A Sequential Stochastic Assignment Problem with Random Number of Jobs

Yaakov Malinovsky
University of Maryland, Baltimore County, USA

joint work with

Alexander Goldenshluger, University of Haifa, Israel

Assaf Zeevi, Columbia University, USA

36th Annual Quality and Productivity Research Conference
June 10-13, 2019
American University
Washington, DC
Outline

- Sequential Stochastic Assignment Problem with Fixed Number of Jobs
- Sequential Stochastic Assignment Problem with Random Number of Jobs
- Selecting one of the $k$ Best Values with Random Number of Alternatives
Part 1: Sequential Stochastic Assignment Problem with Fixed Number of Jobs
Suppose that \( n \) jobs arrive sequentially in time.

The \( t \)th job, \( 1 \leq t \leq n \), is identified with a random variable \( Y_t \) which is observed.

The jobs must be assigned to \( n \) persons which have known “values” \( p_1, \cdots, p_n \).

If the \( t \)th job is assigned to the \( j \)th person then a reward of \( p_j Y_t \) is obtained and the person \( j \) becomes unavailable.

The goal: to maximize expected total reward

\[
S_n(\pi) := E \sum_{t=1}^{n} p_{\pi_t} Y_t.
\]

Assume that $Y_1, \ldots, Y_n$ are integrable independent random variables defined on probability space $(\Omega, \mathcal{F}, P)$.

Let $F_t$ be the distribution function of $Y_t$, $t = 1, \ldots, n$.

Let $\mathcal{Y}_t$ denote the $\sigma$–field generated by $(Y_1, \ldots, Y_t)$:

$$\mathcal{Y}_t = \sigma(Y_1, \ldots, Y_t), 1 \leq t \leq n.$$ 

$\pi = (\pi_1, \ldots, \pi_n)$ is a permutation of $\{1, \ldots, n\}$ defined on $(\Omega, \mathcal{F})$.

We say that $\pi$ is an assignment policy if $\{\pi_t = j\} \in \mathcal{Y}_t$ for every $1 \leq j \leq n$ and $1 \leq t \leq n$:

$\pi$ is a policy if it is non–anticipating relative to the filtration $\mathcal{Y} = \{\mathcal{Y}_t, 1 \leq t \leq n\}$ so that $t$th job is assigned on the basis of information in $\mathcal{Y}_t$.

---

Formal Statement: Problem (AP1)

- Given a vector \( p = (p_1, \ldots, p_n) \), with \( p_1 \leq p_2 \leq \cdots \leq p_n \),
- we want to maximize the total expected reward
  \[ S_n(\pi) := \mathbb{E} \sum_{t=1}^{n} p_{\pi_t} Y_t \]
  with respect to \( \pi \in \Pi(\mathcal{Y}) \).
- The policy \( \pi^* \) is called optimal if \( S_n(\pi^*) = \sup_{\pi \in \Pi(\mathcal{Y})} S_n(\pi) \).

Useful representation:

\[
\sum_{t=1}^{n} p_{\pi_t} Y_t = \sum_{t=1}^{n} \sum_{j=1}^{n} p_j Y_t 1\{\pi_t = j\} = \sum_{j=1}^{n} p_j Y_{\nu_j};
\]

- \( \nu_j \) denotes the index of the job to which the \( j \)th person is assigned: \( \{\nu_j = t\} = \{\pi_t = j\} \), \( 1 \leq t \leq n, 1 \leq j \leq n \).
Theorem (DLR, 1972; Albright, 1972):

- There exist real numbers 
  \(-\infty \equiv a_{0,n} \leq a_{1,n} \leq \cdots \leq a_{n-1,n} \leq a_{n,n} \equiv \infty\) such that on the first step, when \(Y_1 \sim F_1\) is observed, the optimal policy is 
  \[
  \pi_1^* = \sum_{j=1}^{n} j 1\{Y_1 \in (a_{j-1,n}, a_{j,n}]\}.
  \]

- \(\{a_{j,n}\}_{j=1}^{n}\) do not depend on \(p_1, \ldots, p_n\) and are determined by 
  \[
  a_{j,n+1} = \int_{a_{j-1,n}}^{a_{j,n}} zdF_1(z) + a_{j-1,n}F_1(a_{j-1,n}) + a_{j,n}[1 - F_1(a_{j,n})],
  \]
  \(j = 1, \ldots, n\), where \(-\infty \cdot 0 \equiv 0 \equiv \infty \cdot 0\).
Backward Induction Solution (con’t)

• At the end of the first stage the assigned $p$ is removed from the feasible set and the process repeats with the next observation, where the above calculation is then performed relative to the distribution $F_2$ and real numbers $-\infty \equiv a_{0,n-1} \leq a_{1,n-1} \leq \cdots \leq a_{n-2,n-1} \leq a_{n-1,n-1} \equiv \infty$ are determined, and so on.

• Moreover,

$$a_{j,n+1} = \mathbb{E}Y_{\nu_j}, \quad \forall 1 \leq j \leq n,$$

i.e., $a_{j,n+1}$ is the expected value of the job which is assigned to the $j$th person.
Remark and Example

- By **backward induction** we determine a triangular array, where we use $F_{n-t+2}$ to determine $\{a_{.,t}\}$:

  $a_{1,2}$
  $a_{1,3}, a_{2,3}$
  $a_{1,4}, a_{2,4}, a_{3,4}$
  $\vdots$
  $a_{1,n}, a_{2,n}, \ldots, a_{n-1,n}$
  $a_{1,n+1}, a_{2,n+1}, \ldots, a_{n,n+1}$ $\Rightarrow$ $S_n(\pi^*) = p_1 \cdot a_{1,n+1} + \cdots + p_n \cdot a_{n,n+1}$

- **Example:** $X_1 \sim X_2 \sim X_3 \sim Uniform[0,1]$

  $a_{1,2} = 1/2$
  $a_{1,3} = 3/8, a_{2,3} = 5/8$
  $a_{1,4} = 39/128, a_{2,4} = 39/128, a_{3,4} = 89/128$ $\Rightarrow$

  $$S_3 = p_1 \cdot 39/128 + p_2 \cdot 1/2 + p_3 \cdot 89/128.$$
Part 2: Sequential Stochastic Assignment Problem with Random Number of Jobs
Let $N$ be a positive integer-valued random variable with known distribution $\gamma = \{\gamma_k\}$, $\gamma_k = P(N = k), k = 1, \ldots, N_{\text{max}}$, where $N_{\text{max}}$ can be infinite.

Let $Y_1, Y_2, \ldots$ be an infinite sequence of integrable independent random variables with distributions $F_1, F_2, \ldots$, independent of $N$.

Given real numbers $p_1 \leq \ldots \leq p_{N_{\text{max}}}$ the objective is to maximize the expected total reward

$$S_{\gamma}(\pi) = E\sum_{t=1}^{N} p_{\pi_t} Y_t$$

over all policies $\pi \in \Pi(Y)$. 
Theorem: 

In Problem (AP2) assume that $N_{\text{max}} < \infty$ and let

$$\tilde{Y}_t := Y_t \sum_{k=t}^{N_{\text{max}}} \gamma_k, \quad t = 1, \ldots, N_{\text{max}}$$

For any $\pi \in \Pi(\mathcal{Y})$ one has

$$S_\gamma(\pi) = \mathbb{E} \sum_{t=1}^{N_{\text{max}}} p_{\pi_t} \tilde{Y}_t,$$

and the optimal policy in Problem (AP2) coincides with the optimal policy in Problem (AP1) associated with fixed horizon $n = N_{\text{max}}$ and job sizes $\tilde{Y}_1, \ldots, \tilde{Y}_{N_{\text{max}}}$.

---

Proof

- For any \( \pi \in \Pi(\mathcal{Y}) \) we have
  \[
  S_\gamma(\pi) = \mathbb{E}\sum_{t=1}^{N} p_{\pi_t} Y_t = \sum_{t=1}^{N_{\text{max}}} \mathbb{E}[p_{\pi_t} Y_t 1(N \geq t)],
  \]

- and
  \[
  \mathbb{E}[p_{\pi_t} Y_t 1(N \geq t)] = \mathbb{E}\sum_{k=t}^{N_{\text{max}}} \mathbb{E}\left\{[p_{\pi_t} Y_t 1(N = k)] \mid \mathcal{Y}_t\right\} = \mathbb{E}\left\{p_{\pi_t} Y_t \sum_{k=t}^{N_{\text{max}}} \gamma_k\right\}
  \]

  \[
  = \mathbb{E}\left\{p_{\pi_t} \tilde{Y}_t\right\},
  \]

  where we have used the fact that \( \pi_t \) is \( \mathcal{Y}_t \)-measurable, and \( N \) is independent of \( \mathcal{Y}_t \).

- Therefore \( \mathbb{E}\sum_{t=1}^{N} p_{\pi_t} Y_t = \mathbb{E}\sum_{t=1}^{N_{\text{max}}} p_{\pi_t} \tilde{Y}_t \).

- Note that \( \tilde{Y}_t \) are independent random variables, and \( \sigma \)-fields \( \tilde{\mathcal{Y}}_t \) and \( \mathcal{Y}_t \) are identical. This implies the stated result.
Part 3: Selecting one of the $k$ Best Values with Random Number of Alternatives
Sequential Selection Problems

- Let $X_1, X_2, \ldots$ be an infinite sequence of independent identically distributed continuous random variables defined on a probability space $(\Omega, \mathcal{F}, P)$.

$$R_t := \sum_{j=1}^{t} 1(X_t \leq X_j), \quad A_{t,n} := \sum_{j=1}^{n} 1(X_t \leq X_j), \quad t = 1, \ldots, n.$$ 

- Let $\mathcal{R}_t := \sigma(R_1, \ldots, R_t)$ and $\mathcal{X}_t := \sigma(X_1, \ldots, X_t)$ denote the $\sigma$–fields generated by $R_1, \ldots, R_t$ and $X_1, \ldots, X_t$.

- $\mathcal{R} = (\mathcal{R}_t, 1 \leq t \leq n)$ and $\mathcal{X} = (\mathcal{X}_t, 1 \leq t \leq n)$ are the corresponding filtrations.

- The class of all stopping times of a filtration $\mathcal{Y} = (\mathcal{Y}_t, 1 \leq t \leq n)$ will be denoted $\mathcal{I}(\mathcal{Y})$; i.e., $\tau \in \mathcal{I}(\mathcal{Y})$ if $\{\tau = t\} \in \mathcal{Y}_t$ for all $1 \leq t \leq n$. 
**Average Reward**

- **Fixed n: Problem (A1):** Let $n$ be a fixed positive integer, and let $q : \{1, 2, \ldots, n\} \to \mathbb{R}$ be a reward function. The average reward of a stopping rule $\tau \in \mathcal{F}(\mathcal{R})$ is $V_n(q; \tau) := \mathbb{E}(A_{\tau,n})$, and we want to find the rule $\tau_* \in \mathcal{F}(\mathcal{R})$ such that

  $$V_n^*(q) := \max_{\tau \in \mathcal{F}(\mathcal{R})} V_n(q; \tau) = \mathbb{E}(A_{\tau_*,n}).$$

- **Random N: Problem (A2):** $\gamma_k = P(N = k)$, $k = 1, 2, \ldots, N_{\text{max}}$, $N \perp \{X_t, t \geq 1\}$. Let $q : \{1, 2, \ldots, N_{\text{max}}\} \to \mathbb{R}$.

  $$V_{\gamma}(q; \tau) := \mathbb{E}[q(A_{\tau,N})1(\tau \leq N)].$$

  We want to find the stopping rule $\tau_* \in \mathcal{F}(\mathcal{R})$ such that

  $$V_{\gamma}^*(q) := \max_{\tau \in \mathcal{F}(\mathcal{R})} V_{\gamma}(q; \tau) = V_{\gamma}(q; \tau_*).$$
Fixed n: Gusein-Zade Stopping Problem

- **Selecting One of the $k$ Best Values:** \( q(a) = q_{gZ}^{(k)}(a) := 1\{a \leq k\} \), and the problem is to maximize \( P\{A_{\tau,n} \leq k\} \) with respect to \( \tau \in \mathcal{I}(\mathcal{R}) \).

- **The optimal policy:** is determined by $k$ natural numbers

\[ 1 \leq \pi_1 \leq \pi_2 \leq \cdots \leq \pi_k \]

and proceeds as follows: pass the first $\pi_1 - 1$ observations and among the subsequent $\pi_1, \pi_1 + 1, \ldots, \pi_2 - 1$ choose the first best observation; if it does not exists then among the set of observations $\pi_2, \pi_2 + 1, \ldots, \pi_3 - 1$ choose one of the two best, etc.

- **Example (n=30, k=3):** $\pi_1 = 11$, $\pi_2 = 18$, $\pi_3 = 24$ and

---

\[
\max_{\tau \in \mathcal{F}(\mathcal{R})} \mathbb{P}\{A_{\tau,30} \leq 3\} = 0.73492.
\]
An Auxiliary Optimal Stopping Problem: Problem (B)

- Let \( Y_1, \ldots, Y_n \) be a sequence of integrable independent real-valued random variables with corresponding distributions \( F_1, \ldots, F_n \).

- For a stopping rule \( \tau \in \mathcal{T}(\mathcal{Y}) \) define \( W_n(\tau) := \mathbb{E}Y_\tau \). The objective is to find the stopping rule \( \tau_\ast \in \mathcal{T}(\mathcal{Y}) \) such that

\[
W_n^* := \max_{\tau \in \mathcal{T}(\mathcal{Y})} \mathbb{E}Y_\tau = W_n(\tau_\ast) = \mathbb{E}Y_{\tau_\ast}.
\]
Consider Problem (AP1) with $p_1 = 0, p_2 = 0, \ldots, p_n = 1$ and by Theorem (DLR, 1972), at step $t$ the optimal policy assign $p_n$ to the job $Y_t$ only if $Y_t > a_{n-t,n-(t-1)}$ and $\ldots$

Let $\{b_t, t \geq 1\}$ be the sequence of real numbers defined recursively by

* $b_1 = -\infty$, $b_2 = EY_n$,

* $b_{t+1} = \int_{b_t}^{\infty} z dF_{n-t+1}(z) + b_t F_{n-t+1}(b_t)$, $t = 2, \ldots, n$.

Let

$$\tau_* = \min\{1 \leq t \leq n : Y_t > b_{n-t+1}\};$$

then

$$W_n^* = EY_{\tau_*} = \max_{\tau \in \mathcal{T}(\mathcal{Y})} EY_{\tau} = b_{n+1}.$$
Reduction: Problems (A1) $\Rightarrow$ Problem (B)

- **Fixed Horizon**
  
  Let
  
  \[
  I_{t,n}(r) := \sum_{a=r}^{n-t+r} q(a) \frac{(a-1)(n-a)}{(n-t)(t-r)} = \mathbb{E}\{q(A_{t,n})|R_t = r\}, \quad r = 1, \ldots, t. \quad (1)
  \]

  \[
  Y_t := I_{t,n}(R_t), \quad t = 1, \ldots, n. \quad (2)
  \]

- **Theorem**: the optimal stopping rule $\tau_*$ solving Problem (B) with random variables $\{Y_t\}$ given in (1)–(2) also solves Problem (A1):

  \[
  V_n(q; \tau_*) = \max_{\tau \in \mathcal{F}(\mathcal{R})} \mathbb{E}q(A_{\tau,n}) = \max_{\tau \in \mathcal{F}(\mathcal{Y})} \mathbb{E}Y_{\tau}.
  \]
First we note that for any stopping rule $\tau \in \mathcal{I}(\mathcal{R})$ one has $E(q(A_{\tau,n}) = E(Y_{\tau})$, where $Y_t := E[q(A_{t,n})|\mathcal{R}_t]$.

\[ E(q(A_{\tau,n}) = \sum_{k=1}^{n} E(q(A_{\tau,n})1\{\tau = k\} = \sum_{k=1}^{n} E(q(A_{k,n})1\{\tau = k\} \]
\[ = \sum_{k=1}^{n} E[1\{\tau = k\}E[q(A_{k,n})|\mathcal{R}_k]] = \sum_{k=1}^{n} E[1\{\tau = k\}Y_k] = E(Y_{\tau}), \]

where we have used the fact that $\{\tau = k\} \in \mathcal{R}_k$. This implies that $\max_{\tau \in \mathcal{I}(\mathcal{R})} E(q(A_{\tau,n}) = \max_{\tau \in \mathcal{I}(\mathcal{Y})} E(Y_{\tau})$.

To prove the theorem it suffices to show only that
\[ \max_{\tau \in \mathcal{I}(\mathcal{R})} E(Y_{\tau}) = \max_{\tau \in \mathcal{I}(\mathcal{Y})} E(Y_{\tau}). \] (3)
* Clearly,

\[ \mathcal{Y}_t \subseteq \mathcal{R}_t, \quad \forall 1 \leq t \leq n. \quad (4) \]

* Because \( R_1, \ldots, R_n \) are independent random variables, and \( Y_t = I_{t,n}(R_t), \forall t \) we have that for any \( s, t \in \{1, \ldots, n\} \) with \( s < t \)

\[ P\{G_t \mid \mathcal{Y}_s\} = P\{G_t \mid \mathcal{R}_s\}, \quad \forall G_t \in \mathcal{Y}_t. \quad (5) \]

* The statement (3) follows from (4), (5) and Theorem 5.3. \(^a\)


This concludes the proof.
Numerical Values

<table>
<thead>
<tr>
<th>$n$</th>
<th>$k$</th>
<th>$P(n,k)$</th>
<th>$E(n,k)/n$</th>
<th>$n$</th>
<th>$k$</th>
<th>$P(n,k)$</th>
<th>$E(n,k)/n$</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>2</td>
<td>0.57956</td>
<td>0.68645</td>
<td>1,000</td>
<td>2</td>
<td>0.57417</td>
<td>0.68966</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.86917</td>
<td>0.60871</td>
<td></td>
<td>5</td>
<td>0.86123</td>
<td>0.60988</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>0.98140</td>
<td>0.54236</td>
<td></td>
<td>10</td>
<td>0.97703</td>
<td>0.54434</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>0.99755</td>
<td>0.50428</td>
<td></td>
<td>15</td>
<td>0.99609</td>
<td>0.50893</td>
</tr>
<tr>
<td>10,000</td>
<td>2</td>
<td>0.57363</td>
<td>0.68927</td>
<td>50,000</td>
<td>2</td>
<td>0.57358</td>
<td>0.68923</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.86043</td>
<td>0.61014</td>
<td></td>
<td>5</td>
<td>0.86036</td>
<td>0.61018</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>0.97658</td>
<td>0.54496</td>
<td></td>
<td>10</td>
<td>0.97654</td>
<td>0.54500</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>0.99592</td>
<td>0.50947</td>
<td></td>
<td>15</td>
<td>0.99591</td>
<td>0.50950</td>
</tr>
</tbody>
</table>

Table 1: Optimal probabilities $P(n,k)$ and the normalized expected time elapsed until stopping $E(n,k)/n$ for selecting one of the $k$ best values.
Optimal Strategy for $n = 30, k = 3$

$V_n^*(q) = 0.73492$
Reduction: Problems (A2) $\Rightarrow$ Problem (B)

- **Random Horizon**
  
  Let
  
  $J_t(r) := \sum_{k=t}^{N_{\text{max}}} \gamma_k I_{t,k}(r), \ r = 1, \ldots, t. \quad (6)$

  Define

  $Y_t := J_t(R_t) = \sum_{k=t}^{N_{\text{max}}} \gamma_k I_{t,k}(R_t), \ t = 1, \ldots, N_{\text{max}}. \quad (7)$

- **Theorem:** let $N_{\text{max}} < \infty$; then the optimal stopping rule $\tau_*$ solving Problem (B) with fixed horizon $N_{\text{max}}$ and random variables $\{Y_t\}$ given in (6)–(7) provides the optimal solution to Problem (A2):

  $V^*_\gamma(q) = \max_{\tau \in \mathcal{T}(\mathcal{R})} V_\gamma(q; \tau) = \max_{\tau \in \mathcal{T}(\mathcal{Y})} \mathbb{E}Y_\tau = W_{N_{\text{max}}}(\tau^*).$
Proof

* In Problem (A2) the reward for stopping at time $t$ is

$$\tilde{q}(A_{t,N}) = q(A_{t,N})\mathbf{1}\{N \geq t\}.$$ 

$$E\{q(A_{t,N})\mathbf{1}\{N \geq t\} | R_1 = r_1, \ldots, R_{t-1} = r_{t-1}, R_t = r\}$$

$$= \sum_{k=t}^{N_{\max}} E\{q(A_{t,N})\mathbf{1}\{N = k\} | R_1 = r_1, \ldots, R_{t-1} = r_{t-1}, R_t = r\}$$

$$= \sum_{k=t}^{N_{\max}} E\{\mathbf{1}\{N = k\}E[q(A_{t,k}) | N = k, R_t = r]\}$$

$$= \sum_{k=t}^{N_{\max}} \gamma_k \sum_{a=r}^{k-t+r} q(a) \frac{(a-1)(k-a)}{t-r} \binom{k}{t} = \sum_{k=t}^{N_{\max}} \gamma_k I_{t,k}(r) =: J_t(r). \quad (8)$$

* Together with (7) this implies that $E\tilde{q}(A_{\tau,N}) = EJ_\tau(R_\tau) = EY_\tau$ for any $\tau \in \mathcal{T}(\mathcal{R})$. The remainder of the proof proceeds along the lines of the proof of Theorem for fixed horizon $n$. 
Optimal Strategy for $N \sim Uniform\{1, 2, \ldots, 30\}, \ k = 3$
The proposed framework is applicable to sequential selection problems that can be reduced to settings with independent observations and additive reward function. In particular:

- selection problems with no-information, rank-dependent rewards and fixed or random horizon,
- selection problems with full information when the random variables $\{X_t\}$ are observable, and the reward for stopping at time $t$ is a function of the current observation $X_t$ only,
- multiple choice problems with random horizon and additive reward.
The proposed framework is not applicable to the following sequential selection problems:

- for instance, settings with rank-dependent reward and full information as in Gnedin (2007)\(^a\) cannot be reduced to optimal stopping of a sequence of independent random variables

- multiple choice problem with zero-one reward, where the problem of maximizing the probability of selecting \(k\) best alternatives; see, e.g., Rose (1982)\(^b\) where the problem of maximizing the probability of selecting \(k\) best alternatives was considered.

---


Thank You!