Q1

Let X_n denote the number of elements picked when there are n elements left. Note that $|S_n| = X_n$ is a random variable. Then for any $n \geq 1$ we can write the following recurrence relation:

$$\mathbf{E}[X_n] = \sum_{i=1}^{n} \frac{1}{n}(1 + \mathbf{E}[X_{n-i}])$$

where $X_0 = 0$. Similarly for any $n \geq 2$ we can write:

$$\mathbf{E}[X_{n-1}] = \sum_{i=1}^{n} \frac{1}{n-1}(1 + \mathbf{E}[X_{n-1-i}]).$$

Thus we have

$$n\mathbf{E}[X_n] - (n-1)\mathbf{E}[X_{n-1}] = 1 + \mathbf{E}[X_{n-1}]$$

which is equivalent to

$$\mathbf{E}[X_n] = \mathbf{E}[X_{n-1}] + \frac{1}{n}.$$

Recursively expand this expression we get

$$\mathbf{E}[X_n] = \sum_{i=1}^{n} \frac{1}{i} = H_n.$$

Q2

We need the following fact in this problem.

Fact 1 *$\left(1 - \frac{1}{n}\right)^n$ lies between $\frac{1}{4}$ and $\frac{1}{e}$ for $n \geq 2$.*

(a) Let X_i be the number of purchases until we have collected i different toys. Then $X_1 = 1$ and let $X_0 = 0$. Consider the random variable $Y_i = X_i - X_{i-1}$, i.e. the number of purchases to obtain the ith toy when we have $i-1$ toys.

When we have $i-1$ toys, the probability that on the next purchase, we obtain the new ith toy is $\frac{n-i+1}{n}$, and thus the number of purchase needed to obtain the ith toy Y_i follows a geometric distribution with $p_i = \frac{n-i+1}{n}$. Thus

$$\mathbf{E}[Y_i] = \frac{1}{p_i} = \frac{n}{n-i+1}.$$

Therefore, we have

$$\mu = \mathbf{E}[X] = \sum_{i=1}^{n} \mathbf{E}[Y_i] = nH_n.$$

(b) The probability that we have not collected a particular toy within 2μ draws is $\left(1 - \frac{1}{n}\right)^{2\mu} < \frac{1}{n^2}$ where we use the fact that $\left(1 - \frac{1}{n}\right)^n < \frac{1}{e}$ when $n \geq 2$. Taking a union bound for all toys, we have

$$\mathbf{Pr}[X > 2\mu] < n \cdot \frac{1}{n^2} = \frac{1}{n} = o(1).$$

(c) Let A_j denote the event that you have collected a toy of type j in $\frac{\mu}{2}$ draws. We know that for any j

$$\mathbf{Pr}[A_j] = 1 - \left(\frac{n-1}{n}\right)^{\mu/2} \leq 1 - \left(\frac{1}{4}\right)^{\frac{\log n}{2}} = 1 - \frac{1}{n^{\log 2}}.$$

Let $A = \bigcap_{j=1}^{n} A_j$. Our goal is to show $\mathbf{Pr}[A] \leq o(1)$. Observe that the A_j's are "negatively correlated" events in the sense that

$$\mathbf{Pr}[A_j] \geq \mathbf{Pr}\left[A_j \mid \bigcap_{i \in S} A_i\right]$$

for any subset $S \subseteq [n] \setminus \{j\}$. Thus we know

$$\mathbf{Pr}[A] = \mathbf{Pr}[A_1] \prod_{j=2}^{n} \mathbf{Pr}\left[A_j \mid \bigcap_{i \in [j-1]} A_i\right] \leq \prod_{j=1}^{n} \mathbf{Pr}[A_j] \leq \left(1 - \frac{1}{n^{\log 2}}\right)^n < \left(\frac{1}{e}\right)^{n^{1-\log 2}} = o(1).$$

Q3

(a) By construction, when $j < i$, there is no edge (π_j, i), and thus $\mathbf{E}\left[x_{\pi_j,i}\right] = 0$. $\mathbf{E}\left[x_{\pi_j,i}\right] \leq \frac{1}{n-i+1}$ if $j \geq i$. To see this latter fact, note that for any two elements $j, k \in \{i, i+1, ..., n\}$, we have $\mathbf{E}\left[x_{\pi_j,i}\right] = \mathbf{E}[x_{\pi_k,i}]$ by symmetry, since the subgraph consisting of all edges observed up until time i has an automorphism that exchanges j and k. Since $\mathbf{E}[x_{\pi_i,i}] = \mathbf{E}\left[x_{\pi_{i+1},i}\right] = \cdots = \mathbf{E}[x_{\pi_n,i}]$ and the sum of these numbers is at most 1, each of them is at most $\frac{1}{n-i+1}$.

(b)

$$\begin{aligned}
\sum_{i,j\in[n]} \mathbf{E}\big[x_{\pi_j,i}\big] &= \sum_{j=1}^{n}\sum_{i=1}^{j} \mathbf{E}\big[x_{\pi_j,i}\big] \\
&= \sum_{j=1}^{k}\sum_{i=1}^{j} \mathbf{E}\big[x_{\pi_j,i}\big] + \sum_{j=k+1}^{n}\sum_{i=1}^{j} \mathbf{E}\big[x_{\pi_j,i}\big] \\
&\le \sum_{j=1}^{k}\sum_{i=1}^{j} \frac{1}{n-i+1} + \sum_{j=k+1}^{n} 1 \\
&= \sum_{i=1}^{k}\sum_{j=i}^{k} \frac{1}{n-i+1} + \sum_{j=k+1}^{n} 1 \\
&= \sum_{i=1}^{k} \frac{k-i+1}{n-i+1} + \sum_{j=k+1}^{n} 1
\end{aligned}$$

Let $k = n - \lceil \frac{n}{e} \rceil$, and observe that $\frac{5}{n} + \sum_{i=1}^{k} \frac{1}{n-i+1} > 1$. This is because $n-k+1 = \lceil \frac{n}{e} \rceil + 1 < \frac{n+6}{e}$ and thus

$$\frac{5}{n} + \sum_{i=1}^{k} \frac{1}{n-i+1} > \frac{1}{n+5} + \cdots + \frac{1}{n+1} + \sum_{i=1}^{k} \frac{1}{n-i+1} > \int_{(n+6)/e}^{n+6} \frac{1}{x} dx = 1.$$

Therefore, we have

$$\begin{aligned}
\sum_{i,j\in[n]} \mathbf{E}\big[x_{\pi_j,i}\big] &\le \sum_{i=1}^{k} \frac{k-i+1}{n-i+1} + \sum_{j=k+1}^{n} 1 \\
&< \sum_{i=1}^{k} \frac{k-i+1}{n-i+1} + \sum_{j=k+1}^{n} \Big(\frac{5}{n} + \sum_{i=1}^{k} \frac{1}{n-i+1}\Big) \\
&\le 5 + \sum_{i=1}^{k} \Big(\frac{k-i+1}{n-i+1} + \frac{n-k}{n-i+1}\Big) \\
&= 5 + k \\
&< 5 + \left(1 - \frac{1}{e}\right) n.
\end{aligned}$$

Q4

(a) For four balls of weights 0.2, 0.4, 0.6, and 0.8, respectively, the optimal packing only takes two bins. However, if we use FF, the sequence that sees balls from light to heavy would take 3 bins.

(b) Suppose, on the contrary, OPT uses k bins and FF uses f ($f \geq 2k+1$) bins. If there are two or more of these f bins having balls of weight $\leq \frac{1}{2}$, then there is a contradiction because such two bins can be merged into one. Therefore, we assume there is only one bin of weight $\leq \frac{1}{2}$. For the other $f-1$ bins containing balls of weight more than $1/2$, the total weight of these $f-1$ bins is greater than $2k \cdot \frac{1}{2} = k$, and obviously, k is no less than the total weight of these balls. Therefore, there is also a contradiction.

(c) In this part, we prove a competitive ratio that is close to $\frac{7}{4}$.

In order to bound the competitive ratio of FF on arbitrary composition of balls, we first consider the following situation: If all the balls are of weight $\leq \frac{1}{2}$ and their total weights are k, at most how many bins can be used if we apply FF on any possible sequence of these "small" balls? Later we will employ this bound to prove another bound for the general cases.

Lemma 2 *Suppose all balls are of weight $\leq \frac{1}{2}$ and their total weights are k, then given any sequence of them, FF takes at most $\lfloor \frac{3}{2}k + \frac{1}{2} \rfloor$ bins.*

Proof: The idea in this proof is that we abstractly list the final positioning of balls, then show a limit on the number of bins we can have. Assume FF use t bins in total to put all the balls, and also assume the positioning is as follows:

$$\{f_1\}, \{f_2\}, \ldots, \{f_t\}$$

where f_i denotes the total weight of balls in bin i. If we denote the empty space in bin i by e_i, obviously $f_i + e_i = 1$. Here we make two important observations:

- For any two bins i, j such that $1 \leq i < j \leq t$, the weight of any single item in bin j must be larger than e_i.
- For any two bins i, j, $f_i + f_j > 1$.
- Except for the bin t, every bin must contain at least 2 items.

According to the information above, we have

$$f_1 + f_2 + \cdots + f_t = k$$

$$e_1 + e_2 + \cdots + e_t = t - k$$

$$f_2 > 2e_1,\ f_3 > 2e_2,\ \ldots,\ f_{t-1} > 2e_{t-2}$$

with which we can show that

$$\begin{aligned}
k &= f_1 + f_2 + \cdots + f_t \\
&> f_1 + 2(e_1 + e_2 + \cdots + e_{t-2}) + f_t \\
&= f_1 + f_t + 2(t - k - (e_{t-1} + e_t)) \\
&= f_1 + f_t + 2(t - k - (2 - (f_{t-1} + f_t))) \\
&= (f_1 + f_t) + 2(f_{t-1} + f_t) + 2(t-k) - 4 \\
&> 2(t-k) - 1
\end{aligned}$$

which means $t < \frac{3}{2}k + \frac{1}{2}$. This is equivalent to $t \le \lfloor \frac{3}{2}k + \frac{1}{2} \rfloor$ when t is an integer. ■

Now we are ready to prove the main theorem.

Theorem 3 *If a set of balls takes k bins as the optimal solution, then in the worst case FF can use $\frac{5}{3}k$ or more bins, but no more than $\lfloor \frac{7}{4}k + \frac{1}{2} \rfloor$ bins.*

Proof: To show the lower bound, we give an example. Consider the following arrangement of balls:

$$\{\frac{1}{2} + 2\epsilon, \frac{1}{2} - 2\epsilon\}, \{\frac{1}{2} - \epsilon, \frac{1}{4}, \frac{1}{4} + \epsilon\}, \{\frac{1}{2} + 2\epsilon, \frac{1}{4} + 2\epsilon, \frac{1}{4} - 4\epsilon\}$$

where ϵ is a very small positive number. This must be an OPT with 3 bins because every bin is full. With the following order of feeding balls, FF takes 5 bins, which proves the lower bound $\frac{5}{3}$:

$$\{\frac{1}{4} - 4\epsilon, \frac{1}{4} + \epsilon, \frac{1}{2} - 2\epsilon\}, \{\frac{1}{4}, \frac{1}{4} + 2\epsilon\}, \{\frac{1}{2} + 2\epsilon\}, \{\frac{1}{2} + 2\epsilon\}, \{\frac{1}{2} - \epsilon\}.$$

To show the upper bound, we would like to think of a way to use the lemma. It is easy to see that no two "big" balls (with weight $> \frac{1}{2}$) can be put into one bin. We assume in an optimal solution with k bins, there are m bins that contains a big ball. Here we list some observations and the reasons behind:

- In order to prove an upper bound of FF, we can always assume every bin in OPT is full. Suppose, on the contrary, some of the k bins are not full and for some sequence of balls FF occupies t bins, then after we add some balls to fill up all the k bins (still an OPT), we can append these newly added balls to the end of the previous sequence that FF uses. This will take up at least t bins and the competitive ratio could only be the same or increase.
- Instead of looking solely at FF, we come up with FF$'$, a "worse" version of FF, that puts big balls and small balls exclusively. To be precise, FF$'$ maintains two bunches of bins, bunch s_1 for small balls and bunch s_2 for big balls; when a ball comes, FF$'$ first decides the bunch of bins it should be put into, then for the decided bunch of bins use FF as a sub function. We claim that the upper bound on competitive ratio of FF$'$ is an upper bound on FF. To see this, we describe a process that turns a valid positioning of balls in FF to a valid positioning of balls in FF$'$, and each step in this process cannot decrease the number of bins consumed:
  - (I) For FF, we can assume it only uses bunch s_1 and occupies t bins. For any bin i that has a big ball and after which there is still some other bin that only contains small balls, suppose the final positioning of balls in it is as follows: it contains some small balls of total weight x at the bottom, a big ball of weight b in the middle, and some other small balls of total weight y on the top. We rearrange the sequence by picking out all the balls of weight $(x + b + y)$ in bin i (without changing the relative order inside), then for the remaining subsequence we still use s_1. The subsequence would use $t - 1$ bins and all these bins would have exactly the same positioning of balls as before.

(II) For the balls we picked out, we split them into two parts: small balls and the big ball. Then we append the small balls of weight $(x+y)$ to the end of the subsequence that uses s_1. For the big ball, we use bunch s_2. The big ball must take up a bin by itself. The small balls appended at the end of the subsequence might increase the number of bins in s_1 to more than $t-1$. In total, we uses $\geq t$ bins.

(III) Apply the above steps until there are no bins that has a big ball but after it there is still at least one bin that only contains small balls.

After the process, the positioning is a valid positioning as we use FF$'$ without decrementing the number of bins used.

Since we have argued that the upper bound on FF$'$ is an upper bound of FF, now we can try to upper bound the competitive ratio of FF$'$ using the lemma we have proved. Suppose OPT uses k bins and they are all full, and FF$'$ uses s bins, and there are s_1 bins for small balls and s_2 bins for big balls. Since the total weight of those big balls is greater than $\frac{1}{2}s_2$, the total weight of those small balls should be less than $k - \frac{1}{2}s_2$. Then, we have

$$
\begin{aligned}
t &= s_1 + s_2 \\
&< \lfloor \frac{3}{2}(k - \frac{1}{2}s_2) + \frac{1}{2} \rfloor + s_2 \\
&= \lfloor \frac{3}{2}k + \frac{1}{4}s_2 + \frac{1}{2} \rfloor \\
&\leq \lfloor \frac{3}{2}k + \frac{1}{4}k + \frac{1}{2} \rfloor \\
&= \lfloor \frac{7}{4}k + \frac{1}{2} \rfloor
\end{aligned}
$$

■