E-Companion

EC.1. Proof of Theorem 1

Proof of Theorem 1. For any instance of arrival sequence S, we will show $\frac{\mathsf{ALG}_1(S)}{\mathsf{OPT}^+(S)} \geq 3/7$.

First of all, Greedy always accepts at least one item. Denote the set of items accepted by Greedy as G, and we know $G \neq \emptyset$. Denote $\mathsf{size}(G) = g$. If $G = [T]$ then Greedy is optimal. In this case

$$\frac{\mathsf{ALG}_1}{\mathsf{OPT}^+} \geq \Pr(\tau = 0) \cdot 1 + \Pr(\tau > 0) \cdot 0 \geq F(0) = 4/7 \geq 3/7.$$

If $G \subsetneq [T]$, let $M = [T] \backslash G$ denote the set of items blocked by Greedy. Since Greedy always accepts an item as long as it can fill in, any item blocked by Greedy must exceed the remaining space of the knapsack, at the moment it is blocked.

Let m be the size of the smallest item in M, i.e. $m = \min_{t \in M} s_t$. Define index t_m for the smallest item, or the first smallest item, if there are multiple smallest items,

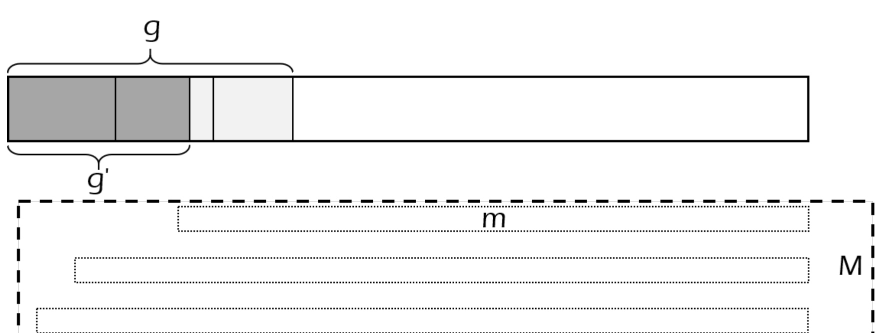
$$t_m = \min\{t \in [T] \,|\, s_t = m\}. \tag{EC.1}$$

Denote G' as the set of items accepted by Greedy, at the moment item t_m is blocked. Let $g' = \mathsf{size}(G')$. See Figure EC.1. A straightforward, but useful observation about m is:

$$g' + m > 1, \tag{EC.2}$$

because m is blocked by Greedy.

Figure EC.1 **Illustration of the items that** Greedy **accepts, and blocks. Note that m is the size of the smallest item in M, not necessarily the first.**



Next, we wish to understand when we can pack an item of size at least m, by selecting a proper threshold τ. We distinguish two cases: $m \geq 1/2$ and $m < 1/2$.

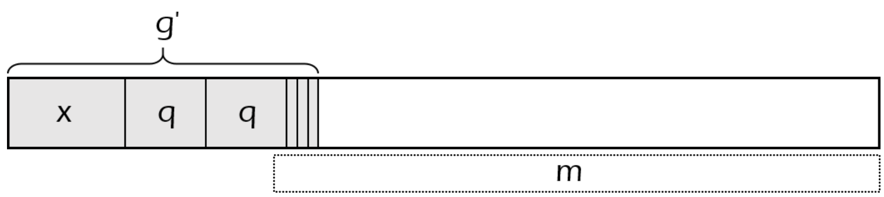
Case 1: $m \geq 1/2$.

Let $S^{\mathsf{THR}}(\tau)$ be the set of items that have sizes at least τ, i.e. $S^{\mathsf{THR}}(\tau) = \{t \in S \,|\, s_t \geq \tau\}$. Now define

$$\begin{aligned} q = \max \quad & \tau \\ s.t. \quad & m + \mathsf{size}(S^{\mathsf{THR}}(\tau) \cap G') > 1. \end{aligned} \tag{EC.3}$$

This means that if we adopt a $\mathsf{THR}(q)$ policy, then item t_m must not be accepted because $m + \mathsf{size}(S^{\mathsf{THR}}(\tau) \cap G') > 1$. That is, the items that a $\mathsf{THR}(q)$ policy accepts before item t_m arrives would exceed $1 - m$, leaving not enough space to accept item t_m the size m item[2].

Figure EC.2 Illustration of Case 1 (and specifically, Case 1.2)



Now consider the items in $S^{\mathsf{THR}}(q) \cap G'$. These items have sizes at least q. We count how many size q items are there, and let n be the number of size q items. Denote the total size of the remaining items in $S^{\mathsf{THR}}(q) \cap G'$ be x. We know that $\mathsf{size}(S^{\mathsf{THR}}(q) \cap G') = nq + x$. See Figure EC.2.

We make the following observations:

1. There must exist some item from G' that is of size q, i.e.

$$\exists t_q \in G' \subseteq [T], s.t.\ s_{t_q} = q. \tag{EC.4}$$

This is because otherwise we can select the smallest item in G' whose size is (strictly) larger than q. This size satisfies (EC.3), and violates the maximum property of q.

2. Size m items can not fit in together with all the items in $S^{\mathsf{THR}}(q) \cap G'$, i.e.,

$$nq + x + m > 1 \tag{EC.5}$$

This is because $\mathsf{size}(S^{\mathsf{THR}}(q) \cap G') = nq + x$. And then the constraint in (EC.3) implies (EC.5).

3. A size m item can fit in together with items $S^{\mathsf{THR}}(\tau) \cap G', \forall \tau > q$, i.e.,

$$x + m \leq 1 \tag{EC.6}$$

This is because otherwise $x + m > 1$, and we could further strictly increase q to still satisfy the constraint in (EC.3). Define

$$\hat{q} = \min \left\{ s_t \,\middle|\, s_t > q, t \in S^{\mathsf{THR}}(q) \cap G' \right\}.$$

[2] This does not exclude the possibility that it is also rejected, due to $q > m$, which leads to the discussion in Case 1.1.

We know (i) $\hat{q} > q$; (ii) $\mathsf{size}(S^{\mathsf{THR}}(\hat{q}) \cap G') + m = x + m > 1$. So $\hat{q}$ violates the maximum property of q.

We further distinguish two cases: $q > m$, and $q \le m$.

Case 1.1: $q > m$.

In this case, if we adopt Greedy then we can get g.

If we adopt $\mathsf{THR}(\tau), \forall \tau \in (0, q]$ then we can get no less than q. This is because due to (EC.4) there must exist some item $t_q \in G'$ of size q. We either accept it, in which case we immediately earn q, or we block it because we have accepted some item $z \in [T]$ from M that arrived earlier and consumed too much space. In the latter case, Greedy blocks item z earlier than it accepts item t_q, which means that s_z is no less than the remaining capacity when z arrives, which is no less than the remaining capacity when t_q arrives at a later time, which is further no less than q, the size of t_q which is accepted. So we have $s_z \ge s_{t_q} = q$ in the second case. Putting both cases together we earn q.

We have the following:

$$
\begin{aligned}
\mathsf{ALG}_1 &\ge \Pr(\tau = 0) \cdot g + \Pr(0 < \tau \le q) \cdot q \\
&= F_1(0) \cdot g + (F_1(q) - F_1(0)) \cdot q \\
&\ge F_1(0) \cdot (1 - 2q) + F_1(q) \cdot q \\
&= 4/7 \cdot (1 - 2q) + 1 \cdot q \\
&= 4/7 - 1/7 \cdot q \\
&\ge 3/7
\end{aligned}
$$

where the second inequality is because $g \ge g' > 1 - m$ (due to (EC.2)) and $1 - m > 1 - q$ (Case 1.1: $q > m$); second equality is because $q > m \ge 1/2$ and the way we defined $F_1(\cdot)$ in (5) so $F_1(q) = 1$; last inequality is because $q \le 1$.

Since $\mathsf{OPT}^+ \le 1$, we have $\frac{\mathsf{ALG}_1}{\mathsf{OPT}^+} \ge \frac{3}{7}$.

Case 1.2: $q \le m$.

In this case, if we adopt Greedy then we can get as much as g.

If we adopt $\mathsf{THR}(\tau), \forall \tau \in (0, q]$ then we get no less than q. This is because due to (EC.4) there must exist some item $t_q \in G'$ of size q. We either accept it, in which case we immediately earn q, or we block it because we have accepted some item $z \in [T]$ from M that arrived earlier and consumed too much space. In the latter case, Greedy blocks item z earlier than it accepts item t_q, which means that $s_z \ge s_{t_q} = q$. So in either case we earn q.

If we adopt $\mathsf{THR}(\tau), \forall \tau \in (q, m]$ then we get no less than m. This is because due to (EC.6), all the items from $S^{\mathsf{THR}}(\tau) \cap G'$ altogether will not block item t_m (from expression (EC.1)); and $\tau \le m$

so we will not reject item t_m. We either accept item t_m, in which case we immediately earn m, or we block it because we have accepted some item $z \in [T]$ from M and consumed too much space. But m is smallest item size in M, which means that $s_z \geq m$. So in either case we earn m.

We have the following:

$$\begin{aligned}
\mathsf{ALG}_1 &\geq \Pr(\tau = 0) \cdot g + \Pr(0 < \tau \leq q) \cdot q + \Pr(q < \tau \leq m) \cdot m \\
&= F_1(0) \cdot g + (F_1(q) - F_1(0)) \cdot q + (F_1(m) - F_1(q)) \cdot m \\
&\geq F_1(0) \cdot (nq + x) + (F_1(q) - F_1(0)) \cdot q + (F_1(m) - F_1(q)) \cdot (1 - (nq + x)) \\
&= (F_1(q) - F_1(0)) \cdot q + 1 - F_1(q) + (F_1(q) - 3/7) \cdot (nq + x) \\
&\geq (F_1(q) - F_1(0)) \cdot q + 1 - F_1(q) + (F_1(q) - 3/7) \cdot q \\
&= F_1(q) \cdot (2q - 1) + 1 - q
\end{aligned}$$

where the second inequality is because $g \geq g' \geq nq + x$ and $m > 1 - (nq + x)$ (due to (EC.5)); second equality is because $m \geq 1/2$ and the way we defined $F_1(\cdot)$ in (5) so $F_1(m) = 1$; the last inequality is because $F_1(q) \geq F_1(0) = 4/7 > 3/7$, so the coefficient in front of $nq + x$ is positive.

Now we plug in the expression of $F_1(q)$ as defined in (5). If $q \leq 3/7$ then $\mathsf{ALG}_1 \geq \frac{4/7 - q}{1 - 2q} \cdot (2q - 1) + 1 - q = 3/7$; If $q > 3/7$ then $\mathsf{ALG}_1 \geq 1 \cdot (2q - 1) + 1 - q = q > 3/7$. So in either case we have shown $\mathsf{ALG}_1 \geq 3/7$.

Since $\mathsf{OPT}^+ \leq 1$, we have $\frac{\mathsf{ALG}_1}{\mathsf{OPT}^+} \geq \frac{3}{7}$.

Case 2: $m < 1/2$.

In this case, a crude analysis is enough. See Figure EC.3.

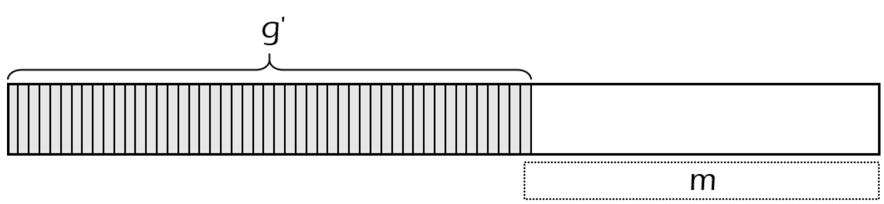


Figure EC.3 Illustration of Case 2

If we adopt Greedy then we can get as much as g.

If we adopt $\mathsf{THR}(\tau), \forall \tau \in (0, m]$ then we either get m, or m is blocked, in which case we must have already earned at least $1 - m$ to block m.

We have the following:

$$\begin{aligned}
\mathsf{ALG}_1 &\geq \Pr(\tau = 0) \cdot g + \Pr(0 < \tau \leq m) \cdot \min\{m, 1 - m\} \\
&\geq \Pr(\tau = 0) \cdot g + \Pr(0 < \tau \leq m) \cdot m
\end{aligned}$$

$$= F_1(0) \cdot g + (F_1(m) - F_1(0)) \cdot m$$
$$\geq F_1(0) \cdot (1-m) + (F_1(m) - F_1(0)) \cdot m$$
$$= F_1(m) \cdot m + 4/7 \cdot (1-2m)$$

where the second inequality is because $m < 1/2$; the last inequality is because $g \geq g' > 1-m$ (due to (EC.2)).

Now we plug in the expression of $F_1(m)$ as defined in (5). If $m > 3/7$ then

$$\mathsf{ALG}_1 \geq 4/7 - 1/7 \cdot m \geq 3/7,$$

because $m < 1/2 \leq 1$. If $m \leq 3/7$ then

$$\mathsf{ALG}_1 \geq \frac{4/7 - m}{1-2m} \cdot m + \frac{4}{7} \cdot (1-2m).$$

Note that

$$\frac{4/7 - m}{1-2m} \cdot m + \frac{4}{7} \cdot (1-2m) = \frac{9}{28} \cdot (1-2m) + \frac{1}{28} \cdot \frac{1}{1-2m} + \frac{3}{14} \geq 2\sqrt{\frac{9}{28} \cdot \frac{1}{28}} + \frac{3}{14} = \frac{3}{7}.$$

So in either case we have $\mathsf{ALG}_1 \geq 3/7$.

Since $\mathsf{OPT}^+ \leq 1$, we have $\frac{\mathsf{ALG}_1}{\mathsf{OPT}^+} \geq \frac{3}{7}$.

In all, we have enumerated all the possible cases, to find $\dfrac{\mathsf{ALG}_1}{\mathsf{OPT}^+} \geq \dfrac{3}{7}$ always holds. □

EC.2. Proof of Theorem 3

Proof of Theorem 3. We are going to show that, for any instance of arrival sequence S, we have $\frac{\mathsf{ALG}_2(S)}{\mathsf{OPT}(S)} \geq c_2$. We lower bound $\mathsf{ALG}_2(S)$ and upper bound $\mathsf{OPT}(S)$ at the same time.

First of all, if the arrival sequence is not empty set, Greedy always accepts something. Denote the set of items accepted by Greedy as G. Denote $\mathsf{size}(G) = g$. If $G = [T]$ then Greedy is optimal. In this case

$$\frac{\mathsf{ALG}_2}{\mathsf{OPT}} \geq \Pr(\tau = 0) \cdot 1 + \Pr(\tau > 0) \cdot 0 \geq F_2(0) = 1 - c_2 \geq c_2.$$

If $G \subsetneqq [T]$, let $M = [T] \backslash G$ denote the set of items blocked by Greedy. Since Greedy always accepts an item as long as it can fill in, any item blocked by Greedy must exceed the remaining space of the knapsack, at the moment it is blocked. We also know that $G \cup M = [T]$, $G \cap M = \emptyset$.

Let m be the smallest size in M, i.e. $m = \min_{t \in M} s_t$. Define index t_m for the smallest item, or the first smallest item, if there are multiple smallest items.

$$t_m = \min \{ t \in [T] \,|\, s_t = m \}. \qquad \text{(EC.7)}$$

Denote G' as the set of items accepted by Greedy, at the moment s_{t_m} is blocked. Let $g' = \mathsf{size}(G')$. See Figure EC.1. A straightforward, but useful observation about m is:

$$g' + m > 1, \tag{EC.8}$$

because m is blocked by Greedy. We wish to understand when we can accept an item of size at least m, by selecting a proper threshold τ.

We distinguish two cases: $m > 1/2$ and $m \leq 1/2$.

Case 1: $m > 1/2$.

Let $S^{\mathsf{THR}}(\tau)$ be the set of items that have sizes at least τ, i.e. $S^{\mathsf{THR}}(\tau) = \{t \in S \,|\, s_t \geq \tau\}$. Now define

$$\begin{aligned} q = \max \quad & \tau \\ s.t. \quad & m + \mathsf{size}(S^{\mathsf{THR}}(\tau) \cap G') > 1 \end{aligned} \tag{EC.9}$$

This means that if we adopt a $\mathsf{THR}(q)$ policy, then the size m item must be blocked (possibly it will also be rejected, due to $q > m$).

Now consider the items in $S^{\mathsf{THR}}(q) \cap G'$. See Figure EC.2. These items have sizes at least q. We count how many size q items are there, and let n be the number of size q items. Denote the total size of the remaining items by x. We know that

$$\mathsf{size}(S^{\mathsf{THR}}(q) \cap G') = nq + x. \tag{EC.10}$$

We make the following observations:

1. There must exist some item from G' that is of size q, i.e.

$$\exists t_q \in G' \subseteq [T], s.t.\ s_{t_q} = q. \tag{EC.11}$$

   This is because otherwise we can select the smallest item size in G' that is also larger than q. This item size satisfies (EC.9), and violates the maximum property of q.

2. Size m items can not fit in together with items $S^{\mathsf{THR}}(q) \cap G'$, i.e.

$$nq + x + m > 1 \tag{EC.12}$$

   This is because $\mathsf{size}(S^{\mathsf{THR}}(q) \cap G') = nq + x$. This is implied by (EC.9).

3. A size m item can fit in together with items $S^{\mathsf{THR}}(\tau) \cap G', \forall \tau > q$, i.e.

$$x + m \leq 1 \tag{EC.13}$$

   This is because otherwise we could further increase q to $\hat{q}$ so that $\mathsf{size}(S^{\mathsf{THR}}(\hat{q}) \cap G') + m > 1$, which violates the maximum property of q.

We further distinguish two cases: $q > m$, and $q \le m$.

Case 1.1: $q > m$.

In this case, if we adopt Greedy then we can get as much as g.

If we adopt $\mathsf{THR}(\tau), \forall \tau \in (0, q]$ then we can get no less than q. This is because due to (EC.11) there must exist some item $t_q \in G'$ of size q. We either accept it, in which case we immediately earn q, or we block it because we have accepted some item $z \in [T]$ from M that arrived earlier and consumed too much space. In the latter case, Greedy blocks item z earlier than it accepts item t_q, which means that $s_z \ge s_{t_q} = q$. So in either case we earn q.

We have the following:

$$\begin{aligned}
\mathsf{ALG}_2 &\ge \Pr(\tau = 0) \cdot g + \Pr(0 < \tau \le q) \cdot q \\
&= F_2(0) \cdot g + (F_2(q) - F_2(0)) \cdot q \\
&\ge F_2(0) \cdot (1 - 2q) + F_2(q) \cdot q \\
&= (1 - c_2) \cdot (1 - 2q) + \left[2(1 - c_2) - \frac{1 - 2c_2}{q} \right] \cdot q \\
&= c_2
\end{aligned}$$

where the second inequality is because $g \ge g' > 1 - m$ (due to (EC.8)) and $1 - m > 1 - q$ (Case 1.1: $q > m$); second equality is because $q > m \ge 1/2 > q_2$, so we plug in $F_2(\cdot)$ as defined in (11).

Since $\mathsf{OPT} \le 1$, we have $\frac{\mathsf{ALG}}{\mathsf{OPT}} \ge c_2$.

Case 1.2: $q \le m$.

First we wish to upper bound OPT. OPT selects some items from $[T] = G \cup M$, where $G \cap M = \emptyset$. Notice that $m > 1/2$ so there is at most 1 item from M that OPT can select. If OPT selects no item from M, then $\mathsf{OPT} \le g$. With probability $F_2(0)$, ALG_2 adopts Greedy and earns g. So we have

$$\frac{\mathsf{ALG}_2}{\mathsf{OPT}} \ge \Pr(\tau = 0) \cdot 1 + \Pr(\tau > 0) \cdot 0 \ge F_2(0) = 1 - c_2 \ge c_2.$$

If OPT selects one item from M, let $t_{m'} \in [T]$ be this item. So $s_{t_{m'}} = m' \ge m$. See Figure EC.4.

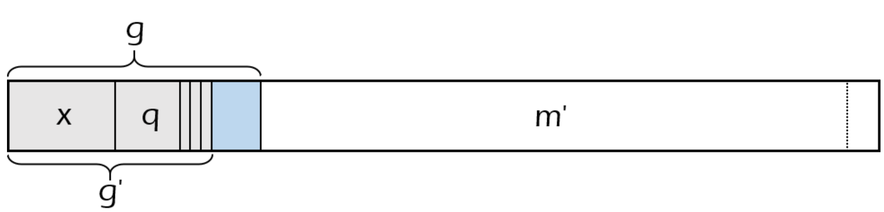


Figure EC.4 Illustration of the items accepted by OPT

We can partition all the items in S into three sets:

$$M; \qquad S^{\mathsf{THR}}(q) \cap G'; \qquad G \backslash (S^{\mathsf{THR}}(q) \cap G')$$

Let $\tilde{g} = \mathsf{size}(G \backslash (S^{\mathsf{THR}}(q) \cap G'))$. Since $S^{\mathsf{THR}}(q) \cap G'$ and $G \backslash (S^{\mathsf{THR}}(q) \cap G')$ form a partition of G, we have $g = (nq + x) + \tilde{g}$. From (EC.12) we know that $m' + \mathsf{size}(S^{\mathsf{THR}}(q) \cap G') \geq m + \mathsf{size}(S^{\mathsf{THR}}(q) \cap G') > 1$. This means that OPT cannot pack item $t_{m'}$ and $S^{\mathsf{THR}}(q) \cap G'$ together. OPT must block at least one item from $\{t_{m'}\} \cup (S^{\mathsf{THR}}(q) \cap G')$ – and the smallest item from this union is of size q (because $q \leq m \leq m'$). So we upper bound OPT by:

$$\begin{aligned} \mathsf{OPT} &\leq \min\left\{1, \left[m' + \mathsf{size}(S^{\mathsf{THR}}(q) \cap G')\right] - q + \mathsf{size}(G \backslash (S^{\mathsf{THR}}(q) \cap G'))\right\} \\ &= \min\left\{1, m' + (nq + x) - q + \tilde{g}\right\}. \end{aligned} \tag{EC.14}$$

In other words, we allow OPT to accept the items of size $\tilde{g}$; but we need to argue that OPT must block at least q from G'.

Then we analyze ALG_2. If we adopt Greedy then we can get as much as g.

If we adopt $\mathsf{THR}(\tau), \forall \tau \in (0, q]$ then we get no less than $nq + x$. This is because due to (EC.10) there must exist some items in $S^{\mathsf{THR}}(q) \cap G'$, which are of size $nq + x$. For any item in $(S^{\mathsf{THR}}(q) \cap G')$, if we accept them all, we immediately earn $(nq + x)$. Suppose we block some item in $(S^{\mathsf{THR}}(q) \cap G')$ because we have accepted some item $z \in [T]$ from M and consumed too much space. But Greedy blocks item z earlier than it accepts the item in $(S^{\mathsf{THR}}(q) \cap G')$, whose smallest possible size is q. This suggests that $s_z \geq q$. But this contradicts with the fact that we do not accept item z using the $\mathsf{THR}(\tau)$ threshold policy when $\tau \leq q$.

If we adopt $\mathsf{THR}(\tau), \forall \tau \in (q, m]$ then we get no less than m. This is because due to (EC.13), any item in $S^{\mathsf{THR}}(\tau) \cap G'$ will not block item t_m (from expression (EC.7)); and $\tau \leq m$ so we will not reject item t_m. We either accept item t_m, in which case we immediately earn m, or we block it because we have accepted some item $z \in [T]$ from M and consumed too much space. But m is smallest item size in M, which means that $s_z \geq m$. So in either case we earn m.

If we adopt $\mathsf{THR}(\tau), \forall \tau \in (m, m']$ then we get no less than τ. This is because $s_{t_{m'}}$ does exist, and $\mathsf{THR}(\tau)$ must accept at least one item. The least that $\mathsf{THR}(\tau)$ can get is τ.

We have the following:

$$\begin{aligned} \mathsf{ALG}_2 &\geq \Pr(\tau = 0) \cdot g + \Pr(0 < \tau \leq q) \cdot (nq + x) + \Pr(q < \tau \leq m) \cdot m + \int_m^{m'} \tau \, \mathrm{d}F_2(\tau) \\ &= F_2(0) \cdot (nq + x + \tilde{g}) + (F_2(q) - F_2(0)) \cdot (nq + x) + (F_2(m) - F_2(q)) \cdot m + \int_m^{m'} \tau \, \mathrm{d}F_2(\tau) \\ &= F_2(0) \cdot \tilde{g} + F_2(q) \cdot (nq + x - m) + F_2(m') \cdot m' - \int_m^{m'} F_2(\tau) \, \mathrm{d}\tau \\ &= F_2(0) \cdot \tilde{g} + F_2(q) \cdot (nq + x - m) + F_2(m') \cdot m' - \int_m^{m'} \left(2(1 - c_2) - \frac{1 - 2c_2}{\tau}\right) \, \mathrm{d}\tau \\ &\geq F_2(0) \cdot \tilde{g} + F_2(q) \cdot (2(nq + x) - 1) + F_2(m') \cdot m' - \int_{1-(nq+x)}^{m'} \left(2(1 - c_2) - \frac{1 - 2c_2}{\tau}\right) \, \mathrm{d}\tau \end{aligned}$$

where the second equality is due to integration by part (our definition of $F_2(\cdot)$ in (11) is a continuous function); the third equality is because $m > 1/2 > q_2$ and we plug in the expression of $F_2(\cdot)$; the last inequality is because $\frac{\partial \mathsf{ALG}_2}{\partial m} = F_2(m) - F_2(q) \geq 0$, (because $q \leq m$, and $F_2(\cdot)$ is a increasing function), so that ALG_2 is increasing in m. Hence, ALG_2 achieves its minimum when m is the smallest, and $m > 1-(nq+x)$ from (EC.12).

Observe that

$$\begin{aligned}\mathsf{ALG}_2 - c_2\mathsf{OPT} \geq\ & F_2(0)\cdot\tilde{g} + F_2(q)\cdot(2(nq+x)-1) + F_2(m')\cdot m' \\ & - \int_{1-(nq+x)}^{m'} \left(2(1-c_2) - \frac{1-2c_2}{\tau}\right)\, d\tau - c_2\cdot\min\{1, m' + (nq+x) - q + \tilde{g}\}\end{aligned}$$

If we focus on the dependence of $\tilde{g}$, we find that

$$\frac{\partial\left(\mathsf{ALG}_2 - c_2\mathsf{OPT}\right)}{\partial \tilde{g}} \geq F_2(0) - c_2 = 1 - 2c_2 \geq 0,$$

where the first inequality is because the subgradient of the subtracted term is either 0 or c_2. Since $\mathsf{ALG}_2 - c_2\mathsf{OPT}$ is a increasing function of $\tilde{g}$, it achieves its minimum when $\tilde{g} = 0$.

Thus,

$$\begin{aligned}\mathsf{ALG}_2 - c_2\mathsf{OPT} \geq\ & F_2(q)\cdot(2(nq+x)-1) + F_2(m')\cdot m' \\ & - \int_{1-(nq+x)}^{m'} \left(2(1-c_2) - \frac{1-2c_2}{\tau}\right)\, d\tau - c_2\cdot\min\{1, m' + (nq+x) - q\}\end{aligned}$$

Now let $y = (n-1)q + x$, and we plug in $F_2(\cdot)$ as we defined in (EC.12).

Case 1.2.1: When $q \leq q_2$, we have:

$$\begin{aligned}& \mathsf{ALG}_2 - c_2\mathsf{OPT} \\ \geq\ & F_2(q)\cdot(2(q+y)-1) + F_2(m')\cdot m' - \int_{1-(q+y)}^{m'} \left(2(1-c_2) - \frac{1-2c_2}{\tau}\right)\, d\tau - c_2\cdot\min\{1, m'+y\} \\ =\ & (1-c_2)(2(q+y)-1) + \frac{(2q-1+2y)(1-2c_2)\ln(1-q)}{2q-1} + 2(1-c_2)m' - (1-2c_2) \\ & - 2(1-c_2)m' + 2(1-c_2)[1-(q+y)] + (1-2c_2)\left[\ln m' - \ln(1-(q+y))\right] - c_2\cdot\min\{1, m'+y\} \\ =\ & c_2 + \frac{(2q-1+2y)(1-2c_2)\ln(1-q)}{2q-1} + (1-2c_2)\left[\ln m' - \ln(1-(q+y))\right] - c_2\cdot\min\{1, m'+y\}\end{aligned}$$

If we focus on the dependence of m', we will see that $\mathsf{ALG}_2 - c_2\mathsf{OPT}$ has only one local minimum: when $m' < 1-y$ we have

$$\frac{\partial\left(\mathsf{ALG}_2 - c_2\mathsf{OPT}\right)}{\partial m'} = \frac{1-2c_2}{m'} - c_2 \leq \frac{1-2c_2}{1/2} - c_2 = 2 - 5c_2 < 0,$$

because $m' \geq m \geq 1/2$. So $\mathsf{ALG}_2 - c_2\mathsf{OPT}$ is decreasing on m' when $m' < 1-y$. When $m' > 1-y$ we have

$$\frac{\partial\left(\mathsf{ALG}_2 - c_2\mathsf{OPT}\right)}{\partial m'} = \frac{1-2c_2}{m'} > 0,$$

so $\mathsf{ALG}_2 - c_2\mathsf{OPT}$ is increasing on m'. Hence, $\mathsf{ALG}_2 - c_2\mathsf{OPT}$ achieves its minimum when $m' = 1-y$.

Plugging in $m' = 1-y$, we have further

$$\mathsf{ALG}_2 - c_2\mathsf{OPT} \geq \frac{(2q-1+2y)(1-2c_2)\ln(1-q)}{2q-1} + (1-2c_2)\left[\ln(1-y) - \ln(1-(q+y))\right]$$

If we focus on the dependence of y, we find that

$$\frac{\partial(\mathsf{ALG}_2 - c_2\mathsf{OPT})}{\partial y} = (1-2c_2)\left[2\frac{\ln(1-q)}{2q-1} - \frac{1}{1-y} + \frac{1}{1-y-q}\right] > 0,$$

because $\ln(1-q) < 0, 2q-1 < 2q_2-1 < 0, \frac{1}{1-y-q} - \frac{1}{1-y} \geq 0$. Since $\mathsf{ALG}_2 - c_2\mathsf{OPT}$ is increasing on y, it achieves its minimum when $y = 0$.

Finally, plugging in $y = 0$, we have

$$\mathsf{ALG}_2 - c_2\mathsf{OPT} \geq \frac{(2q-1)(1-2c_2)\ln(1-q)}{2q-1} - (1-2c_2)\ln(1-q) = 0$$

Case 1.2.2: When $q > q_2$, we have:

$$\begin{aligned}
&\mathsf{ALG}_2 - c_2\mathsf{OPT} \\
\geq& F_2(q)\cdot(2(q+y)-1) + F_2(m')\cdot m' - \int_{1-(q+y)}^{m'} F_2(\tau)\ d\tau - c_2\cdot\min\{1, m'+y\} \\
=& 2(1-c_2)(2(q+y)-1) - \frac{(1-2c_2)(2(q+y)-1)}{q} + 2(1-c_2)m' - (1-2c_2) \\
&- 2(1-c_2)m' + 2(1-c_2)[1-(q+y)] + (1-2c_2)\left[\ln m' - \ln(1-(q+y))\right] - c_2\cdot\min\{1, m'+y\} \\
=& 2(1-c_2)(y+q) - (1-2c_2) - \frac{(1-2c_2)(2(q+y)-1)}{q} \\
&+ (1-2c_2)\left[\ln m' - \ln(1-(q+y))\right] - c_2\cdot\min\{1, m'+y\}
\end{aligned}$$

Again, if we focus on the dependence of m', we will see that $\mathsf{ALG}_2 - c_2\mathsf{OPT}$ has only one local minimum when $m' = 1-y$.

Plugging in $m' = 1-y$, we have further

$$\begin{aligned}
&\mathsf{ALG}_2 - c_2\mathsf{OPT} \geq \\
&\qquad (1-c_2)(2(y+q)-1) - \frac{(1-2c_2)(2(q+y)-1)}{q} + (1-2c_2)\left[\ln(1-y) - \ln(1-(q+y))\right]
\end{aligned}$$

Again, if we focus on the dependence of y, we find that

$$\frac{\partial(\mathsf{ALG}_2 - c_2\mathsf{OPT})}{\partial y} = 2(1-c_2-\frac{1-2c_2}{q}) + (1-2c_2)\left[-\frac{1}{1-y} + \frac{1}{1-y-q}\right] > 0,$$

because $1-c_2-\frac{1-2c_2}{q} \geq 1-c_2-\frac{1-2c_2}{q_2} \approx 0.142 > 0, \frac{1}{1-y-q} - \frac{1}{1-y} \geq 0$. Since $\mathsf{ALG}_2 - c_2\mathsf{OPT}$ is increasing on y, it achieves its minimum when $y = 0$.

Finally, plugging in $y=0$, we have

$$\begin{aligned}
\mathsf{ALG}_2 - c_2\mathsf{OPT} \geq & (1-c_2)(2q-1) - \frac{(1-2c_2)(2q-1)}{q} - (1-2c_2)\ln(1-q) \\
= & (2q-1)\left[(1-c_2) - \frac{(1-2c_2)}{q}\right] - (1-2c_2)\ln(1-q) \\
\geq & (2q-1)\left[\frac{(1-2c_2)\cdot(-\ln(1-q))}{1-2q}\right] - (1-2c_2)\ln(1-q) \\
= & 0
\end{aligned}$$

where the second inequality is because $H(c_2,q) = \frac{1-2c_2}{q} - \frac{(1-2c_2)\ln(1-q)}{1-2q} - (1-c_2) \geq 0, \forall q \in (0,1/2)$ from (10), and when $q \in [1/2,1]$, the second line expression is an increasing function of q (because $2q-1; (1-c_2) - \frac{(1-2c_2)}{q}$; and $-(1-2c_2)\ln(1-q)$ are all increasing in q), thus plugging in $q=1/2$ we have $\mathsf{ALG}_2 - c_2\mathsf{OPT} \geq -(1-2c_2)\ln(1-q) > 0$.

In all, $\mathsf{ALG}_2 \geq c_2\mathsf{OPT}$.

Case 2: $m \leq 1/2$.

In this case, a crude analysis is enough. See Figure EC.3.

If we adopt Greedy then we can get as much as g.

If we adopt $\mathsf{THR}(\tau), \forall \tau \in (0,m]$ then we either get m, or m is blocked, in which case we must have already earned at least $1-m$ to block m.

We have the following:

$$\begin{aligned}
\mathsf{ALG} &\geq \Pr(\tau=0)\cdot g + \Pr(0<\tau\leq m)\cdot \min\{m, 1-m\} \\
&\geq \Pr(\tau=0)\cdot g + \Pr(0<\tau\leq m)\cdot m \\
&= F_2(0)\cdot g + (F_2(m) - F_2(0))\cdot m \\
&\geq F_2(0)\cdot(1-m) + (F_2(m) - F_2(0))\cdot m \\
&= F_2(0)\cdot(1-2m) + F_2(m)\cdot m \\
&\geq (1-c_2)(1-2m) + \left[2(1-c_2) - \frac{1-2c_2}{m}\right]\cdot m \\
&= c_2
\end{aligned}$$

where the second inequality is because $m \leq 1/2$; the third inequality is because $g \geq g' > 1-m$ (due to (EC.8)); the last inequality is because we plug in $F_2(\cdot)$ as defined in (11), and using the fact that $(1-c_2) - \frac{(1-2c_2)\ln(1-x)}{1-2x} > 2(1-c_2) - \frac{1-2c_2}{x}$ (the first piece is larger than the second piece) in (11).

Since $\mathsf{OPT} \leq 1$, we have $\frac{\mathsf{ALG}}{\mathsf{OPT}} \geq c_2$.

In all, we have enumerated all the possible cases, to find $\dfrac{\mathsf{ALG}}{\mathsf{OPT}} \geq c_2$ always holds. □

EC.3. Proof of Theorem 4

Proof of Theorem 4. Let $u(\cdot)$ be a continuous function that describes the density in $[1-q_2, 1]$. Let the random arrival sequence be S:

$$S = \begin{cases} (\underbrace{\varepsilon, \varepsilon, \ldots, \varepsilon}_{(1-q)/\varepsilon+1 \text{ many}}, q), & \text{where } q \in [1-q_2, 1] \text{ conforms to } u(\cdot); \\ (q_2, 1-q_2+\varepsilon, 1-q_2+2\varepsilon, \ldots, 1), & \text{with prob. } x; \\ (\underbrace{\varepsilon, \varepsilon, \ldots, \varepsilon}_{(1-q_2)/\varepsilon+1 \text{ many}}, q_2), & \text{with prob. } y; \\ (\varepsilon, 1), & \text{with prob. } z; \end{cases} \tag{EC.15}$$

where

$$x = \frac{1-2c_2}{1-2q_2} \approx 0.3711; \qquad y = \frac{1-2c_2}{q_2} \approx 0.4231; \qquad z = c_2 - x \approx 0.0613; \qquad u(q) = \frac{x}{q}.$$

We can verify that

$$x + y + z + \int_{1-q_2}^{1} \frac{x}{q} \, \mathrm{d}q = x + \frac{1-2c_2}{q_2} + (c_2 - x) - \frac{1-2c_2}{1-2q_2} \cdot \ln(1-q_2) = 1,$$

by plugging the expressions into the equation and using $H(c_2, q_2) = 0$ from (9). This equation shows that our construction conforms to a legitimate probability measure.

Following each realization of S, $\mathsf{OPT}(S) = 1$. So we have $\mathbb{E}_S[\mathsf{OPT}(S)] = 1$.

For any $\mathsf{ALG} = \mathsf{THR}(\tau), \tau \in [0, 1]$, we enumerate all the potential values of τ in the following.

Case 1: $0 \le \tau \le \varepsilon$. In this case,

$$\begin{aligned} \mathbb{E}_S[\mathsf{ALG}(S)] &= q_2 \cdot x + (1-q_2+\varepsilon) \cdot y + \varepsilon \cdot z + \int_{1-q_2}^{1} (1-q+\varepsilon) \cdot u(q) \mathrm{d}q \\ &= q_2 \cdot x + (1-q_2) \cdot \frac{1-2c_2}{q_2} - \frac{1-2c_2}{1-2q_2} \cdot \ln(1-q_2) - x \cdot q_2 + \varepsilon \cdot (1-x) \\ &= \frac{1-2c_2}{q_2} - \frac{1-2c_2}{1-2q_2} \cdot \ln(1-q_2) - (1-2c_2) + \varepsilon \cdot (1-x) \\ &= c_2 + \varepsilon \cdot (1-x), \end{aligned}$$

where the coefficient of $(1-x)$ next to the ϵ term is simplified using expression (9); and the last equality is due to (9).

Case 2: $\varepsilon < \tau \le q_2$. In this case,

$$\begin{aligned} \mathbb{E}_S[\mathsf{ALG}(S)] &= q_2 \cdot x + q_2 \cdot y + 1 \cdot z + \int_{1-q_2}^{1} q \cdot u(q) \mathrm{d}q \\ &= q_2 \cdot \frac{1-2c_2}{1-2q_2} + q_2 \cdot \frac{1-2c_2}{q_2} + c_2 - \frac{1-2c_2}{1-2q_2} + q_2 \cdot \frac{1-2c_2}{1-2q_2} \\ &= c_2 + (1-2c_2)\left(2\frac{q_2}{1-2q_2} + 1 - \frac{1}{1-2q_2}\right) \\ &= c_2 \end{aligned}$$

Case 3: $q_2 < \tau \le 1 - q_2 + \varepsilon$. In this case,

$$\begin{aligned}\mathbb{E}_S[\mathsf{ALG}(S)] &= (1-q_2+\varepsilon)\cdot x + 1\cdot z + \int_{1-q_2}^{1} q\cdot u(q)\mathrm{d}q \\ &= (1-q_2)\cdot\frac{1-2c_2}{1-2q_2} + c_2 - \frac{1-2c_2}{1-2q_2} + q_2\cdot\frac{1-2c_2}{1-2q_2} + \varepsilon\cdot x \\ &= c_2 + \varepsilon\cdot x\end{aligned}$$

Case 4: $1 - q_2 + \varepsilon < \tau \le 1$. In this case,

$$\begin{aligned}\mathbb{E}_S[\mathsf{ALG}(S)] &\ge \tau\cdot x + 1\cdot z + \int_{\tau}^{1} q\cdot u(q)\mathrm{d}q \\ &= \tau\cdot x + c_2 - x + (1-\tau)\cdot x \\ &= c_2\end{aligned}$$

In all, we have enumerated all the values that a threshold can take. In all cases, the performance of the threshold $\mathsf{THR}(\tau)$ policy has an expected performance of no more than $c_2 + \varepsilon\cdot(1-x)$. But $\mathbb{E}_S[\mathsf{OPT}(S)] = 1$. By taking $\varepsilon \to 0^+$ we finish the proof. $\square$

EC.4. Proofs of Theorems 5 and 6

Theorem 6 is a generalization of Theorem 5. However, for better exposition, we prove both Theorems 5 and 6 separately. We start with Theorem 5.

Proof of Theorem 5 For any knapsack j, let I_j be the set of items routed to it in Step 2 of Definition 4 (I_j includes items that are later discarded by the threshold of knapsack j). Note that I_j does not depend on the adoption of single-knapsack algorithms from Step 3.

Denote $\mathsf{OPT}_j^+ = \min\{\sum_{t\in I_j} s_{tj}, B_j\}, \forall j \in [N]$. Note that $\mathsf{ALG}_{\mathsf{AW}} = \sum_{j\in[N]} \mathsf{OPT}_j^+$, due to the allowance of truncation in $\mathsf{ALG}_{\mathsf{AW}}$.

From Definition 1, we earn at least $\frac{3}{7}\cdot\mathsf{OPT}_j^+$ from knapsack j in expectation. So when we focus on all knapsacks,

$$\mathsf{ALG} \ge \frac{3}{7}\sum_{j\in[N]}\mathsf{OPT}_j^+ \ge \frac{3}{7}(\frac{1}{2}\mathsf{OPT}_{\mathsf{AW}}(S)) \ge \frac{3}{14}\mathsf{OPT},$$

where the first inequality is because on each knapsack ALG earns at least $\frac{3}{7}$ fraction of what $\mathsf{ALG}_{\mathsf{AW}}$ does; the second inequality is from Proposition 1; and the third inequality is simply the fact that $\mathsf{OPT}_{\mathsf{AW}} \ge \mathsf{OPT}$, because any optimal assignment when truncation is not allowed is a feasible solution to the problem when truncation is allowed. $\square$

Following Theorem 5, we generalize Definition 4 and present Definition EC.1 below. Let $\mathsf{ALG}_{\mathsf{AW}}$ be any algorithm designed for the AdWords problem; let $\mathsf{ALG}_{\mathsf{SK}}$ be any algorithm designed for the single knapsack problem. Similar to Definition 4, we give a formal statement of combining the above two algorithms as follows.

Definition EC.1. Let d_{tj} denote the "phantom" capacity filled in each knapsack $j \in [N]$ at time $t \in [T]$, if all the items routed to knapsack j were accepted, with truncation allowed. The algorithm proceeds as follows:

1. Initialization: at time $t=0$, set $b_{1j}=0, d_{1j}=0, \forall j \in [N]$.
2. For each item t, use $\mathsf{ALG}_{\mathsf{AW}}$ to route item t to knapsack $\tilde{j}$, and update (i) $d_{(t+1)\tilde{j}} = d_{t\tilde{j}} + s_{t\tilde{j}}$; (ii) $d_{(t+1)j} = d_{tj}, \forall j \neq \tilde{j}$.
3. For item t that is routed to knapsack $\tilde{j}$, use $\mathsf{ALG}_{\mathsf{SK}}$ to accept or discard item t. If item t is accepted, update (i) $b_{(t+1)\tilde{j}} = b_{t\tilde{j}} + s_{t\tilde{j}}$; (ii) $b_{(t+1)j} = b_{tj}, \forall j \neq \tilde{j}$; if item t is discarded, $b_{(t+1)j} = b_{tj}, \forall j \in [N]$.

Proof of Theorem 6 For any knapsack j, let I_j be the set of items routed to it in Step 2 of Definition EC.1 (I_j includes items that are later discarded by $\mathsf{ALG}_{\mathsf{SK}}$). Note that I_j does not depend on the adoption of single-knapsack algorithms from Step 3.

Denote $\mathsf{OPT}_j^+ = \min\{\sum_{t \in I_j} s_{tj}, B_j\}, \forall j \in [N]$. Note that $\mathsf{ALG}_{\mathsf{AW}} = \sum_{j \in [N]} \mathsf{OPT}_j^+$, due to the allowance of truncation in $\mathsf{ALG}_{\mathsf{AW}}$.

Note that $\mathsf{ALG}_{\mathsf{SK}}$ is Γ_{SK}-competitive for the single knapsack problem relative to the optimal fractional packing. We earn at least $\Gamma_{\mathsf{SK}} \cdot \mathsf{OPT}_j^+$ from knapsack j in expectation. So when we focus on all knapsacks,

$$\mathsf{ALG} \geq \Gamma_{\mathsf{SK}} \sum_{j \in [N]} \mathsf{OPT}_j^+ \geq \Gamma_{\mathsf{SK}}(\Gamma_{\mathsf{AW}} \mathsf{OPT}_{\mathsf{AW}}(S)) \geq \Gamma_{\mathsf{AW}} \Gamma_{\mathsf{SK}} \mathsf{OPT},$$

where the first inequality is because on each knapsack ALG earns at least Γ_{SK} fraction of what $\mathsf{ALG}_{\mathsf{AW}}$ does; the second inequality is from Proposition 1; and the third inequality is simply the fact that $\mathsf{OPT}_{\mathsf{AW}} \geq \mathsf{OPT}$, because any optimal assignment when truncation is not allowed is a feasible solution to the problem when truncation is allowed. $\square$

EC.5. Proof of Theorem 7

Proof of Theorem 7. Consider the example we described above, where $N=4$, and $\alpha = 1 - \frac{12}{7}\varepsilon$. By equation (12) above,

$$\mathbb{E}_S[\mathsf{OPT}(S)] = (1 - \frac{12}{7}\varepsilon)N\varepsilon + (\frac{12}{7}\varepsilon)N = \frac{76}{7}\varepsilon - \frac{48}{7}\varepsilon^2.$$

Now we analyze the maximum possible value of $\mathbb{E}_S[\mathsf{ALG}(S)]$. As discussed before, any algorithm is characterized by $e \in \{0, 1, \dots, 4\}$, which is the number of size-ε items accepted.

Case 1: $e=0$. With probability α, the arrival sequence terminates with 0 accepted; with probability $1-\alpha$, there are N more items. We enumerate all the 24 possibilities, to find there are 6 cases that a deterministic algorithm accepts 2 of them, 17 cases that accepts 3, and 1 case that accepts 4. In expectation we fill 67/24 into the knapsacks. In this case

$$\frac{\mathbb{E}_S[\mathsf{ALG}(S)]}{\mathbb{E}_S[\mathsf{OPT}(S)]} = \frac{\frac{12}{7}\cdot\varepsilon\cdot\frac{67}{24}}{\frac{76}{7}\cdot\varepsilon - \frac{48}{7}\cdot\varepsilon^2} = \frac{67}{152-96\varepsilon}.$$

Case 2: $e=1$. With probability α, the arrival sequence terminates with ε accepted; with probability $1-\alpha$, there are N more items. Out of all the 24 possibilities, there are 16 cases that a deterministic algorithm accepts 2 of them, and 8 cases that accepts 3. In expectation we fill 56/24 into the knapsacks. In this case

$$\frac{\mathbb{E}_S[\mathsf{ALG}(S)]}{\mathbb{E}_S[\mathsf{OPT}(S)]} = \frac{1\cdot\varepsilon + \frac{12}{7}\cdot\varepsilon\cdot\frac{56}{24}}{\frac{76}{7}\cdot\varepsilon - \frac{48}{7}\cdot\varepsilon^2} = \frac{35}{76-48\varepsilon}.$$

Case 3: $e=2$. With probability α, the arrival sequence terminates with 2ε accepted; with probability $1-\alpha$, there are N more items. Out of all the 24 possibilities, there are 6 cases that a deterministic algorithm accepts 1 of them, and 18 cases that accepts 2. In expectation we fill 42/24 into the knapsacks. In this case

$$\frac{\mathbb{E}_S[\mathsf{ALG}(S)]}{\mathbb{E}_S[\mathsf{OPT}(S)]} = \frac{2\cdot\varepsilon + \frac{12}{7}\cdot\varepsilon\cdot\frac{42}{24}}{\frac{76}{7}\cdot\varepsilon - \frac{48}{7}\cdot\varepsilon^2} = \frac{35}{76-48\varepsilon}.$$

Case 4: $e=3$. With probability α, the arrival sequence terminates with 3ε accepted; with probability $1-\alpha$, there are N more items. The algorithm must be able to fill one item into the unfilled knapsack in the first round of phase two. In this case

$$\frac{\mathbb{E}_S[\mathsf{ALG}(S)]}{\mathbb{E}_S[\mathsf{OPT}(S)]} = \frac{3\cdot\varepsilon + \frac{12}{7}\cdot\varepsilon\cdot 1}{\frac{76}{7}\cdot\varepsilon - \frac{48}{7}\cdot\varepsilon^2} = \frac{33}{76-48\varepsilon}.$$

Case 5: $e=4$. With probability α, the arrival sequence terminates with 4ε accepted; with probability $1-\alpha$, there are N more items. But the algorithm cannot fill in any because all the knapsacks are all occupied with ε's. In this case

$$\frac{\mathbb{E}_S[\mathsf{ALG}(S)]}{\mathbb{E}_S[\mathsf{OPT}(S)]} = \frac{4\cdot\varepsilon + \frac{12}{7}\cdot\varepsilon\cdot 0}{\frac{76}{7}\cdot\varepsilon - \frac{48}{7}\cdot\varepsilon^2} = \frac{28}{76-48\varepsilon}.$$

In all cases, any policy has an expected performance of no more than $\frac{35}{76-48\varepsilon}$. By taking $\varepsilon\to 0^+$ we finish the proof. □

EC.6. Additional Simulations

Recall that in Section 2.1 we discussed the intuition that we should choose a CDF that lives at the intersection. Different CDFs that live at the intersection will have different performance. In this section, we compare two different CDFs through simulations.

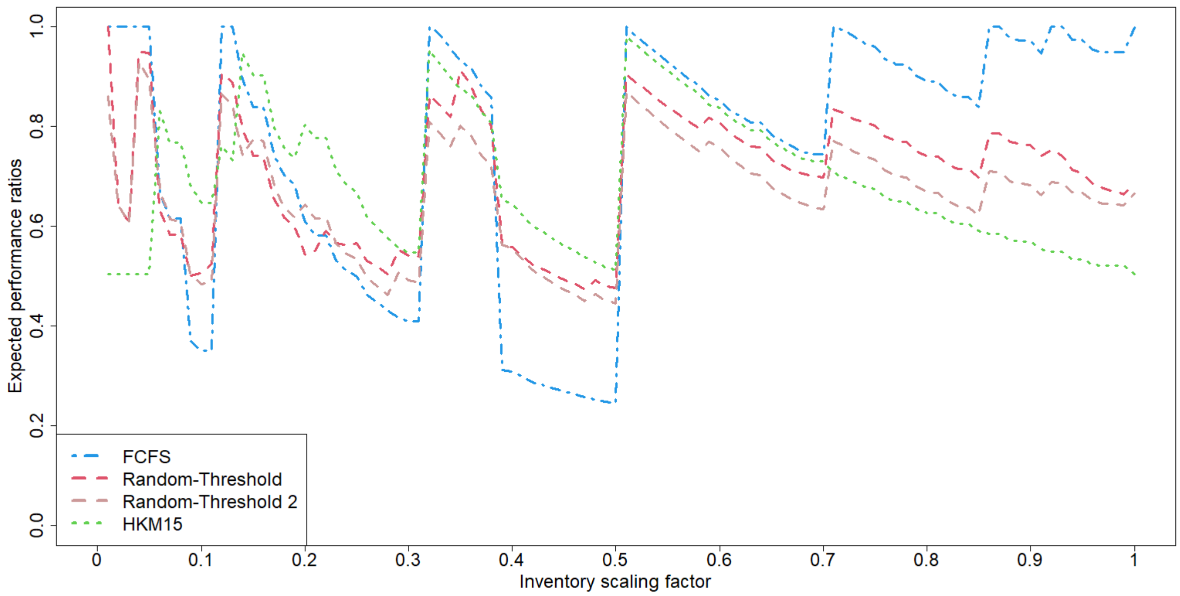


Figure EC.5 Computational performance on one real arrival sequence shown in Equation (13)**.**

Similar to the simulation setup as in Section 4, we consider the arrival sequence as defined in (13). We compute the expected revenue from our proposed random threshold algorithm from Definition 1, named **Random-Threshold**; the (deterministic) revenue from first-come-first-serve policy, named **FCFS**; and the expected revenue from the algorithm suggested in Han et al. (2015), named **HKM15**. Additionally, we compute the expected revenue from a random threshold algorithm whose CDF is specified as

$$F(x) = \begin{cases} \dfrac{\frac{4}{7} - x}{1 - 2x}, & x \in [0, 1/3] \\ \dfrac{8}{7} - \dfrac{1}{7x}, & x \in (1/3, 1] \end{cases}$$

We refer to this algorithm as **Random-Threshold 2** in our simulation.

The results are shown in Figure EC.5, where we have divided all the numbers by its corresponding offline optimal integer packing. The offline optimal packing serves as an upper bound, so that the performance ratio is always between 0 and 1, with higher ratios indicating better performance. As we can see in Figure EC.5, although **Random-Threshold** and **Random-Threshold 2** have two CDFs that live at the same intersection, they have different performance. **Random-Threshold** slightly outperforms **Random-Threshold 2**.