthcare allocation and food distribution; see [6,16] for example. Furthermore, Savas [15] argues that equity (fairness or impartiality of service), among other performance measures, deserves much more attention. This is true especially for resource allocation problems; Bertsimas et al. [5] study a class of efficiency–fairness objective functions and indicate applications in allocating resources. However, only few papers incorporate fairness into SRA problem. We refer the readers to the doctoral thesis by Lien [9], which is closest to our work, and the references therein. In this paper, we focus our attention to SRA problems with equity (fairness) as the objective, and we call our problem the *equity based* SRA problem.

As one may imagine, the term equity is amorphous and its meaning varies according to the context. Although no single criterion is universally accepted in every setting, Bertsimas et al. [5] discuss some general theories on justice and fairness that serve as the basis of most equity measures. They are all related to maximizing the social welfare function (SWF) defined by (1). A general form of SWF that incorporates fairness measure is introduced in [2] and studied in many classic textbooks [12]. Let $U = (u_1, \ldots, u_n)$ be a utility vector of n agents and $\rho \leq 1$ be a real number, then this SWF has the form

$$W(U) = \left(\sum_i u_i^{\rho}\right)^{1/\rho} \quad \text{for } \rho \neq 0, \tag{1}$$

and

$$W(U) = \sum_i \log u_i \quad \text{for } \rho = 0.$$

To see the relation between maximizing $W(U)$ and fairness criteria, we take three special values of ρ for example. First, when $\rho = 1$, $W(U)$ is simply the sum of all utilities. Maximizing $W(U)$ is in the principle of utilitarianism. However, Young [19] has argued that it fails to achieve fairness. Second, maximizing $W(U)$ when $\rho = 0$ results in the Nash solution, proposed by Nash [13]. The Nash solution is considered to be a fair allocation by many researchers (e.g., [8,4]). Finally, when $\rho \to -\infty$, $W(U)$ equals the

* Correspondence to: 2329 W Mall, Vancouver, Canada V6T 1Z4.
E-mail addresses: xin.geng@sauder.ubc.ca (X. Geng), tim.huh@sauder.ubc.ca (W.T. Huh), mahesh.nagarajan@sauder.ubc.ca (M. Nagarajan).

0167-6377/$ – see front matter
http://dx.doi.org/10.1016/j.orl.2013.12.004

minimum utility. Then the objective is to make the minimum utility as large as possible, which is in the principle of *Rawlsian justice* (proposed by Rawls [14]). This max–min criterion has been widely used in data network [3] and has initiated applications in bandwidth allocation problems [11] as well as general resource allocation problems [1,10]. In general, the parameter ρ indicates the level of inequity. The aversion to inequity increases as ρ decreases to negative infinity [5,12]. Hence, Rawlsian justice retains the most fairness. Chevaleyre et al. [7] claim that the minimum utility "offers a level of fairness and may be a suitable performance indicator when we have to satisfy the minimum needs of a large number of customers" (p. 17). This fits to our setting of non-profit food allocation very well. Consequently, we will use a max–min objective which is in line with the commonly used Rawlsian justice. Although our objective function is based on (1) with $\rho \to -\infty$, the result also holds for finite negative ρ values (see Section 3.1). However, we will not consider the cases $0 \le \rho \le 1$. Furthermore, we model the customer's utility as the ratio of the allocated amount to the demand, which is named *fill rate*. Fill rate captures the proportion of demand satisfied for each customer, and customers tend to compare this measure with one another after allocation is completed. Indeed, both [9,17] adopt this measurement in their models, and we believe it is a suitable candidate.

Our work provides theoretical results concerning the basic structure of the optimal solutions. The limitation of two-customer special case notwithstanding, our results serve as a reference for further theoretical studies and heuristic development. In addition, our contribution differs from that of the theoretical work on profit-based SRA (e.g. [18]); since the objective functions have different forms and properties, their results or methods are not directly applicable to the equity based SRA.

2. Model formulation

A supplier needs to allocate to N customers with a fixed amount s of resource. The customers are sequentially ordered. Each customer's demand is random and will be known to the supplier only after all demands of the previous customers have been realized and allocation decisions to those customers have been made, but before the allocated amount to him is decided. As discussed in the previous section, we aim to maximize the minimal fill rate of all customers to achieve Rawlsian justice. Since the demands are random, our objective is therefore to maximize the expected minimum fill rate over all the customers. In this paper, we focus on the two-customer case only. Studying this special case simplifies the problem while keeping the inherent challenges of sequential decision making. Besides, it is straightforward to extend some of the main results to multiple customers. Hence, we aim at finding structural properties that help understanding the N-customer case.

Let x_i $(i = 1, 2)$ be the allocation to customer i. Since $x_2 = s - x_1$, x_1 is the only decision variable. Let D_i $(i = 1, 2)$ be the random variable representing the demand from the customer and d_i $(i = 1, 2)$ be the realized demand. Throughout this paper we assume that the two demands are independent (but not necessarily identical), which is commonly assumed in the literature. Moreover, we assume that $D_i > 0$ $(i = 1, 2)$ almost surely. Then, the fill rate takes the form of x_i/D_i $(i = 1, 2)$.

Let $a \wedge b$ represent $\min\{a, b\}$. Given an initial supply $s > 0$ and a realized first demand d_1, define

$$R(x_1, d_1) = E_{D_2}\left(\frac{x_1}{d_1} \wedge \frac{s - x_1}{D_2}\right) \tag{2}$$

to be the expectation of the minimum of the two fill rates, where $0 \le x_1 \le d_1$ and the expectation is taken with respect to D_2. It is straightforward to see that function $R(x_1, d_1)$ is jointly concave in x_1 and s. Further, let

$$v(d_1) = \max_{0 \le x_1 \le \min\{s, d_1\}} R(x_1, d_1), \tag{3}$$

then the optimal expected minimum fill rate is given by $u = E_{D_1} v(D_1)$. Since x_1 is decided after D_1 is realized, we need only to focus our attention on the random variable D_2 and how it affects the structure of the optimal decision conditioned on the realized value of the first demand d_1.

Our interest, therefore, is in solving (3). Let $x_1^* = x_1^*(d_1)$ be an optimal solution to (3). Note that the constraint $x_1 \le d_1$ simply says that there is no need to give the first customer more than demanded. As a result, $v(d_1)$ cannot exceed 1. Equivalently, one can remove the constraint $x_1 \le d_1$ and modify the objective function as $E_{D_2}\left(1 \wedge \frac{x_1}{d_1} \wedge \frac{s-x_1}{D_2}\right)$ (see [9] for an example). Although they are equivalent formulation, we will use the former objective function, which turns out to be easier to work with. To further simplify the problem, we enlarge the feasible region for x_1 in (3), and consider the following relaxed problem.

$$\tilde{v}(d_1) = \max_{0 \le x_1 \le s} R(x_1, d_1). \tag{4}$$

Let $\phi(d_1)$ be an optimal solution to (4). Comparing the two problems, the feasible region in (4) is not bounded by d_1. Therefore, $x_1^*(d_1)$ is at most d_1 while $\phi(d_1)$ may exceed d_1; similarly, $v(d_1)$ is bounded above by 1 while $\tilde{v}(d_1)$ may exceed 1. However, the two problems are closely related. In fact, if $\phi(d_1) \le d_1$, then it is easy to see that the relaxed problem has an optimum that is feasible for the original problem so $x_1^*(d_1) = \phi(d_1)$. Otherwise, if $s \ge \phi(d_1) > d_1$, then by concavity of $R(x_1, d_1)$, the objective function is increasing in x_1 on the interval $[0, \phi(d_1)]$. Hence it is also increasing on $[0, d_1]$, which implies that the optimum of the original problem is achieved at the boundary d_1, i.e., $x_1^*(d_1) = d_1$. To summarize, we always have $v(d_1) \le \tilde{v}(d_1)$ and

$$x_1^*(d_1) = \phi(d_1) \wedge d_1. \tag{5}$$

For our SRA problem, we are primarily interested in the optimal allocation policy rather than the optimal value of the objective function. To understand the structure of $x_1^*(d_1)$, it is convenient for us to first study $\phi(d_1)$ in the relaxed problem (4) since we can recover the information about the optimal allocation through (5). To the best of our knowledge, this method of relaxation has not been used in such equity based SRA problem before. We ask the following question: what happens to the optimal allocation as d_1 increases? In other words, if the realized demand d_1 is higher, do we always allocate more to the first customer? This is one of the research questions analysed in this paper, and is discussed in the next section.

3. Structure of optimal solution

Preliminary intuition drives us to believe that the larger d_1 is, the more we should allocate to it, resulting in larger x_1^*. This is asserted to be true in [9] and later corrected in [10]. In this section, we give another example where this is not true and provide sufficient conditions under which this assertion is true.

3.1. Example: non-monotonicity of $x_1^*(d_1)$ in d_1

Suppose the second period demand D_2 is such that $D_2 = 1$ or 4 with equal probability. Let the initial supply $s = 1$. We then compute the optimal allocation for two cases, in which $d_1 = 3$ and $d_1 = 5$, respectively. Using (2) and some algebra, we have

$$R(x_1, 3) = \frac{1}{2}\left(\frac{x_1}{3} \wedge (1 - x_1)\right) + \frac{1}{2}\left(\frac{x_1}{3} \wedge \frac{1 - x_1}{4}\right).$$