A Stochastic Assignment Problem

David T. Wu, Sheldon M. Ross

Department of Industrial and Systems Engineering, University of Southern California, Los Angeles, California

Received 4 February 2014; revised 14 November 2014; accepted 18 November 2014
DOI 10.1002/nav.21611
Published online 11 December 2014 in Wiley Online Library (wileyonlinelibrary.com).

Abstract: There are n boxes with box i having a quota value $m_i, i = 1 \ldots n$. Balls arrive sequentially, with each ball having a binary vector $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ attached to it, with the interpretation being that if $X_i = 1$ then that ball is eligible to be put in box i. A ball's vector is revealed when it arrives and the ball can be put in any alive box for which it is eligible, where a box is said to be alive if it has not yet met its quota. Assuming that the components of a vector are independent, we are interested in the policy that minimizes, either stochastically or in expectation, the number of balls that need arrive until all boxes have met their quotas.

Keywords: sequential assignment problem; dynamic programming; stochastic optimization; stochastic process modeling; probability; online optimization

1. INTRODUCTION

We study a stochastic assignment problem in which one assigns balls to boxes. Balls arrive sequentially; each ball has a binary vector $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ attached to it, with the interpretation that if $X_i = 1$ the ball is eligible to be put in box $i, i = 1 \ldots n$. A ball's vector is revealed when it arrives; after observing the vector one decides which box the ball is to be assigned to. A ball that is ineligible for all currently alive boxes is discarded, where a box is said to be alive if it has not yet met its quota. This process continues until there are m_i balls in box i for each $i = 1 \ldots n$. Let N denote the total number of balls that have arrived. We assume that successive eligibility vectors are independent and identically distributed (iid) and that the components of the eligibility vector are independent Bernoulli random variables. We are interested in the policy that, either stochastically or in expectation, minimizes N.

This model is an extension of [16], which assumed that all $m_i = 1$, and showed that N is stochastically minimized by the policy which puts an arriving ball in the alive and eligible box i having the smallest value of $p_i \equiv P(X_i = 1)$. Although the model in this article could be interpreted as one in which there are $t = \sum_{i=1}^{n} m_i$ boxes; each box needing a single ball and with each ball having a vector $\mathbf{X} = (X_1, X_2, \ldots, X_t)$ such that the first m_1 components of $\mathbf{X}$ are equal, the next m_2 components are equal, and so on, under this interpretation the X_i are no longer independent and so the policy proposed in [16] need not be optimal.

The model of this article can be applied to an organizational employment problem, where the organization is putting together a team that requires m_i workers for job type i, for $i = 1, \ldots, n$. Potential team members arrive sequentially; upon arrival, the job types for which the candidate is qualified are determined, and the candidate is then either rejected or assigned to one of the not yet filled job types. Assuming that work cannot begin until the team is complete, the problem is to minimize the expected time to get there. By interpreting a box as a job type, a box quota as the number of required workers of that job type, and a ball vector as indicating which jobs the candidate is qualified for, we see that this application fits our model.

The model of this article is related to the generalized sequential assignment model studied in [10]. The model of [10] assumes that sequentially arriving job candidates are to be assigned to n jobs. Each candidate is assumed to have a vector $(X_1, \ldots, X_n)$, with the interpretation that X_i is earned if that seeker is assigned to job $i, i = 1, \ldots, n$. The candidate can then either be rejected or assigned to a not yet assigned job. Assuming that the vector $(X_1, \ldots, X_n)$ has an arbitrary distribution, and that a cost per job applicant is incurred, a computational approach was given to find the optimal policy when n is of moderate size.

The paper [10] was a generalization of the classical stochastic sequential assignment problem introduced in [9] (denoted as DLR), which assumed that each of the n jobs had a given value and that each arriving candidate had a skill

Correspondence to: David T. Wu (tengwu@usc.edu)

index which is the value of a random variable with fixed distribution. When a candidate is assigned to a job, a reward equal to the product of the candidate's skill index and the job's value is earned. Assuming that each job can be assigned once and each arriving worker must be assigned to some job, the paper [8] explicitly derived the optimal policy. For applications and variations of the DLR model, see for instance [1, 4, 6, 8, 12, 15, 17, 18, 20].

Our model can also be applied to channel scheduling in telecommunication systems. Assume in a time division multiple access network that n users share one channel, and that user i initially has m_i packets to transmit, $i = 1 \dots n$. In each time slot, each user is independently either connected or not, with user i being connected with probability $p_i, i = 1, \dots, n$. Although multiple users can be connected, only one is allowed to transmit at a time, with a central controller making the assignment decisions. Under the objective of minimizing, the mean time until all users have transmitted all their packets, it follows that this model is equivalent to our problem on interpreting the boxes as the users and the successive ball vectors as the successive indicators of which users are connected. This channel scheduling problem has been studied in [19] where it was shown that when all $p_i = p$ the policy π_o that assigns to any connected user with the largest number of packets waiting, is optimal. Moreover, it was also shown in [19] that if new packets arrive to all the users according to iid Bernoulli processes, then π_o minimizes the long-term average number of packets in the system. The paper [11] generalized the model by allowing more than one user to transmit in each time slot, and showed that π_o remained the optimal policy for a variety of optimality criteria, even when the assumption of iid vectors was replaced by the weaker assumption of exchangeability. Both [19, 11] made strong use of their symmetry assumptions, and neither considered the case where the p_i need not be equal.

In Section 2, we restrict attention to priority policies, which are policies that are determined by a permutation of $1, \dots, n$, with the interpretation that an arriving ball is assigned to the first alive and eligible box in this permutation. We first derive the distribution of N when a priority policy is used. We then prove that if all quotas are equal, the policy that assigns box priorities in increasing order of the box probabilities p_i stochastically minimizes N among all priority policies; we also show that if all p_i are equal, the policy that assigns box priorities in decreasing order of their quotas, stochastically minimizes N among all priority policies.

In Section 3, we consider dynamic assignment policies and show that in searching for a policy that minimizes $E[N]$, one never needs to consider putting a ball in box i if there is another eligible box j such that $p_j < p_i$ and $r_j \geq r_i$ where r_i and r_j are the remaining quotas of boxes i and j. In addition, we introduce a heuristic policy which puts an arriving ball in an alive and eligible box which, among all such boxes j, maximizes $\frac{r_j}{p_j}$. Starting with this heuristic policy, we then improve it using the policy improvement algorithm of dynamic programming. To benchmark the performance of our policies, a lower bound on the optimal value of $E[N]$ is determined. When n and the quota values are not too large, an optimal policy can be computed, thus enabling us to assess the performance of the heuristic policies.

In Section 4, we consider the special case where there are only two boxes. Using the idea of a one step look ahead policy, we explicitly derive the optimal policy when $m_1 = 1$. We also prove a structural result about the optimal policy when $m_1 = 2$, and conjecture that this structural result holds for general m_1.

2. STATIC PRIORITY POLICIES

A static priority policy $j_1, \dots, j_n$ is a permutation of $1 \dots n$ with the instruction that an arriving ball having vector values $(x_1, \dots, x_n)$ should be put in box j_i if i is the smallest integer for which box j_i is alive and $x_{j_i} = 1$. In other words, such a policy initially chooses a priority ordering and then follows that ordering (ignoring earlier actions) throughout in making its choices. Suppose the static priority policy $1, 2, \dots, n$ is to be used. Let $s_j = \sum_{i=1}^{j} m_i, j = 1, \dots, n$. Let Y_1 be the number of balls needed to obtain m_1 that are eligible for box 1; ignoring the balls that are put in box 1, let Y_2 be the number needed to obtain m_2 balls that are eligible for box 2; ignoring the balls that go either in box 1 or box 2, let Y_3 be the number of balls needed to obtain m_3 balls that are eligible for box 3; and so on. (For instance, if $m_1 = m_2 = 3$ and balls 1, 6, 8 are put in box 1, and 4, 5, 7 are put in box 2, then $Y_1 = 8$ and $Y_2 = 5$.) Because the balls that arrive before box 1 is filled that are not put in that box will, by the independence of X_1 and X_2, be eligible for box 2 with probability p_2, it follows that Y_2 is independent of Y_1 and is a negative binomial random variable with parameters (m_2, p_2). Indeed, the independence of the components of an eligibility vector implies that $Y_i, i = 1, \dots, n$, are independent negative binomial random variables with parameters $(m_i, p_i), i = 1, \dots, n$. Let N_k be the number of balls needed to fill all the boxes $1, \dots, k$. Then, with $N_0 = s_0 = 0$, we have

$$N_j = N_{j-1} + (Y_j - N_{j-1} + s_{j-1})^+, j = 1, \dots, n. \quad (1)$$

The preceding follows because if N_{j-1} is the number of balls needed to fill all of the boxes $1, \dots, j-1$, then $N_{j-1} - s_{j-1}$ of these balls would be put in box j if they were eligible for box j and box j were alive upon their arrival. Consequently, the additional number of balls after the first N_{j-1} that are needed for box j to meet its quota is $(Y_j - (N_{j-1} - s_{j-1}))^+$.

Let

$$\bar{B}_{t,q}(k) = P(\mathrm{Bin}(t,q) \geq k)$$

where $\mathrm{Bin}(t,q)$ is a binomial random variable with parameters (t,q).

PROPOSITION 1: For $r \geq s_n$,

$$P(N_n \leq r) = \prod_{i=1}^{n} \bar{B}_{r-s_{i-1},p_i}(m_i)$$

PROOF: With $N_0 = 0$, and with N_i denoting the number of balls until all boxes $1,\ldots,i$ are filled, we have

$$\begin{aligned} P(N_i \leq r) &= P(N_i \leq r | N_{i-1} \leq r)P(N_{i-1} \leq r) \\ &= \bar{B}_{r-s_{i-1},p_i}(m_i)P(N_{i-1} \leq r) \end{aligned}$$

where the preceding follows because given that all boxes $1,\ldots,i-1$ are filled by the time r balls are observed (where we suppose that we continue to observe balls even after all boxes are filled), each of the $r - s_{i-1}$ balls that have not gone into any of the boxes $1,\ldots,i-1$ will, independently, be eligible for box i with probability p_i. Iterating the preceding proves the proposition.

We now obtain the optimal priority ordering policy in two cases: the first where all box quotas are equal, and the second where all box eligibility probabilities are equal. □

THEOREM 2: Suppose $m_i = m$ for all $i = 1,\ldots,n$. If $(i_1,\ldots,i_n)$ is a permutation of $(1,\ldots,n)$ such that $p_{i_j} \leq p_{i_{j+1}}$, then the static priority policy $(i_1,\ldots,i_n)$ stochastically minimizes, among all static priority policies, the number of balls needed to meet all quotas.

PROOF: We will prove this result by showing that, starting with any static priority policy, any pairwise interchange of adjacent elements that moves a box having a smaller probability in front of one with a larger probability will result in a stochastically smaller number of balls needed to fill all boxes. Continually repeating this operation proves the theorem. So, let N and N^* be the number of balls needed when using, respectively, the priority policies $(1,\ldots,j-1,j,j+1,\ldots,n)$ and $(1,\ldots,j-1,j+1,j,\ldots,n)$. Assume that $p_j < p_{j+1}$. Now, from Proposition 1, the result will follow if we can show that

$$\begin{aligned} &\bar{B}_{r-s_{j-1},p_j}(m)\bar{B}_{r-s_{j-1}-m,p_{j+1}}(m) \\ &\quad \geq \bar{B}_{r-s_{j-1},p_{j+1}}(m)\bar{B}_{r-s_{j-1}-m,p_j}(m) \end{aligned}$$

But the preceding is equivalent to stating that in the two box problem with equal box quotas, where balls are eligible for box 1 with probability p_j and to box 2 with probability p_{j+1}, that the probability the number of balls needed to fill both boxes is at most $r - s_{j-1}$ is at least as large under policy (1, 2) as it is under policy (2, 1). Consequently, it suffices to prove the theorem when there are only two boxes. So suppose $n = 2$ and that $p_1 < p_2$. Let N be the number of balls needed to fill both boxes if the policy (1, 2) is employed and let N^* be the number if policy (2, 1) is employed. Let $Y_i, i = 1,2$ be independent negative binomial random variables with respective parameters $(m, p_i), i = 1,2$, and note from (1) that

$$\begin{aligned} N &=_{st} Y_1 + (Y_2 - Y_1 + m)^+ \\ N^* &=_{st} Y_2 + (Y_1 - Y_2 + m)^+ \end{aligned}$$

where $U =_{st} V$ means that U and V have the same distribution. Now, let

$$h(x,y) = x + (y - x + m)^+$$

and note that if $Y_1 = x, Y_2 = y$ then $Y_1 + (Y_2 - Y_1 + m)^+ = h(x,y)$ and $Y_2 + (Y_1 - Y_2 + m)^+ = h(y,x)$. Also, note that $x > y$ implies that

$$h(x,y) = x + (y - x + m)^+ \leq x + m = h(y,x).$$

However, as shown in [5], any function h for which $x > y$ implies $h(x,y) \leq h(y,x)$, has the property that $h(X, Y)$ is stochastically smaller than $h(Y,X)$ when X and Y are independent and X is likelihood ratio larger than Y. (This is proven by conditioning on the pair of values $\min(X,Y)$ and $\max(X,Y)$.) Because it is easy to check that a negative binomial with parameters (m, p) increases in likelihood ratio as p decreases, it follows that Y_1 is likelihood ratio larger than Y_2, giving that

$$N =_{st} Y_1 + (Y_2 - Y_1 + m)^+ \leq_{st} Y_2 + (Y_1 - Y_2 + m)^+ =_{st} N^*.$$

□

THEOREM 3: If $p_i = p, i = 1,\ldots,n$, then the static priority policy that orders the boxes in decreasing order of their quotas stochastically minimizes, among all static priority policies, the number of balls needed to meet all quotas.

PROOF: As in the proof of Theorem 2, it suffices to prove the result when $n = 2$. So, consider two scenarios; the first in which the quotas for boxes one and two are, respectively, i and j and the second where they are j and i. Suppose $i > j > 0$ and that policy (1, 2) is to be used in both scenarios. We will prove the theorem by using a coupling argument to show that $N(1)$, the number of balls needed in the first scenario, is stochastically smaller than $N(2)$, the number needed in the second scenario. At any point of time, let $\mathbf{s} = (s_1, s_2)$ be the remaining quotas of the boxes in the first scenario, and let $\mathbf{t} = (t_1, t_2)$ be the remaining quotas of the boxes in the second scenario. Say that $\mathbf{s}$ dominates $\mathbf{t}$ if

$$s_1 > t_1, \quad s_2 \geq t_1, \quad s_1 + s_2 = t_1 + t_2, \quad \min(s_1, s_2, t_1, t_2) > 0.$$

Note that, because $i > j > 0$, the initial vector $\mathbf{s} = (i, j)$ dominates the initial vector $\mathbf{t} = (j, i)$. If $\mathbf{s}$ and $\mathbf{t}$, the current remaining quota vectors for the two scenarios, are such that $\mathbf{s}$ dominates $\mathbf{t}$ then couple (X_1, X_2) and (X_1^*, X_2^*), the eligibility vectors for the next ball in the two scenarios, so that $X_1^* = X_2, X_2^* = X_1$. (Because $p_1 = p_2$, (X_1^*, X_2^*) will have the distribution of an eligibility vector.) The remaining quotas in the two scenarios (denoted by primes), as a function of (X_1, X_2) are as follows:

$$\begin{aligned}
&(X_1, X_2) \\
&\quad = (1,1) \Rightarrow s_1' = s_1 - 1, s_2' = s_2, t_1' = t_1 - 1, t_2' = t_2 \\
&(X_1, X_2) \\
&\quad = (1,0) \Rightarrow s_1' = s_1 - 1, s_2' = s_2, t_1' = t_1, t_2' = t_2 - 1 \\
&(X_1, X_2) \\
&\quad = (0,1) \Rightarrow s_1' = s_1, s_2' = s_2 - 1, t_1' = t_1 - 1, t_2' = t_2 \\
&(X_1, X_2) \\
&\quad = (0,0) \Rightarrow s_1' = s_1, s_2' = s_2, t_1' = t_1, t_2' = t_2.
\end{aligned}$$

It follows from the preceding that either s' dominates t' or one of the following is true: (a) $s'_1 = t'_1$ or (b) $s'_1 > t'_1 = 0$. If (a) occurs then because $s'_1 + s'_2 = t'_1 + t'_2$, it follows that $\mathbf{s}' = \mathbf{t}'$ and so if we couple the remaining pairs of eligibility vectors in the two scenarios to be identical it will result that $N(1) = N(2)$. If (b) occurs, then the remaining eligibility vectors in the two scenarios will be of the form (r, k) and $(0, r + k)$. In this case couple, the eligibility vectors to be identical unless at some time the remaining quota vectors are of the form $(u,0)$ and $(0,v)$ where, necessarily, $u \leq v$. At this point couple, all remaining eligibility vector pairs so that $X_1^* = X_2, X_2^* = X_1$, resulting in $N(1) \leq N(2)$. Because eventually one of the cases (a) or (b) will occur, it follows that there is a way of coupling the two scenarios so that $N(1) \leq N(2)$. □

3. DYNAMIC POLICIES: A QUALITATIVE RESULT AND A HEURISTIC POLICY

In this section, we allow for dynamic policies, which are policies whose decisions can be based on all that has previously occurred. With $p_i = P(X_i = 1)$, assume that $0 < p_i \leq p_{i+1}, i = 1, \ldots, n-1$. Say that the dynamic programming state of the system is $\mathbf{m} = (m_1, \ldots, m_n)$ if m_r is the current quota for box r for $r = 1, \ldots, n$, and let $V(\mathbf{m})$ be the minimal expected number of balls needed to arrive to meet all quotas when the state is $\mathbf{m}$.

We show that in searching for an optimal policy one need never consider putting a ball into a box for which there is a smaller numbered eligible box whose quota is at least as large. However, before doing so we need some definitions.

DEFINITION: Say that $\mathbf{s}$ dominates $\mathbf{t}$, written as $\mathbf{s} \geq_d \mathbf{t}$, if for some state vector $\mathbf{m} = (m_1, \ldots, m_n)$

$$\begin{aligned}
\mathbf{s} &= (m_1, \ldots, (m_i - k)^+, \ldots, m_j, \ldots\ldots, m_n) \\
\mathbf{t} &= (m_1, \ldots\ldots, m_i, \ldots\ldots, (m_j - k)^+, \ldots, m_n)
\end{aligned}$$

where $i < j, m_i \geq m_j$ and $k \geq 0$. If $\mathbf{s} \geq_d \mathbf{t}$, let $b(\mathbf{s}, \mathbf{t}) = \sum_{r=1}^n m_r$, and call $b(\mathbf{s}, \mathbf{t})$ the b-value of the pair $\mathbf{s}, \mathbf{t}$,

THEOREM 4: If $\mathbf{s} \geq_d \mathbf{t}$, then $V(\mathbf{s}) \leq V(\mathbf{t})$.

PROOF: Although a simple coupling argument can be used to show that it is never optimal to either put a ball in a dead box or to discard it when there is an alive box for which it is eligible, to facilitate the proof we allow both of these possibilities. The proof is by induction on $b(\mathbf{s}, \mathbf{t})$. As it is immediate when $\mathbf{s} \geq_d \mathbf{t}$ and $b(\mathbf{s}, \mathbf{t}) = 1$, assume it is true whenever $\mathbf{s} \geq_d \mathbf{t}$ and $b(\mathbf{s}, \mathbf{t}) < w$, and now suppose that $\mathbf{s} \geq_d \mathbf{t}$ and $b(\mathbf{s}, \mathbf{t}) = w$. Also, as the result is immediate either if $k = 0$ or $m_i = 0$ suppose that both are positive. Consider two scenarios, the first where the initial state is $\mathbf{t}$ and the second where it is $\mathbf{s}$. Let $\mathbf{X}$ and $\mathbf{Y}$ be the eligibility vectors for the initial ball in, respectively, scenarios $\mathbf{s}$ and $\mathbf{t}$, and couple these vectors as follows. Let $U, U_r, r \neq i, j$ be $n-1$ independent uniform (0, 1) random variables, and let $X_r = Y_r = I\{U_r < p_r\}, r \neq i, j$. Y_i, Y_j and X_i, X_j are defined in terms of U as following:

$$\begin{aligned}
U \leq p_i p_j &\Rightarrow Y_i = 1, Y_j = 1 \\
p_i p_j < U \leq p_i p_j + p_i(1 - p_j) = p_i &\Rightarrow Y_i = 1, Y_j = 0 \\
p_i < U \leq p_i + (1 - p_i)p_j &\Rightarrow Y_i = 0, Y_j = 1 \\
p_i + p_j - p_i p_j < U &\Rightarrow Y_i = 0, Y_j = 0.
\end{aligned}$$

Also,

$$\begin{aligned}
U \leq p_i p_j &\Rightarrow X_i = 1, X_j = 1 \\
p_i p_j < U \leq p_i p_j + (1 - p_i)p_j = p_j &\Rightarrow X_i = 0, X_j = 1 \\
p_j < U \leq p_j + p_i(1 - p_j) &\Rightarrow X_i = 1, X_j = 0 \\
p_i + p_j - p_i p_j < U &\Rightarrow X_i = 0, X_j = 0.
\end{aligned}$$

Note, because $p_i \leq p_j$, that $Y_i = 1 \Rightarrow X_j = 1$, and also note that $X_i + X_j = Y_i + Y_j$. Suppose the optimal policy is to be used in the $\mathbf{t}$-scenario. Let $n(\mathbf{t})$ be the next state after $\mathbf{t}$ and let $n(\mathbf{s})$ be the next state after $\mathbf{s}$, and consider the following cases:

- The initial ball is either discarded or put into box $r, r \neq i, j$ in the $\mathbf{t}$-scenario.

In this case do the same thing with the ball in the **s**-scenario. Note that $n(\mathbf{s}) \geq_d n(\mathbf{t})$ and that $b(n(\mathbf{s}), n(\mathbf{t})) \leq w$.

- The ball is put in box i in the **t**-scenario.
  In this case, put the ball in box j in the **s**-scenario. Because, with
$$m_i^* = m_i - 1, m_j^* = (m_j - 1)^+$$
$$n(\mathbf{s}) = (...(m_i^* - (k-1))^+, ..., m_j^*, ...)$$
$$n(\mathbf{t}) = (........m_i^*, ..., (m_j^* - (k-1))^+, ...)$$
  we see that $n(\mathbf{s}) \geq_d n(\mathbf{t})$ and that $b(n(\mathbf{s}), n(\mathbf{t})) < w$. Hence, by the induction hypothesis, $V(n(\mathbf{s})) \leq V(n(\mathbf{t}))$.
- The ball is put in box j in the **t**-scenario; thus, $n(\mathbf{t}) = (..., m_i, ..., (m_j - k - 1)^+, ...)$
  In this case, put the ball in either box i or j in the **s**-scenario (it has to be eligible for at least one of them). Thus, either $n(\mathbf{s}) = (..., (m_i - k - 1)^+, ..., m_j, ...)$ or $n(\mathbf{s}) = (..., (m_i - k)^+, ..., (m_j - 1)^+, ...)$, showing in either case that $n(\mathbf{s}) \geq_d n(\mathbf{t})$ and $b(n(\mathbf{s}), n(\mathbf{t})) \leq w$.

Hence, in all situations $n(\mathbf{s}) \geq_d n(\mathbf{t})$ and $b(n(\mathbf{s}), n(\mathbf{t})) \leq w$. In all the cases where $b(n(\mathbf{s}), n(\mathbf{t})) = w$, repeat the process for the next states, coupling the eligibility vectors and the resulting actions in the same manner as before, and continue to do so until $b(n(\mathbf{s}), n(\mathbf{t})) < w$, then the induction hypothesis will yield that the additional expected number of arriving balls needed to meet all remaining quotas in the **s**-scenario is less than or equal to its value in the **t**-scenario. Because eventually $b(n(\mathbf{s}), n(\mathbf{t}))$ will be less than w, it follows that the expected number of balls needed that must arrive to fill all quotas when starting in state **s** and using the preceding policy is less than or equal to what it is when the initial state is **t** and the optimal policy is used. □

COROLLARY 5: There is an optimal policy that, for any state **m**, never puts a ball into box j if the ball is also eligible for box $i<j$ and $m_i \geq m_j > 0$.

PROOF: Use that $(\ldots, m_i - 1, \ldots m_j, \ldots) \geq_d (\ldots, m_i, \ldots m_j - 1, \ldots)$ and apply Theorem 4. □

Because the needed computations grow exponentially in n, the standard dynamic programming recursion cannot be effectively used to find the optimal policy. As a result, we consider the heuristic policy π_h which puts an arriving ball in box i if, among all those boxes j that are eligible for the ball, i maximizes $\frac{r_j}{p_j}$ where r_j is the remaining quota of box j.

Starting with the policy π_h we then use the policy improvement algorithm; that is, given the current quota vector $(r_1, \ldots, r_n)$ put an arriving ball into that box j for which

$$E[N_{\pi_h}(r_1, \ldots, r_j - 1, \ldots, r_n)] = \min_{i: X_i = 1} \left\{ E[N_{\pi_h}(r_1, \ldots, r_i - 1, \ldots, r_n)] \right\}$$

where $E[N_{\pi_h}(r_1, \ldots, r_i - 1, \ldots, r_n)]$ denotes the expected number of balls needed to fill the quotas $(r_1, \ldots, r_i - 1, \ldots, r_n)$ when policy π_h is employed. The determination of the values $E[N_{\pi_h}(r_1, \ldots, r_i - 1, \ldots, r_n)]$ is done via simulation. We call this new policy π_h*. Thus, π_h* is an online policy.

To help measure how "good" π_h and π_h* are, we now derive a lower bound for the minimal value of $E[\mathbf{N}]$. To begin, note that

$$\min_\pi \{E[N_\pi(m_1, \ldots, m_n)]\} \geq \max_{i=1,\ldots,n} \left\{ \frac{m_i}{p_i} \right\}$$

To identify a second lower bound for $E[N]$, let $t = \sum_{i=1}^n m_i$, and let $\mathbf{B_i}, i = 0, \ldots t-1$ denote the set of the alive boxes after i balls were saved, where a ball is said to be saved if it is put in a box that was alive when it was put in. Also, let A_i denote the additional number of balls collected after i balls have been saved until $i+1$ balls have been saved, and note that $N = \sum_{i=0}^{t-1} A_i$.

LEMMA 6: Conditional on $(\mathbf{B}_0, \ldots, \mathbf{B_{t-1}})$, $A_0, \ldots, A_{t-1}$ are independent geometric random variables with $E[A_j | \mathbf{B}_0, \ldots, \mathbf{B_{t-1}}] = \frac{1}{P(\mathbf{B_j})}$, where

$$P(\mathbf{B_j}) = P(\text{a ball is eligible for at least one box in } \mathbf{B_j}) = 1 - \prod_{i \in \mathbf{B_j}} (1 - p_i).$$

PROOF: The result is similar to one proved in [16]. Its argument is based on the fact that the number of balls needed to obtain one that is eligible for at least one of the boxes in **B**, is independent of which boxes in **B** it is eligible for. To show this, let $\mathbf{X}_i, i \geq 1$, be the eligibility vectors of the successive balls, and let A denote the smallest i for which ball i is eligible for at least one of the boxes in **B**. Let $p = P(\sum_{j \in \mathbf{B}} X_j > 0)$. Then, for any $\mathbf{x} = (x_1, \ldots, x_n)$ for which $\sum_{j \in \mathbf{B}} x_j > 0$,

$$P(A = i | \mathbf{X}_A = \mathbf{x}) = \frac{P(A = i, \mathbf{X}_i = \mathbf{x})}{\sum_{i=1}^\infty P(A = i, \mathbf{X}_i = \mathbf{x})} = \frac{(1-p)^{i-1} P(\mathbf{X} = \mathbf{x})}{\sum_{i=1}^\infty (1-p)^{i-1} P(\mathbf{X} = \mathbf{x})} = p(1-p)^{i-1}$$

□

Because $p_1 \leq p_2 \cdots \leq p_n$ it follows from Lemma 6 that $N|(\mathbf{B}_0,\ldots,\mathbf{B}_{t-1})$ is stochastically minimized when $\mathbf{B_k}=\{1,\ldots,n\}, k=0,\ldots,t-n, \mathbf{B_{t-n+1}}=\{2,\ldots,n\},\ldots,\mathbf{B}_{t-1}=\{n\}$. Consequently,

$$E[N|(\mathbf{B}_0,\ldots,\mathbf{B}_{t-1})] \geq \frac{t-n+1}{1-\prod_{i=1}^{n}(1-p_i)} + \sum_{j=2}^{n}\frac{1}{1-\prod_{i=j}^{n}(1-p_i)}$$

Combining this with the earlier lower bound yields that

$$\min_{\pi}\{E[N_\pi]\} \geq \max\left\{\frac{t-n+1}{1-\prod_{i=1}^{n}(1-p_i)} + \sum_{j=2}^{n}\frac{1}{1-\prod_{i=j}^{n}(1-p_i)}, \max_{i=1\ldots n}\frac{m_i}{p_i}\right\}$$

EXAMPLE 3.1: For $\mathbf{p}=[0.1\ 0.15\ 0.2\ 0.25\ 0.3\ 0.35\ 0.4\ 0.45\ 0.5]$; $\mathbf{m}=[5\ 8\ 10\ 12\ 15\ 18\ 20\ 23\ 26]$, the following compares simulation based estimators of $E[N_{\pi_h}]$ and $E[N_{\pi_{h*}}]$ with the lower bound on the optimal $E[N]$.

$E[N_{\pi_h}]$	σ^2	$E[N_{\pi_{h*}}]$	σ^2	$L.B.\{E[N]\}$
146.1140	7.9322	144.8000	12.6222	143.5344

The variance listed in the table is per simulation run based. At each decision point, in order to determine the action chosen by π_{h*} we did 10 simulation runs for each of the (up to 9) different possible initial actions which are then followed by utilizing π_h. We then chose the action by using the average of the 10 runs. (We felt reasonably comfortable with only 10 runs for each possible action as it intuitively would not seem to make much difference which, say of 2, initial actions is chosen when their relevant expectations are roughly equal.) It is easy to see that the heuristic policy performs well and its improved version has $E[N_{\pi_{h*}}]$ very close to the lower bound.

EXAMPLE 3.2: When $n=2$, we can numerically compute the optimal value function by the standard dynamic programming algorithm when m_1 and m_2 are not too large. For $i,j=1,..,200$, $E[N_{\pi_h}(i,j)]$ and $E[N_{\pi_{h*}}(i,j)]$ denote the expected numbers of balls needed when the initial quota vector is (i,j) when using the heuristic and the improved heuristic policies; and $V(i,j)$ denotes the minimal expected number found by dynamic programming.

(p_1,p_2)	$\max_{(i,j)}\left\{\frac{E[N_{\pi_h}(i,j)]}{V(i,j)}\right\}$	$\max_{(i,j)}\left\{\frac{E[N_{\pi_{h*}}(i,j)]}{V(i,j)}\right\}$
(0.1, 0.9)	1.1033	1.0055
(0.2, 0.8)	1.0710	1.0094
(0.3, 0.7)	1.0333	1.0066
(0.4, 0.6)	1.0139	1.0021
(0.5, 0.5)	1	1
(0.005, 0.5)	1.0410	1
(0.0005, 0.5)	1	1

4. THE TWO-BOX SCENARIO

We start by finding the optimal policy for states of the form (1,m).

THEOREM 7: Suppose $n=2$ and that $p_1 < p_2$. Also, let $q_i = 1-p_i, i=1,2$. When in state (1,m) the optimal policy puts a (1, 1) labeled ball in box 1 if and only if $m \leq M_1$, where

$$M_1 = \frac{\log\left(\frac{p_1q_2}{p_2q_1}\right)}{\log\left(1-\frac{p_1}{1-q_1q_2}\right)}.$$

PROOF: The proof uses a one stage look ahead argument. Assume the current state is (1,$m+1$). Because no additional decisions need to be made after a ball is put in box 1, we can regard this as a stopping problem where the stopping action is to put the ball in box 1. We claim that the one stage look ahead policy is the optimal policy for this problem. To show this, suppose the initial ball has a (1, 1) vector. If put in box 1 (e.g., the stopping action) then the expected number of additional balls needed is $\frac{m+1}{p_2}$. Letting $N(1,m)$ be the number of additional balls needed if we put the current one in box 2, and then put the next box 1 eligible ball to arrive into box 1, then the one stage look ahead policy would put the ball into box 1 if and only if

$$\frac{m+1}{p_2} \leq E[N(1,m)].$$

Now, say that a ball is eligible if it is eligible for at least one of the boxes, and let X be the number of eligible balls that need to arrive until one is eligible for box 1. Letting $\bar{p}=\frac{p_1}{1-q_1q_2}, \bar{q}=1-\bar{p}$, and conditioning on X yields

$$E[N(1,m)] = \sum_{r=1}^{m} E[N(1,m)|X = r]\bar{p}\bar{q}^{r-1} + E[N(1,m)|X > m]\bar{q}^{m} \quad (2)$$

Now,

$$E[N(1,m)|X=r] = \frac{r}{1-q_1q_2} + \frac{m-r+1}{p_2}, \quad r \le m$$
$$E[N(1,m)|X>m] = \frac{m}{1-q_1q_2} + \frac{1}{p_1}$$

where the preceding used that the expected number of balls needed to obtain one eligible for at least one of the boxes is $\frac{1}{1-q_1q_2}$. Substituting the preceding back into (2) yields, after some elementary algebra, that

$$E[N(1,m)] = 1 + \frac{m}{p_2} + \left(\frac{q_1}{p_1}\right)\bar{q}^m \tag{3}$$

Thus, the one stage look ahead policy puts a (1, 1) ball into box 1 when the state is (1,$m+1$) if and only if

$$\left(\frac{q_1}{p_1}\right)\bar{q}^m - \frac{q_2}{p_2} \ge 0$$

Because $\left(\frac{q_1}{p_1}\right)\bar{q}^m - \frac{q_2}{p_2}$ decreases in m, and the components of a state vector cannot increase as the problem progresses, it follows that once the stopping region of the one stage look ahead policy is entered it can never be left (e.g., the monotonic case), which implies that it is the optimal policy. Solving $\left(\frac{q_1}{p_1}\right)\bar{q}^m = \frac{q_2}{p_2}$, yields the critical value $M_1 = \frac{\log\left(\frac{p_1q_2}{p_2q_1}\right)}{\log\left(1-\frac{p_1}{1-q_1q_2}\right)}$. □

We now prove a structural result about the optimal policy when the state is (2,m).

THEOREM 8: There exists a value M_2 such that, when in state (2,m), the optimal policy puts a (1, 1) labeled ball in box 1 if and only if $m \le M_2$.

PROOF: Again, the proof is made by a one-stage look ahead argument. Because once box 1 receives a ball the problem becomes one where an optimal policy is known and therefore the minimal expected additional number of balls needed is obtainable, we can treat the (2,m) problem as a stopping time problem where putting a ball in box 1 corresponds to stopping. As in the (1,m) problem, we show the (2,m) model also constitutes a monotonic stopping time problem. Suppose that the initial state is (2,$m+1$) and that the current ball has a (1, 1) vector. If that ball is put in box 1 then the minimal expected additional number of balls needed is V (1,m). Letting $H(2, m)$ be the minimal expected additional number of balls if we put the current ball in box 2, the next box 1 eligible ball in box 1, and then continue optimally, then the one-stage look ahead policy will put the current ball into box 1 if $V(1,m+1) \le H(2,m)$. Thus, to prove its optimality, it suffices to show that $H(2,m) - V(1,m+1)$ decreases monotonically in m. Letting $\bar{p} = \frac{p_1}{1-q_1q_2}, \bar{q} = 1-\bar{p}$, we have, by the same argument as used in establishing (3), that

$$H(2,m) = \sum_{r=1}^{m} \bar{p}\bar{q}^{r-1}\left[\frac{r}{1-q_1q_2} + V(1,m+1-r)\right] + \bar{q}^m\left[\frac{m}{1-q_1q_2} + \frac{2}{p_1}\right] \tag{4}$$

As we have already showed in state (1,m) that there exists an M_1 such that we should put a (1, 1) labeled ball into box 1 if and only if $m \le M_1$, we consider two cases.

CASE 1:

$$m \le M_1$$

With $N(1,\ m)$ as defined in the proof of Theorem 7, we have for $r = 1, \ldots, m$ that

$$V(1,m+1-r) = E[N(1,m+1-r)] = 1 + \frac{m+1-r}{p_2} + \left(\frac{q_1}{p_1}\right)\bar{q}^{m+1-r}.$$

Substituting the preceding into (4) gives, after some algebra, that

$$H(2,m) = 2 + \frac{m}{p_2} + \bar{q}^m\left[m\frac{q_1}{1-q_1q_2} + 2\frac{q_1}{p_1}\right]$$

Using this, it follows that

$$\begin{aligned} H(2,m) - V(1,m+1) &= -\frac{q_2}{p_2} + \bar{q}^m\left[\frac{q_1}{1-q_1q_2}m + \frac{q_1}{p_1} + \frac{q_1}{1-q_1q_2}\right] \\ &= -\frac{q_2}{p_2} + \frac{q_1}{1-q_1q_2}\bar{q}^m\left[m + \frac{1-q_1q_2}{p_1} + 1\right] \\ &= -\frac{q_2}{p_2} + E[XI\{X>m\}] + \frac{q_1}{1-q_1q_2}\bar{q}^m \end{aligned}$$

where X follows a geometric distribution with parameter $\bar{p}$. Thus $H(2,m) - V(1,m+1)$ decreases monotonically with respect to m. It is also easily verified that $H(2,M_1) - V(1,M_1+1) > 0$. Thus, given $m \le M_1$, in state (2,m) the one stage look ahead policy puts a (1, 1) ball in box 1.

CASE 2:

$$m = M_1 + k, k > 0$$

Consider the balls that are eligible for at least one box and condition on the first of them to have a 1, 0 vector, to obtain

$$V(1, M_1+k) = \sum_{r=1}^{k} \frac{p_1 q_2}{1-q_1q_2}\left(\frac{p_2}{1-q_1q_2}\right)^{r-1} \times \left[\frac{r}{1-q_1q_2} + \frac{M_1+k-r+1}{p_2}\right] + \left(\frac{p_2}{1-q_1q_2}\right)^{k}\left[\frac{k}{1-q_1q_2} + V(1, M_1)\right]$$

After some elementary algebra this yields that

$$V(1, M_1+k) = \frac{M_1+k}{p_2} + \left(\frac{p_2}{1-q_1q_2}\right)^{k}\left(V(1, M_1) - \frac{M_1}{p_2}\right)$$

Also,

$$\begin{aligned} H(2, M_1+k) &= \sum_{r=1}^{k} \frac{p_1}{1-q_1q_2}\left(\frac{q_1p_2}{1-q_1q_2}\right)^{r-1} \times \left[\frac{r}{1-q_1q_2} + V(1, M_1+k+1-r)\right] \\ &\quad + \left(\frac{q_1p_2}{1-q_1q_2}\right)^{k}\left(\frac{k}{1-q_1q_2} + H(2, M_1)\right) \\ &= 1 + \frac{M_1+k}{p_2} + \left(\frac{q_1p_2}{1-q_1q_2}\right)^{k} \times \left(\frac{M_1q_1q_2}{p_2(1-q_1q_2)} + H(2, M_1) - 1 - \frac{V(1, M_1)}{1-q_1q_2}\right) \\ &\quad + \left(\frac{p_2}{1-q_1q_2}\right)^{k}\left(\frac{V(1, M_1)}{1-q_1q_2} - \frac{M_1}{p_2(1-q_1q_2)}\right) \end{aligned}$$

After some algebra, we obtain upon using the preceding, along with the formulas for $V(1, M_1)$ and and $H(2, M_1)$ given in Case 1, that

$$H(2, M_1+k) - V(1, M_1+k+1) = \left(\frac{q_1p_2}{1-q_1q_2}\right)^{k}\left(A + \left(\frac{q_1p_2}{1-q_1q_2}\right)^{M_1} B\right) - \frac{q_2}{p_2}$$

where

$$A = \frac{q_2}{q_1^k(1-q_1q_2)} - \frac{q_1q_2}{1-q_1q_2}$$

$$B = \frac{M_1q_1}{1-q_1q_2} + \frac{2q_1}{p_1} - \frac{q_1}{p_1(1-q_1q_2)} + \frac{q_1q_2}{q_1^k p_1(1-q_1q_2)}$$

Hence, $H(2, M_1+k) - V(1, M_1+k+1)$ monotonically decreases with respect to k, with a final value $-\frac{q_2}{p_2}$ and an initial value $H(2, M_1) - V(1, M_1+1) > 0$.

Combining the two cases shows that $H(2, m) - V(1, m+1)$ decreases monotonically in m, thus establishing the optimality of the one stage lookahead policy, and enabling us to conclude that there exists a value M_2 such that it is optimal in state (2,m) to put a (1, 1) ball into box 1 if and only if $m \le M_2$. Moreover, it follows from the results of Case 1 that $M_2 \ge M_1$. □

ACKNOWLEDGMENT

This material is based on work supported by, or in part by, the US Army Research Laboratory and the US Army Research Office under contract/grant number W911NF-14-1-0166 and by the National Science Foundation under contract/grant number CMMI1233337.

REFERENCES

[1] C. Albright and C. Derman, Asymptotic optimal policies for stochastic sequential assignment problem, Manage Sci 19 (1972), 46–51.
[2] S.C. Albright, Markov-decision-chain approach to a stochastic assignment problem, Oper Res 22 (1974), 61–64.
[3] S.C. Albright, Optimal sequential assignments with random arrival times, Manage Sci 21 (1974), 60–67.
[4] F.T. Bruss and T.S. Ferguson, Multiple buying or selling with vector offers, J Appl Probab 34 (1997), 959–973.
[5] M. Brown and H. Solomon, Optimal issuing policies under stochastic field lives, J Appl Probab 10 (1973), 761–768.
[6] Y.H. Chun and R.T. Sumichrast, A rank-based approach to the sequential selection and assignment problem, Eur J Oper Res 174 (2006), 1338–1344.
[7] I. David and O. Levi, A new algorithm for the multi-item exponentially discounted optimal selection problem, Eur J Oper Res 153 (2004), 782–789.
[8] C. Derman, G.J. Lieberman, and S.M. Ross, Stochastic sequential allocation model, Oper Res 23 (1975), 1120–1130.
[9] C. Derman, G.J. Lieberman, and S.M. Ross, A sequential stochastic assignment problem, Manage Sci 18 (1972), 349–355.
[10] R. Dizon-Ross and S.M. Ross, Dynamic models for selling multiple items, Working Paper, Available at http://scholar.harvard.edu/files/rdr/files/multiple_items_2013nov24.pdf. Last accessed date: December 2, 2014.
[11] A. Ganti, E. Modiano, and J. Tsitsiklis, Optimal transmission scheduling in symmetric communication models with intermittent connectivity, IEEE Trans Inf Theory 53 (2007), 998–1008.
[12] D.P. Kennedy, Optimal sequential assignment, Math Oper Res 11 (1986), 619–626.
[13] A.J. Kleywegt and J.D. Papastavrou, The dynamic and stochastic knapsack problem with random sized items, Oper Res 49 (2001), 26–41.
[14] A.G. Nikolaev and S.H. Jacobson, Stochastic sequential decision-making with a random number of jobs, Oper Res 58 (2010), 1023–1027.
[15] R. Righter, Stochastic sequential assignment problem with arrivals, Probab Eng Inf Sci 25 (2011), 477–485.
[16] S.M. Ross and D.T. Wu, A generalized coupon collecting model as a parsimonious optimal stochastic assignment model, Ann Oper Res 208 (2013), 133–146.

[17] V. Shestak, E.K.P. Chong, A.A. Maciejewski, and H.J. Siegel, Probabilistic resource allocation in heterogeneous distributed systems with random failures, J Parallel Distrib Comput 72 (2012), 1186–1194.
[18] X.M. Su and S.A. Zenios, Patient choice in kidney allocation: A sequential stochastic assignment model, Oper Res 53 (2005), 443–455.
[19] L. Tassiulas and A. Ephremides, Dynamic server allocation to parallel queues with randomly varying connectivity, IEEE Trans Inf Theory 39 (1993), 466–478.
[20] S.A. Zenios, G.M. Chertow, and L.M. Wein, Dynamic allocation of kidneys to candidates on the transplant waiting list, Oper Res 48 (2000), 549–569.