OPERATIONS RESEARCH
DOI 10.1287/opre.1090.0778ec

e-companion

ONLY AVAILABLE IN ELECTRONIC FORM

Electronic Companion—"Stochastic Sequential Decision-Making with a Random Number of Jobs" by Alexander G. Nikolaev and Sheldon H. Jacobson, *Operations Research*, DOI 10.1287/opre.1090.0778.

Proofs of Theorems

EC.1. Proof of Theorem 1.

THEOREM 1. *Let π'^* be an optimal policy for AP. Then, an optimal policy for BP, π^*, is obtained using rules: (1) whenever a job arrives and π'^* assigns a worker with success rate zero, discard the job, (2) whenever a job arrives and π'^* assigns a worker with success rate $p > 0$, assign a worker with the same success rate.*

Proof: Note that π^* and π'^* have the same expected rewards for BP and AP, respectively. Suppose that π^* is not optimal for BP, and that there exists a policy π_{better} that yields a higher expected reward. Then define policy π'_{better} for AP as follows:

- whenever a job arrives and π_{better} discards it, assign a worker with success rate zero
- whenever a job arrives and π_{better} assigns a worker with success rate p, assign a worker with the same success rate.

Consider the reward earned in BP by making assignments for sequence $s = \{x_j\}_{j=1}^{N}$ according to policy π_{better}. Then, the same reward is earned in AP by making assignments for sequence $s' = \{\{x'_j\}_{j=1}^{N_{max}} : x'_j = x_j \text{ for } j = 1, 2, ..., N \text{ and } x'_j = 0 \text{ for } j = N+1, N+2, ..., N_{max}\}$ according to policy π'_{better}. Also, by (1),

$$P(\{X'_n > 0\} \bigcap \{X'_{n+1} = 0\}) = 1 \cdot \frac{\sum_{i=2}^{N_{max}} P_i}{\sum_{i=1}^{N_{max}} P_i} \cdot \frac{\sum_{i=3}^{N_{max}} P_i}{\sum_{i=2}^{N_{max}} P_i} \cdot ...... \cdot \frac{\sum_{i=n}^{N_{max}} P_i}{\sum_{i=n-1}^{N_{max}} P_i} \cdot \frac{P_n}{\sum_{i=n}^{N_{max}} P_i} = P_n,$$

which means that the probability of encountering sequence s' in AP is equal to the probability of encountering sequence s in BP. Therefore, π_{better} and π'_{better} yield the same expected rewards for BP and AP, respectively, and hence, π'_{better} is a better policy than π'^*, which contradicts the assumption that π'^* is optimal. □

EC.2. Proof of Theorem 3.

THEOREM 3. *Whenever job $n = 1, 2, ..., N_{max} - 1$ arrives in AP, the optimal assignment policy is to assign the n^{th} job to the worker with the m^{th} highest success rate (available at the time the assignment decision has to be made) if x'_n lies in the m^{th} highest of the intervals, defined by the* *fixed breakpoints* $E[Z_{1,n+1}^{N_{max}}|X_n' > 0]$, $E[Z_{2,n+1}^{N_{max}}|X_n' > 0]$,..., $E[Z_{N_{max}-n,n+1}^{N_{max}}|X_n' > 0]$. *These breakpoints are computed recursively,*

$$E[Z_{m,n+1}^{N_{max}}|X_n' > 0] = \frac{\sum_{i=n+1}^{N_{max}} P(N=i)}{\sum_{i=n}^{N_{max}} P(N=i)}(F_{n+1}(E[Z_{m,n+2}^{N_{max}}|X_{n+1}' > 0])E[Z_{m,n+2}^{N_{max}}|X_{n+1}' > 0]$$

$$+\int_{E[Z_{m,n+2}^{N_{max}}|X_{n+1}'>0]}^{E[Z_{m-1,n+1}^{N_{max}}|X_{n+1}'>0]} x dF_{n+1}(x) + (1 - F_{n+1}(E[Z_{m-1,n+1}^{N_{max}}|X_{n+1}' > 0]))E[Z_{m-1,n+1}^{N_{max}}|X_{n+1}' > 0]). \tag{EC.1}$$

Proof: Theorem 2 can be used to establish an optimal policy for AP. Observe that for any $1 \leq m \leq N_{max} - n + 1$ and $1 \leq n \leq N_{max} - 1$, computing the expectation $E[Z_{m,n+1}^{N_{max}}|\mathcal{F}_n]$ (see Theorem 2) involves only the random variables X_j', $j = n+1, n+2, ..., N_{max}$. The distributions of these variables depend only on the value X_n' (which is known at the time $E[Z_{m,n+1}^{N_{max}}|\mathcal{F}_n]$ is computed), and not on $X_1', X_2', ..., X_{n-1}'$. Therefore, the stochastic process $\{X_j'\}_{j=1}^{N_{max}}$ has the Markovian property, and hence, $E[Z_{m,n+1}^{N_{max}}|\mathcal{F}_n] = E[Z_{m,n+1}^{N_{max}}|X_n']$ Next, note that by (1), $P[X_j' = 0|X_n' = 0] = 1$ for all $j = n+1, n+2, ..., N_{max}$. Then by definition, $Z_{1,N_{max}}^{N_{max}} = 0$ and $P[Z_{m,r}^{N_{max}} = 0|X_n' = 0] = 1$ for any $1 \leq m \leq N_{max} - r + 1$, $r = n+1, n+2, ..., N_{max} - 1$, and hence, $E[Z_{m,n+1}^{N_{max}}|X_n' = 0] = 0$. Also note that $E[Z_{m,n+1}^{N_{max}}|X_{n+1}' = 0] = 0$. Therefore, conditioning on X_{n+1}, $E[Z_{m,n+1}^{N_{max}}|X_n' > 0]$ can be expressed as

$$E[Z_{m,n+1}^{N_{max}}|X_n' > 0] = E[Z_{m,n+1}^{N_{max}}|X_{n+1}' > 0]P(X_{n+1}' > 0|X_n' > 0)$$

$$+E[Z_{m,n+1}^{N_{max}}|X_{n+1}' = 0]P(X_{n+1}' = 0|X_n' > 0) = \frac{\sum_{i=n+1}^{N_{max}} P_i}{\sum_{i=n}^{N_{max}} P_i}E[Z_{m,n+1}^{N_{max}}|X_{n+1}' > 0]. \tag{EC.2}$$

By definition, $Z_{m,n+1}^{N_{max}}$ can be expressed through $Z_{m,n+2}^{N_{max}}$ and $Z_{m+1,n+2}^{N_{max}}$, and hence, (EC.2) becomes

$$\frac{\sum_{i=n+1}^{N_{max}} P_i}{\sum_{i=n}^{N_{max}} P_i}E[[X_{n+1}'|X_{n+1}' > 0] \vee E[Z_{m,n+2}^{N_{max}}|X_{n+1}' > 0] \wedge E[Z_{m-1,n+1}^{N_{max}}|X_{n+1}' > 0]]$$

$$= \frac{\sum_{i=n+1}^{N_{max}} P_i}{\sum_{i=n}^{N_{max}} P_i}E[X_{n+1} \vee E[Z_{m,n+2}^{N_{max}}|X_{n+1}' > 0] \wedge E[Z_{m-1,n+1}^{N_{max}}|X_{n+1}' > 0]]. \tag{EC.3}$$

Note that in (EC.3), $P[X_{n+1}' = X_{n+1}|X_{n+1}' > 0] = 1$ by (1). Since the distribution of X_{n+1} is given, then (EC.3) becomes (EC.1). □

EC.3. Proof of Theorem 6.

THEOREM 6. *The optimal expected accumulated reward EV_1^C can be computed using the recursion,*

$$EV_j^c = \frac{\sum_{i=j}^{N_{max}} P_i}{\sum_{i=j-1}^{N_{max}} P_i}\,[\,P(W_j \le c,\, R_j + EV_{j+1}^{c-W_j} \ge EV_{j+1}^c)$$

$$\times E[R_j + EV_{j+1}^{c-W_j}\,|\,W_j \le c,\, R_j + EV_{j+1}^{c-W_j} \ge EV_{j+1}^c]$$

$$+[P(R_j + EV_{j+1}^{c-W_j} < EV_{j+1}^c,\, W_j \le c) + P(W_j > c)]\,EV_{j+1}^c\,], \tag{EC.4}$$

with boundary conditions $EV_j^c = 0$ for any c and $j \ge N_{max}$.

Proof: By the definition of EV_j^c and conditioning on the event that job j does arrive,

$$EV_j^c \equiv E(V_j^c\,|\,\text{job } j-1 \text{ has arrived}) = E(V_j^c\,|\,\text{job } j \text{ arrives})\,P(\text{job } j \text{ arrives}\,|\,\text{job } j-1 \text{ has arrived})$$

$$+\,E(V_j^c\,|\,\text{job } j \text{ does not arrive})\,P(\text{job } j \text{ does not arrive}\,|\,\text{job } j-1 \text{ has arrived}).$$

Note that $P(\text{job } j \text{ arrives}\,|\,\text{job } j-1 \text{ has arrived}) = \frac{\sum_{i=j}^{N_{max}} P_i}{\sum_{i=j-1}^{N_{max}} P_i}$. Also, if job j does not arrive, then no reward can be collected, and hence, $E(V_j^c\,|\,\text{job } j \text{ does not arrive}) = 0$. Then, (EC.4) follows from Theorem 5, since the optimal assignment policy, described by (3), is used for making each allocation decision. Since no job can arrive after N_{max} jobs, no reward can be collected, which establishes the boundary conditions. □

EC.4. Proof of Theorem 7.

THEOREM 7. *Assume that*

$$B \equiv E[\sup_j |\frac{X_j}{W_j}|] < +\infty, \text{ and } P[N < +\infty] = 1. \tag{EC.5}$$

Then for any $j = 1, 2, \ldots$ and $c \in [0, C]$, the infinite sequence $\{EV_j^c(N_{max})\}_{N_{max}=1}^{+\infty}$ converges to the finite limit $EV_j^c(\infty) \equiv \lim_{N_{max}\to+\infty} EV_j^c(N_{max})$, and Theorem 5 establishes an optimal policy for DSKP, where the pmf *of the number of jobs has infinite support, with EV_j^c replaced by $EV_j^c(\infty)$, $j = 1, 2, \ldots,\ c \in [0, C]$.*

Proof: For any $j = 1,2,...,$ $c \in [0,C]$, and $N_{max} = 1,2,...,$ $EV_j^c(N_{max}+1) \geq EV_j^c(N_{max})$ and $EV_j^c(N_{max}) \leq B\,c$. The infinite sequence $\{EV_j^c(N_{max})\}_{N_{max}=1}^{+\infty}$ is monotonically increasing and uniformly bounded, and hence, $\lim_{N_{max}\to+\infty} EV_j^c(N_{max})$ exists and is finite.

For any $j = 1,2,...$ and $c \in [0,C]$, let OV_j^c denote the optimal *conditional* expected accumulated reward from the allocation of resource capacity c to jobs $j, j+1, ...$, given that job $j-1$ has arrived. Theorem 5 establishes an optimal policy for DSKP, where the *pmf* of the number of jobs has infinite support, with EV_j^c replaced by OV_j^c, $j = 1,2,...$, $c \in [0,C]$. It suffices to show that for any $j = 1,2,...$ and $c \in [0,C]$, $OV_j^c = EV_j^c(\infty) \equiv \lim_{N_{max}\to+\infty} EV_j^c(N_{max})$.

For any $j = 1,2,...,$ $c \in [0,C]$ and $N_{max} = 1,2,...,$ $OV_j^c \geq EV_j^c(N_{max})$ and $\lim_{N_{max}\to+\infty} EV_j^c(N_{max}) \geq EV_j^c(N_{max})$, and hence, $OV_j^c - EV_j^c(\infty) > 0$.

Suppose that an optimal policy for accepting jobs is followed up until the arrival of some job N, and then the jobs $N+1, N+2, ...$ can be accepted without regard to their weights. Then, the expected total accumulated reward is higher than OV_j^c, and the following expression holds,

$$OV_j^c - EV_j^c(\infty) \leq EV_j^c(N) + \sum_{i=N+1}^{\infty} [X_i\, P(\text{job } i \text{ arrives})] - EV_j^c(\infty) \leq B\,C \sum_{i=N+1}^{\infty} [P(\text{job } i \text{ arrives})].$$

By (EC.5), for any $\Delta > 0$, there exists N_Δ such that $\sum_{i=N_\Delta+1}^{\infty} [P(\text{job } i \text{ arrives})] < \frac{\Delta}{B\,C}$, which means that with $N = N_\Delta$, $OV_j^c - EV_j^c(\infty) < \Delta$. Since Δ can be chosen arbitrarily, then $OV_j^c = EV_j^c(\infty)$.

□