

# LIMITING BEHAVIOR OF THE TARGET-DEPENDENT STOCHASTIC SEQUENTIAL ASSIGNMENT PROBLEM

GOLSHID BAHARIAN\* AND  
SHELDON H. JACOBSON,\*\* *University of Illinois at Urbana-Champaign*

## Abstract

The stochastic sequential assignment problem assigns distinct workers to sequentially arriving tasks with stochastic parameters. In this paper the assignments are performed so as to minimize the *threshold probability*, which is the probability of the long-run reward per task failing to achieve a target value (threshold). As the number of tasks approaches infinity, the problem is studied for independent and identically distributed (i.i.d.) tasks with a known distribution function and also for tasks that are derived from  $r$  distinct unobservable distributions (governed by a Markov chain). Stationary optimal policies are presented, which simultaneously minimize the threshold probability and achieve the optimal long-run expected reward per task.

*Keywords:* Sequential assignment; threshold probability; stationary policy; hidden Markov chain

2010 Mathematics Subject Classification: Primary 60K30  
Secondary 60J20

## 1. Introduction

Consider the sequential stochastic assignment problem (SSAP) introduced in [6], where  $n$  workers are available to perform  $n$  independent, identically distributed (i.i.d.) sequentially arriving tasks. The random variable  $X_j$  denotes the  $j$ th task value, and a value (success rate)  $p_i$  is associated with each worker. Whenever the  $i$ th worker is assigned to the  $j$ th task that worker becomes unavailable for future assignments, with  $p_i x_j$  denoting the expected reward due to this assignment. The objective is to assign these  $n$  workers to  $n$  tasks so as to maximize the expected total reward. It is shown in [6] that there exists numbers

$$-\infty = a_{0,n} \leq a_{1,n} \leq a_{2,n} \leq \cdots \leq a_{n,n} = +\infty, \quad (1.1)$$

such that the optimal choice in the initial stage is to assign the  $i$ th best available worker if the random variable  $X_1$  falls within the  $i$ th highest interval. The SSAP has applications in several areas and various extensions to the problem have been discussed in the literature. For example, [10] studied a variation of the SSAP in aviation security screening systems, while [11] addressed the problem of allocating sequentially arriving donor kidneys to patients on a transplant waiting list. Another application of the SSAP is the asset selling problem [3], where there is a need to choose the best offers out of a sequence of bids from potential buyers. Moreover, Albright [1]

---

Received 19 June 2013; revision received 17 December 2013.

\* Postal address: Research Center of CHU Sainte-Justine, 3175 Cote Sainte-Catherine, local A-714, Montreal, Québec H3T 1C5 Canada. Email address: gbahari2@illinois.edu

\*\* Postal address: Department of Computer Science, 201 N. Goodwin Ave., Urbana, IL 61801, USA.  
Email address: shj@illinois.edu

studied the SSAP with various task-arrival-time distributions. Nikolaev and Jacobson [9] considered a variation of the SSAP in which the number of tasks are unknown until after the arrival of the final task which followed a given probability distribution.

Implementing the optimal assignment policy for the SSAP, as described in [6], involves calculating a new set of breakpoints upon the arrival of each task. The computation of these breakpoints takes polynomial time but is cumbersome for large scale problems. For example, consider an SSAP which allows passengers at an airport check-in to be assigned to the available security resources based on their perceived individual threat level [9]. Assigning passengers to security resources by recalculating the breakpoints every time a passenger arrives is not practical, even for a small airport. Therefore, this paper focuses on the limiting behavior of the  $\{a_{i,n}\}$ , as  $n$  approaches  $\infty$ , so as to obtain practical solutions that can be implemented in real world situations. The existing SSAP literature focuses on a risk neutral objective function, seeking an assignment policy that maximizes the expected total reward. However, a risk neutral policy is not always desirable since the probability distribution function (*pdf*) of the total reward may carry with it a high probability of low values. Consequently, there are instances when a decision maker is interested in a stable reward and looks for a risk sensitive optimal assignment policy.

Taking the above mentioned issues into consideration, this paper studies the limiting behavior of the SSAP under a different objective function, called the *threshold criterion*. For a given *threshold* (or target)  $\tau$ , the goal is to find a policy  $\phi^*$  that minimizes the *threshold probability*: the probability (or risk) of the long-run reward per task failing to achieve the target  $\tau$ . Specifically, the threshold criterion can be expressed as

$$\inf_{\phi \in \Phi} \mathbb{P} \left\{ \limsup_{n \rightarrow +\infty} \frac{1}{n} R_n^\phi \leq \tau \right\},$$

where  $\Phi$  is the set of all admissible policies and  $R_n^\phi$  is the total reward obtained after assigning all  $n$  tasks under policy  $\phi$ . For simplicity, this problem is denoted as the LTSSAP since it studies the limiting behavior of the target dependent SSAP. In this paper we study two versions of this problem. The first version assumes that the sequentially arriving tasks are i.i.d. with a known distribution function. The second problem then considers the case that the task values are derived from  $r$  different distributions, where the successive distributions are governed by an ergodic Markov chain. Once a task arrives in a given time period its value is observed; however, the source of the distribution is unobservable. In both problems it is assumed that there exist  $k$  worker categories, where the  $i$ th category consists of  $r_i$  workers each with value  $p_i$  such that  $\sum_{i=1}^k r_i = n$ . Stationary policies are presented for both problems, which apart from minimizing the threshold probability, achieve the optimal long run expected reward per task. As opposed to the policy described by [6] (see (1.1)) in which the number of breakpoints increases with  $n$  and the (time-dependent) breakpoints are recalculated each time a task arrives, the assignment policies presented in this paper are characterized by  $k - 1$  fixed (time-independent) breakpoints. Comparing the value of each arriving task with these fixed breakpoints (regardless of the arrival time) determines the worker class to which that task must be assigned.

In the existing literature, [4] and [2] addressed the limiting behavior of a special case of SSAP (called the secretary problem) where  $p_i \in \{0, 1\}$ . Specifically, [2] assumed that the task values were generated by  $r$  different distributions, and once a task arrived, both its value and its distribution were observed. However, the assumption that the underlying task distributions are observable does not hold in most real world problems. Deviating from the existing literature, the present paper considers the task distributions to be both unobservable and forming a hidden

Markov chain. As an application of the problem with a hidden Markov chain, consider the case where the task worth at time period  $j$  is denoted by  $X_j$ . It is safe to assume that the task worth is dependent on the economic conditions upon the arrival of that task [8]. Clearly, the economic conditions vary from time to time. Suppose that the changes in these economic conditions are modeled by a Markov chain with state space  $\mathcal{S}$ , where the states of the chain represent the (unobservable) economic conditions. Specifically, assume that if the economic condition upon the arrival of  $X_j$  is given by  $k \in \mathcal{S}$ , then  $X_j$  is derived from the distribution function  $F_k$ . Hence, we are dealing with an observable sequence of arriving tasks, where the underlying (unobservable) economic conditions affect the tasks' worth, with a Markov chain modeling the transitions in economical conditions. Although [4] laid out the framework for studying the limiting behavior of the SSAP, this paper takes on a different approach to the proof for two reasons.

- (i) Dealing with the LTSSAP involves studying the almost sure (a.s.) convergence of the long-run reward per task, while the existing literature focuses on the convergence of the long-run expected value of reward per task.
- (ii) As mentioned in [5], the generalization of the problem with  $p_i \in \{0, 1\}$  to the case with arbitrary worker values is not possible using the approach applied in [4].

Consider an application of LTSSAP in aviation security, where sequentially arriving passengers at an airport check-in are assigned to the available security resources. A random variable  $X_j$  is associated with passenger  $j$ , denoting their threat (risk) value. The threat value is defined as the probability of a passenger carrying a threat item. Upon arrival, a prescreening system determines the passenger threat value and assigns them to either a non-selectee class (i.e. a class of passengers who have been cleared of posing a threat) or a selectee class (i.e. the class who have not been cleared). A security level is assigned to each class, denoting the probability of detecting a passenger with a threat item. Let  $L_S$  and  $L_{NS}$  be the security levels associated with the selectee and the non-selectee classes. Moreover, let  $\gamma_j = 1$  and  $\gamma_j = 0$  denote the  $j$ th passenger assignment as a selectee and a non-selectee, respectively. The *total security* for this setting is defined as

$$\sum_{j=1}^n X_j [L_S \gamma_j + L_{NS} (1 - \gamma_j)].$$

At any airport it is critical to maintain a stable and reasonable level of continuous security with a high probability of detecting threats. Therefore, the objective is to find a policy for assigning passengers to classes as they check in so as to minimize the probability of the long-run average security failing to achieve the target  $\tau$ . In most of the real world settings the distributions generating the passenger threat values are not observable. Moreover, threat values of different passengers are likely to have been generated from different distributions. To see how the model with unobservable distributions (involving a Markov chain) can be useful in this setting consider a group of two or more passengers who are planning an attack. The passenger threat values depend on one another in the sense that if one passenger is classified as a high risk passenger (i.e. their threat value distribution assigns large probabilities to higher values), then there is a high (or low, depending on the attack strategy) probability that the person standing right behind them in the line is also a threat to the safety of the airport and other passengers. In other words, the distribution generating the threat value of such a passenger has an affect on the threat value of the passengers standing close to them in the line. This effect, and the

underlying process governing the transition from one distribution to the other, can be captured by a Markov chain.

This paper is organized as follows. Section 2 provides an optimal assignment policy for LTSSAP with i.i.d. task values from a known observable distribution. The problem of unobservable task distributions, forming a hidden Markov chain, is addressed in Section 3. Section 4 presents concluding remarks and future research directions.

### 2. The model: observable task distributions

Consider the SSAP with  $n$  tasks and  $k$  (fixed) worker categories, where the  $i$ th worker-category consists of  $r_i$  workers each with value  $p_i$ . Let  $[.]$  denote the floor function where  $[y] := \max\{m \in \mathbb{Z} \mid m \leq y\}$ . Moreover, let  $\pi_i$  be the fraction of total number of workers that belong to categories  $i + 1$  to  $k$ , and hence,  $\alpha_i := \pi_{i-1} - \pi_i$  denotes the fraction of workers assigned to class  $i$ , for  $i = 1, 2, \dots, k$ . For simplicity, the  $i$ th worker category is referred to as type- $i$  workers (or equivalently, class- $i$  workers), with the size of the  $i$ th category given by  $r_i = [n\pi_{i-1}] - [n\pi_i]$  for  $i = 1, 2, \dots, k$ , where  $\pi_0 = 1, \pi_k = 0$ , and  $\pi_{i+1} < \pi_i$  for  $i = 0, 1, \dots, k - 1$ . Also, assume that  $p_{i+1} < p_i$  for  $i = 1, 2, \dots, k$ .

Let  $\mathcal{A}_i := (F^{-1}(\pi_i), F^{-1}(\pi_{i-1}))$  for  $i = 1, 2, \dots, k$ , where  $F$  is the distribution function of task values. A task  $X_j$  is labeled a type- $i$  task if  $X_j \in \mathcal{A}_i$ . Consider a policy  $\phi_L$  that assigns the  $j$ th task to a class- $i$  worker if  $X_j \in \mathcal{A}_i$ . If the goal is to maximize the expected reward per task as  $n \rightarrow +\infty$ , then it can be deduced from [4] that the optimal long-run expected reward per task is given by

$$r^* := \sum_{j=1}^k \left( p_j \int_{F^{-1}(\pi_j)}^{F^{-1}(\pi_{j-1})} x F(dx) \right), \tag{2.1}$$

with  $F^{-1}(1) = +\infty$  and  $F^{-1}(0) = -\infty$ . Moreover, [5] proved that  $\phi_L$  is the stationary policy that achieves the optimal long-run expected reward per task,  $r^*$ , for this SSAP. In what follows, we prove that this policy also optimizes the threshold probability for a given target value  $\tau$ ; specifically, it solves the LTSSAP and achieves the infimum in the following expression

$$\inf_{\phi \in \Phi} \mathbb{P} \left\{ \limsup_{n \rightarrow +\infty} \frac{1}{n} R_n^\phi \leq \tau \right\}, \tag{2.2}$$

where  $R_n^\phi$  is the total reward obtained after assigning all  $n$  tasks under policy  $\phi$ . Note that this policy consists of  $k - 1$  (time independent) fixed breakpoints, where  $k$  does not change as  $n \rightarrow +\infty$ , as opposed to the policy described by [6] in which the number of breakpoints increases with  $n$  and the breakpoints are recalculated at each time period.

For a fixed  $n$ , let  $U_{r_i}^{(n)}$  denote the number of tasks that arrive until  $r_i$  tasks of type- $i$  are obtained and define  $U^{(n)} := \min\{U_{r_1}^{(n)}, U_{r_2}^{(n)}, \dots, U_{r_k}^{(n)}\}$ . Observe that  $U_{r_i}^{(n)}$  follows a negative binomial distribution with parameters  $(r_i, \pi_{i-1} - \pi_i)$ , since tasks are assumed to be i.i.d. Proposition 2.1 presents a useful property of  $U^{(n)}$ .

**Proposition 2.1.** *It holds that*

$$\frac{U^{(n)}}{n} \rightarrow 1 \text{ almost surely as } n \rightarrow +\infty$$

*Proof.* Fix  $i \in \{1, 2, \dots, k\}$ , and recall that  $U_{r_i}^{(n)}$  is a negative binomial random variable with parameters  $(r_i, \pi_{i-1} - \pi_i)$ , and hence, it can be represented as the sum of  $r_i$  i.i.d.

geometric random variables each having mean  $1/(\pi_{i-1} - \pi_i)$ . The strong law of large numbers (SLLN) implies that  $(U_{r_i}^{(n)}/r_i) \rightarrow 1/(\pi_{i-1} - \pi_i)$  almost surely as  $n \rightarrow +\infty$ , which leads to  $(U_{r_i}^{(n)}/n) = (U_{r_i}^{(n)}/r_i)(r_i/n) \rightarrow 1$  as  $n \rightarrow +\infty$  almost surely. Recall that the minimum of any two arbitrary functions  $f$  and  $g$  can be represented as  $\min\{f, g\} = \frac{1}{2}(f + g - |f - g|)$ . Moreover,  $(U^{(n)}/n)$  is the minimum over a finite number of almost surely convergent functions, and hence,  $(U^{(n)}/n) \rightarrow 1$  as  $n \rightarrow +\infty$  almost surely, which completes the proof.

Applying the result in Proposition 2.1 we show that Theorem 2.1 proves the optimality of  $\phi_L$  for  $\tau \in [-\infty, r^*)$ .

**Theorem 2.1.** Assume that  $\tau < r^*$ . The infimum in (2.2) is achieved by a policy  $\phi_L$  that assigns the  $j$ th task to a type- $i$  worker if  $X_j \in \mathcal{A}_i$  and  $j \leq U^{(n)}$ .

*Proof.* To prove the result, the total reward under  $\phi_L$ ,  $R_n^{\phi_L}$ , is split into the reward obtained up to time  $U^{(n)}$  and the reward obtained after  $U^{(n)}$ , which are denoted by  $R_n^{(1)}$  and  $R_n^{(2)}$ , respectively. The superscript  $\phi_L$  is dropped to simplify the notation. Observe that

$$\frac{1}{n}R_n^{(1)} = \frac{1}{n} \sum_{i=1}^{U^{(n)}} \sum_{j=1}^k p_j X_i I_{\{X_i \in \mathcal{A}_j\}} = \frac{U^{(n)}}{n} \cdot \frac{1}{U^{(n)}} \sum_{i=1}^{U^{(n)}} \sum_{j=1}^k p_j X_i I_{\{X_i \in \mathcal{A}_j\}},$$

and note that

$$\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^k p_j X_i I_{\{X_i \in \mathcal{A}_j\}} \rightarrow r^* \quad \text{as } n \rightarrow +\infty, \tag{2.4}$$

almost surely by the SLLN. Moreover,  $U^{(n)} \geq \min\{r_1, r_2, \dots, r_k\}$ , and hence,  $U^{(n)} \rightarrow +\infty$  almost surely as  $n \rightarrow +\infty$ . This fact combined with (2.4) results in

$$\frac{1}{U^{(n)}} \sum_{i=1}^{U^{(n)}} \sum_{j=1}^k p_j X_i I_{\{X_i \in \mathcal{A}_j\}} \rightarrow r^* \quad \text{as } n \rightarrow +\infty \quad \text{almost surely.} \tag{2.5}$$

Therefore,  $(R_n^{(1)}/n \rightarrow r^*)$  almost surely as  $n \rightarrow +\infty$ , by (2.5) and Proposition 2.1. Moreover,

$$0 \leq \frac{1}{n}R_n^{(2)} \leq p_1 \left( \frac{1}{n} \sum_{i=1}^n X_i - \frac{1}{n} \sum_{i=1}^{U^{(n)}} X_i \right) \rightarrow 0 \quad \text{as } n \rightarrow +\infty,$$

almost surely, by a similar argument. Therefore,  $(R_n/n) \rightarrow r^*$  almost surely as  $n \rightarrow +\infty$  under  $\phi_L$ , and it follows that  $\mathbb{P}_{\phi_L} \{\limsup_{n \rightarrow +\infty} (R_n/n) \leq \tau\} = 0$  since  $\tau < r^*$ , completing the proof.

Note that Theorem 2.1 along with the results in [5] imply that the policy  $\phi_L$  achieves the maximum long-run expected reward per task, while minimizing the risk of the long-run reward per task failing to achieve a given target level. Theorem 2.2 proves a result which is useful in solving the LTSSAP for target values greater than  $r^*$ . Before proceeding to Theorem 2.2 we discuss the boundedness and integrability of task values in Assumption 2.1.

**Assumption 2.1.** There exists a random variable  $Y$  that is independent of the task values and  $\mathbb{P}\{X_j \leq Y\} = 1$  for  $j = 1, 2, \dots$ , where  $E[Y] < +\infty$ .

**Theorem 2.2.** Consider a policy  $\phi_B$  that assigns  $X_{(j)}$  (the  $j$ th order statistic of  $X_1, X_2, \dots, X_n$ ) to the  $j$ th best available worker. Under Assumption 2.1, it follows that

$$\frac{1}{n}R_n^{\phi_B} \rightarrow r^* \text{ almost surely as } n \rightarrow +\infty. \tag{2.6}$$

*Proof.* Note that the reward per task under  $\phi_B$  can be expressed as

$$\frac{1}{n}R_n^{\phi_B} = \sum_{j=1}^k p_j \left( \frac{1}{n} \sum_{i=[n\pi_j]+1}^{[n\pi_{j-1}]} X_{(i)} \right),$$

while  $r^* = \sum_{j=1}^k p_j E[X_1 I_{\{X_1 \in \mathcal{A}_j\}}]$ . Therefore, to prove (2.6), it is sufficient to prove that  $(1/n) \sum_{i=[n\pi]+1}^n X_{(i)} \rightarrow E[X_1 I_{\{X_1 \in \mathcal{A}_1\}}]$  as  $n \rightarrow +\infty$  almost surely, for arbitrarily fixed  $0 < \pi < 1$ . Observe that  $(1/n) \sum_{i=1}^n X_i I_{\{X_i \geq F^{-1}(\pi)\}} \rightarrow E[X_1 I_{\{X_1 \geq F^{-1}(\pi)\}}]$  as  $n \rightarrow +\infty$  almost surely by SLLN, and hence, it remains to show that

$$\frac{1}{n} \left| \sum_{i=[n\pi]+1}^n X_{(i)} - \sum_{j=1}^n X_j I_{\{X_j \geq F^{-1}(\pi)\}} \right| \rightarrow 0 \text{ as } n \rightarrow +\infty, \tag{2.7}$$

almost surely. Define  $N(n) := \sum_{i=1}^n I_{\{X_i \geq F^{-1}(\pi)\}}$ , and note that  $N(n)/n \rightarrow 1 - \pi$  almost surely as  $n \rightarrow +\infty$ . Consider the following two cases.

- (i) If  $N(n) \geq n - [n\pi]$ , then the left hand numerator in (2.7) contains  $N(n) - n + [n\pi]$  terms, all less than or equal to  $Y$ .
- (ii) If  $N(n) < n - [n\pi]$ , then the left hand numerator in (2.7) contains  $n - [n\pi] - N(n)$  terms, each less than  $F^{-1}(\pi)$ .

Therefore,

$$\frac{1}{n} \left| \sum_{i=[n\pi]+1}^n X_{(i)} - \sum_{j=1}^n X_j I_{\{X_j \geq F^{-1}(\pi)\}} \right| \leq \left| \frac{n - [n\pi] - N(n)}{n} \right| (Y \vee F^{-1}(\pi)) \rightarrow 0,$$

as  $n \rightarrow +\infty$  almost surely by SLLN and the fact that  $Y < +\infty$  with probability one. This completes the proof.

Using the result from Theorem 2.2, Corollary 2.1 solves (2.2) for  $\tau \in [r^*, +\infty)$ .

**Corollary 2.1.** If  $\tau \geq r^*$ , then

$$\inf_{\phi \in \Phi} \mathbb{P} \left\{ \limsup_{n \rightarrow +\infty} \frac{1}{n} R_n^\phi \leq \tau \right\} = 1. \tag{2.8}$$

*Proof.* Observe that  $\mathbb{P}_{\phi_B} \{ \limsup_{n \rightarrow +\infty} (R_n/n) \leq \tau \} = 1$ , by Theorem 2.2. Also

$$\mathbb{P}_{\phi_B} \left\{ \limsup_{n \rightarrow +\infty} \frac{1}{n} R_n \leq \tau \right\} \leq \mathbb{P}_\phi \left\{ \limsup_{n \rightarrow +\infty} \frac{1}{n} R_n \leq \tau \right\},$$

for any  $\phi \in \Phi$ , since no admissible policy can outperform  $\phi_B$ , and hence, (2.8) follows.

By Corollary 2.1, if the target value is greater than or equal to  $r^*$ , then all the admissible policies perform the same in terms of minimizing the threshold probability, and hence, the decision maker is indifferent in choosing between any two such policies. If the decision maker intends to optimize the assignments so as to achieve the maximum long-run expected reward per task, along with controlling risk level in the sense of (2.2), then they can opt to apply policy  $\phi_L$ .

### 3. The model: unobservable task distributions

Consider the LTSSAP with  $k$  worker categories, where the  $i$ th category consists of  $r_i = \lceil n\pi_{i-1} \rceil - \lceil n\pi_i \rceil$  workers, each with value  $p_i$ . Assume that the task values are generated from  $r$  different distributions, where the successive distributions are governed by an irreducible ergodic time-homogeneous Markov chain with (known) transition probability matrix  $Q = (q_{ij})$  and an invariant (stationary) distribution  $\mu$ . The state of the Markov chain at time period  $j$  is denoted by  $Z_j$ , with  $\mathcal{S} = \{1, 2, \dots, r\}$  being the state space of the Markov chain. Specifically,  $Z_j = k$  means that the  $j$ th task  $X_j$  is a random variable with distribution function  $F_k$  having support  $\mathcal{B} \subseteq [0, +\infty)$ . Upon the arrival of each task, its value is observed; however, the state of the Markov chain (and hence, the distribution associated with the task value) is unobservable. The goal is to arrive at an assignment policy that minimizes the threshold probability in (2.2).

Since task values are derived from  $r$  distinct distributions that are linked together through a Markov chain, it follows that task values are no longer i.i.d., and hence, the approach presented in Section 2 cannot be used. To solve this problem, let  $W = \{W_j, j = 1, 2, \dots\}$  be a discrete-time Markov chain with state space  $\mathcal{S} \times \mathcal{B}$  where  $W_j := (Z_j, X_j)$ , and note that only  $X_j$  is observable at time period  $j$ . In the following, the chain  $W$  is proven to be positive recurrent, implying that the strong law of large numbers holds for  $W$ . Then, it is shown that a stationary policy similar to  $\phi_L$  achieves the infimum in (2.2) and is optimal. Lemma 3.1 provides an invariant distribution  $\bar{\mu}$  for  $W$ .

**Lemma 3.1.** *The chain  $W$  admits an invariant measure  $\bar{\mu}$ , where*

$$\bar{\mu}(k, E) := \mu(k)F_k(E) \quad \text{for any } k \in \mathcal{S} \quad \text{and} \quad E \subseteq \mathcal{B}. \tag{3.1}$$

*Proof.* Let  $\mathbb{P}\{(l, x), (k, E)\}$  denote the probability of transitioning from state  $(l, x)$  to state  $(k, E)$ , where  $(l, x) \in \mathcal{S} \times \mathcal{B}$ . To prove the result we need to verify that for any  $k \in \mathcal{S}$  and  $E \subseteq \mathcal{B}$

$$\bar{\mu}(k, E) = \int_{\mathcal{B}} \sum_{l=1}^r \mathbb{P}\{(l, x), (k, E)\} \bar{\mu}(l, dx), \tag{3.2}$$

where  $\bar{\mu}$  is given by (3.1). To do so, (3.1) is substituted into the right hand side of (3.2) as follows

$$\begin{aligned} \int_{\mathcal{B}} \sum_{l=1}^r \bar{\mu}(l, dx) \mathbb{P}\{(l, x), (k, E)\} &= \sum_{l=1}^r \left[ \mu(l) \left( \int_{\mathcal{B}} F_l(dx) \right) q_{lk} F_k(E) \right] \\ &= F_k(E) \sum_{l=1}^r \mu(l) q_{lk} \\ &= \bar{\mu}(k, E) \end{aligned}$$

where the last equality follows from the fact that  $\mu$  is an invariant measure for  $Z := \{Z_j, j = 1, 2, \dots\}$ , and (3.2) is verified, completing the proof.

Before we proceed to Corollary 3.2 some notation and definitions must be introduced, and auxiliary results must be presented. In Corollary 3.2 we show that the SLLN holds for the chain  $W$  (and hence, simplifies the proof of the desired results for the unobservable distributions case).

**Definition 3.1.** Let  $\tau_{(k,E)} := \min\{j \geq 1 : W_j \in (k, E)\}$  denote the first hitting of the set  $(k, E)$ , and define

$$L((l, x), (k, E)) := \mathbb{P}_{(l,x)}\{\tau_{(k,E)} < +\infty\} = \mathbb{P}\{W \text{ ever enters } (k, E), \text{ starting from } (l, x)\},$$

for  $(l, x) \in \mathcal{S} \times \mathcal{B}$  and  $(k, E) \subset \mathcal{S} \times \mathcal{B}$ .

**Definition 3.2.** The chain  $W$  is  $\psi$ -irreducible, see [7, Chapter 4, page 89], if there exists a measure  $\psi$  such that if  $\psi(k, E) > 0$ , then  $L((l, x), (k, E)) > 0$  for any state  $(l, x) \in \mathcal{S} \times \mathcal{B}$ .

**Lemma 3.2.** Fix an arbitrary state  $s_0 \in \mathcal{S}$ , and define  $\psi(k, E) := \delta_{s_0}(k)F_{s_0}(E)$  for all  $k \in \mathcal{S}$  and  $E \subseteq \mathcal{B}$ , where  $\delta_{s_0}(k) := I_{\{k=s_0\}}$ . The chain  $W$  is  $\psi$ -irreducible.

*Proof.* Fix  $k \in \mathcal{S}$  and  $E \subseteq \mathcal{B}$  such that  $\psi(k, E) > 0$ , and note that  $k = s_0$  and  $F_k(E) > 0$ . Since  $Z$  is irreducible, there exists  $n \geq 1$  such that  $Q_{lk}^{(n)} := \mathbb{P}\{Z_{t+n} = k \mid Z_t = l\} > 0$ , for any  $l \in \mathcal{S}$ . Therefore,  $\mathbb{P}_{(l,x)}^{(n)}\{W \in (k, E)\} = Q_{lk}^{(n)}F_k(E) > 0$ , for  $x \in \mathcal{B}$ , and hence,  $L((l, x), (k, E)) > 0$ , which proves that  $W$  is  $\psi$ -irreducible.

**Assumption 3.1.** There exists a small set  $(\bar{A}, \bar{E}) \subset \mathcal{S} \times \mathcal{B}$ , see [7, Chapter 5, page 109], such that  $L((l, x), (\bar{A}, \bar{E})) = 1$  for all  $(l, x) \in \mathcal{S} \times \mathcal{B}$ .

**Corollary 3.1.** The chain  $W$  is positive Harris recurrent under Assumption 3.1, see [7, Chapter 9, page 204, and Chapter 10, page 235].

*Proof.*  $W$  is a positive chain since it is  $\psi$ -irreducible and admits an invariant probability measure  $\bar{\mu}$ . Moreover, Assumption 3.1 along with  $\psi$ -irreducibility of  $W$  implies that the chain is Harris recurrent (Proposition 9.1.7 in [7]).

Corollary 3.2 follows from Theorem 17.0.1 in [7] and proves that the SLLN holds true for chain  $W$  under Assumption 3.1.

**Corollary 3.2.** For any function  $g$  defined on  $\mathcal{S} \times \mathcal{B}$ ,

$$\frac{1}{n} \sum_{j=1}^n g(W_j) \rightarrow \bar{\mu}(g) \text{ almost surely as } n \rightarrow +\infty, \text{ if } g \text{ satisfies } \bar{\mu}(|g|) < +\infty.$$

To define the optimal assignment policy, we first introduce some notation. Define  $F : \mathcal{B} \rightarrow [0, 1]$  as  $F(a) := \sum_{j=1}^r \mu(j)F_j(a)$ , and note that  $F$  is a distribution function on  $\mathcal{B}$  with  $F^{(-1)}(1) := +\infty$  and  $F^{(-1)}(0) := -\infty$ . Let  $r^*$  be defined as in (2.1), with  $F$  being the distribution function introduced above. As defined in Section 2, a task  $X_j$  is labeled type- $i$  if  $X_j \in \mathcal{A}_i := (F^{-1}(\pi_i), F^{-1}(\pi_{i-1})]$ . Moreover, for a fixed  $n$ , let  $t_{r_i}^{(n)}$  denote the number of tasks that must arrive until  $r_i$  tasks of type  $i$  are obtained and define  $t^{(n)} := \min\{t_{r_1}^{(n)}, t_{r_2}^{(n)}, \dots, t_{r_k}^{(n)}\}$ . Note that unlike  $U_{r_i}^{(n)}$ ,  $t_{r_i}^{(n)}$  is not distributed negative binomial, since the task values are no longer assumed to be i.i.d. Unless otherwise mentioned, Assumption 3.1 holds throughout Section 3. Proposition 3.1 discusses a useful property of  $t^{(n)}$ .

**Proposition 3.1.** It holds that

$$\frac{t^{(n)}}{n} \rightarrow 1 \text{ almost surely as } n \rightarrow +\infty$$

*Proof.* It suffices to prove that for all  $i \in \{1, 2, \dots, k\}$ ,  $(t_{r_i}^{(n)}/n) \rightarrow 1$  almost surely as  $n \rightarrow +\infty$ . To this end, arbitrarily fix  $i \in \{1, 2, \dots, k\}$ , and note that  $(r_i/n) \rightarrow \pi_{i-1} - \pi_i$  as  $n \rightarrow +\infty$ , and hence, it is enough to prove that

$$\frac{t_{r_i}^{(n)}}{r_i} \rightarrow \frac{1}{\pi_{i-1} - \pi_i} \quad \text{as } n \rightarrow +\infty,$$

almost surely. Define  $\mathcal{B}_i := \{(Z_j, X_j) : X_j \in \mathcal{A}_i\}$ , and note that  $t_{r_i}^{(n)}$  is the random time of the  $r_i$ th visit to  $\mathcal{B}_i$  for the chain  $W$ . On the other hand, since

$$\bar{\mu}(I_{\mathcal{B}_i}) = \sum_{u=1}^r \mu(u)(F_u(F^{-1}(\pi_{i-1})) - F_u(F^{-1}(\pi_i))) = \pi_{i-1} - \pi_i,$$

it follows from Corollary 3.2 that

$$\frac{1}{n} \sum_{j=1}^n I_{\mathcal{B}_i}(W_j) \rightarrow \pi_{i-1} - \pi_i \quad \text{as } n \rightarrow +\infty \quad \text{almost surely.}$$

Moreover,  $t_{r_i}^{(n)} \geq r_i$ , and hence,  $t_{r_i}^{(n)} \rightarrow +\infty$  almost surely as  $n \rightarrow +\infty$ . Therefore

$$\frac{1}{t_{r_i}^{(n)}} \sum_{j=1}^{t_{r_i}^{(n)}} I_{\mathcal{B}_i}(W_j) \rightarrow \pi_{i-1} - \pi_i \quad \text{as } n \rightarrow +\infty,$$

almost surely, but  $\sum_{j=1}^{t_{r_i}^{(n)}} I_{\mathcal{B}_i}(W_j) = r_i$ , implying that

$$\frac{r_i}{t_{r_i}^{(n)}} \rightarrow \pi_{i-1} - \pi_i \quad \text{as } n \rightarrow +\infty,$$

almost surely, which completes the proof.

Applying the result in Proposition 3.1 we will show that Theorem 3.1 presents the optimal policy for  $\tau \in [-\infty, r^*)$ .

**Theorem 3.1.** *Assume that  $\tau < r^*$ . A policy  $\tilde{\phi}_L$  that assigns the  $j$ th task to a type- $i$  worker if  $X_j \in \mathcal{A}_i$  and  $j \leq t^{(n)}$ , achieves the infimum in (2.2).*

*Proof.* The proof proceeds along the same lines as that of Theorem 2.1 and follows from Corollary 3.2 and Proposition 3.1.

As in the observable distributions case, Theorem 3.1 along with the results in [5], imply that the policy  $\tilde{\phi}_L$  achieves the maximum long-run expected reward per task, while minimizing the risk of the long-run reward per task failing to achieve a given target level. Lemma 3.3 presents a result, which helps with solving the problem for target values greater than or equal to  $r^*$ .

**Lemma 3.3.** *Let  $X_{(j)}$  denote the  $j$ th order statistic of tasks  $X_1, X_2, \dots, X_n$ , coming from  $r$  different distributions  $\{F_1, F_2, \dots, F_r\}$ , where the successive distributions are unobservable and governed by an irreducible ergodic Markov chain with invariant distribution  $\mu$ . Thus,*

$$\frac{1}{n} \sum_{j=[n\pi]+1}^n X_{(j)} \rightarrow \int_{F^{-1}(\pi)}^{+\infty} x F(dx) \quad \text{as } n \rightarrow +\infty,$$

almost surely for any  $\pi \in (0, 1)$ , where  $F(a) := \sum_{j=1}^r \mu(j)F_j(a)$ .

*Proof.* Taking Corollary 3.2 into consideration, the proof is similar to that of Theorem 2.2.

Using the result in Lemma 3.3, Theorem 3.2 solves (2.2) and addresses the problem for target values greater than or equal to  $r^*$ .

**Theorem 3.2.** *If  $\tau \geq r^*$ , then  $\inf_{\phi \in \Phi} \mathbb{P}\{\limsup_{n \rightarrow +\infty} R_n^\phi/n \leq \tau\} = 1$ , under Assumption 2.1.*

*Proof.* In light of Corollary 3.2, Proposition 3.1, and Lemma 3.3, the proof is analogous to that of Theorem 2.2 and Corollary 2.1.

According to Theorem 3.2, and similar to the observable distributions case, the decision maker is indifferent in the choice between any two arbitrary admissible policies, when minimizing the threshold probability with  $\tau \in [r^*, +\infty)$ . A prudent choice is to apply the policy  $\tilde{\phi}_L$  in this case.

Note that the optimal policy for LTSSAP (i.e.  $\phi_L$  for i.i.d. task values and  $\tilde{\phi}_L$  for tasks with unobservable distributions) assigns the best  $\alpha_1$  percent of the tasks to the best workers, the second best  $\alpha_2$  percent of the tasks to the second best workers, and so on. Moreover, [5] proved that the same policy maximizes the long-run expected reward per task. A natural question to ask is whether there exist other classes of objective functions, for which this policy is optimal. To answer this question, Theorem 3.3 analyzes SSAP under the following objective function

$$\inf_{\phi \in \Phi} \limsup_{n \rightarrow +\infty} \mathbb{P}\left\{\frac{1}{n}R_n^\phi \leq \tau\right\}, \tag{3.4}$$

and verifies that  $\phi_L$  and  $\tilde{\phi}_L$  optimize (3.4) for  $\tau \in (-\infty, r^*)$ .

**Theorem 3.3.** *If  $\tau < r^*$ , then  $\phi_L$  and  $\tilde{\phi}_L$  achieve the optimality in (3.4) for the observable and the unobservable distributions case, respectively.*

*Proof.* Recall from Theorem 2.1 and Theorem 3.1 that the reward per task under  $\phi_L$  and  $\tilde{\phi}_L$  converges to  $r^*$  almost surely as  $n \rightarrow +\infty$ , which implies convergence in probability. Therefore,

$$\mathbb{P}\left\{\frac{1}{n}R_n^{\phi_L} \leq \tau\right\} \leq \mathbb{P}\left\{\left|\frac{1}{n}R_n^{\phi_L} - r^*\right| > r^* - \tau\right\} \rightarrow 0 \quad \text{as } n \rightarrow +\infty,$$

implying  $\lim_{n \rightarrow +\infty} \mathbb{P}\{R_n^{\phi_L}/n \leq \tau\} = 0$ . Therefore,  $\phi_L$  achieves the infimum in (3.4). A similar argument proves the optimality of  $\tilde{\phi}_L$  for the unobservable distributions case.

Theorem 3.4 proves that if the given target value is greater than  $r^*$ , then the decision maker has no preference in choosing between the set of admissible policies, under the objective function introduced in (3.4). Recall that this is also the case when solving the LTSSAP.

**Theorem 3.4.** *If  $\tau > r^*$ , then  $\inf_{\phi \in \Phi} \limsup_{n \rightarrow +\infty} \mathbb{P}\{R_n^\phi/n \leq \tau\} = 1$ , under Assumption 2.1, for both the observable and the unobservable distributions cases.*

*Proof.* Recall from Theorem 2.2 that the long-run reward per task under  $\phi_B$  equals  $r^*$  almost surely. On the other hand, no policy can do better than the infeasible policy  $\phi_B$ . Therefore,

$$\limsup_{n \rightarrow +\infty} \frac{1}{n}R_n^\phi \leq \lim_{n \rightarrow +\infty} \frac{1}{n}R_n^{\phi_B} = r^* < \tau \tag{3.5}$$

almost surely for any arbitrarily fixed policy  $\phi \in \Phi$ . It follows from (3.5) and the definition of limsup that  $\mathbb{P}\{\liminf_{n \rightarrow +\infty} \{R_n^\phi/n \leq \tau\}\} = 1$ , which implies that

$$1 = \limsup_{n \rightarrow +\infty} \mathbb{P}\left\{\frac{1}{n}R_n^\phi \leq \tau\right\} \geq \mathbb{P}\left\{\liminf_{n \rightarrow +\infty} \left\{\frac{1}{n}R_n^\phi \leq \tau\right\}\right\} \quad \text{for all } \phi \in \Phi,$$

and the claim is proven for the observable distributions case. The same argument proves the result for the unobservable distributions case.

#### 4. Conclusion

This paper studies the SSAP with  $k$  fixed worker classes as the number of tasks (denoted by  $n$ ) approaches  $\infty$ , where the aim is to minimize the threshold probability. This objective function diverges from the risk-neutral objective function studied in the existing SSAP literature. Two versions of the problem are studied that are based on the distribution function of task values being either observable or unobservable. Simple stationary optimal policies are presented for both problems, as opposed to the time-dependent policy described in [6]. These stationary optimal policies not only maximize the long-run expected reward per task but also minimize the risk of failing to achieve a given threshold value.

This paper incorporates a Markov-modulated dependency that is simultaneous between the task values and the uncertainty in the task distribution functions of the model studied. Further research is required to address other types of uncertainty in the SSAP where there are a large number of tasks. One example can be the case where the number of arriving tasks is not known *a priori* and follows a given probability distribution. Another challenge is shifting the attention from the i.i.d. sequence of tasks to a more general case with dependent task values, where the value of the current task depends on the whole sequence of its preceding tasks.

#### Acknowledgements

This research has been supported in part by the Air Force Office of Scientific Research under Grant No. FA9550-10-1-0387, and the National Science Foundation under Grant No. CMMI-0900226. This material is based upon work supported in part by the second author while serving at the National Science Foundation. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the United States Government, the Air Force Office of Scientific Research, or the National Science Foundation.

#### References

- [1] ALBRIGHT, S. C. (1974). A Markov-decision-chain approach to a stochastic assignment problem. *Operat. Res.* **22**, 61–64.
- [2] ALBRIGHT, S. C. (1976). A Markov chain version of the secretary problem. *Naval Res. Logistics Quart.* **23**, 151–159.
- [3] ALBRIGHT, S. C. (1977). A Bayesian approach to a generalized house selling problem. *Manag. Sci.* **24**, 432–440.
- [4] ALBRIGHT, S. C. AND DERMAN, C. (1972). Asymptotic optimal policies for the stochastic sequential assignment problem. *Manag. Sci.* **19**, 46–51.
- [5] BAHARIAN, G. AND JACOBSON, S. H. (2013). Limiting behavior of the stochastic sequential assignment problem. *Naval Res. Logistics* **60**, 321–330.
- [6] DERMAN, C., LIEBERMAN, G. J. AND ROSS, S. M. (1972). A sequential stochastic assignment problem. *Manag. Sci.* **18**, 349–355.
- [7] MEYN, S. P. AND TWEEDIE, R. L. (1993). *Markov Chains and Stochastic Stability*. Springer, London.
- [8] NAKAI, T. (1986). A sequential stochastic assignment problem in a partially observable Markov chain. *Math. Operat. Res.* **11**, 230–240.
- [9] NIKOLAEV, A. G. AND JACOBSON, S. H. (2010). Stochastic sequential decision-making with a random number of jobs. *Operat. Res.* **58**, 1023–1027.
- [10] NIKOLAEV, A. G., JACOBSON, S. H. AND MCLAY, L. A. (2007). A sequential stochastic security system design problem for aviation security. *Transportation Sci.* **41**, 182–194.
- [11] SU, X. AND ZENIOS, S. A. (2005). Patient choice in kidney allocation: a sequential stochastic assignment model. *Operat. Res.* **53**, 443–455.