CS787: Advanced Algorithms	**Scribe:** Group 4
Title: hw3	

0.1 Chain Length

Let A_i be the subset $\{i, i+1, ..., n\}$, l_i be the size of original set A_i, then $l_i = |A_i| = n + 1 - i$.
Let $f(n) = f(l_1) = E(|S|)$, where S denote the random subset of elements picked in this manner, $S \subset A_1$.
Since $E(|S|) = 1 + \frac{1}{n}\sum_{j=2}^{n} E(|S_j|)$, Where $S_j \subset A_j$,
which means we randomly pick an element e, and then get the average expectation of all possible subset $\{e+1, e+2, ..., n\}$, which is $\frac{1}{n}\sum_{j=2}^{n} E(|S_j|)$.

Then $f(n) = f(l_1) = 1 + \frac{1}{n}\sum_{j=2}^{n} f(l_j) = 1 + \frac{1}{n}\sum_{k=1}^{n-1} f(k)$

$=> n(f(n) - 1) = \sum_{k=1}^{n-1} f(k)$, $(n-1)(f(n-1) - 1) = \sum_{k=1}^{n-2} f(k)$,

$=> n(f(n) - 1) = \sum_{k=1}^{n-2} f(k) + f(n-1) = (n-1)(f(n-1) - 1) + f(n-1)$,

$=> f(n) = f(n-1) + \frac{1}{n}$.

Thus $E(|S|) = f(n) = 1 + \frac{1}{2} + ... + \frac{1}{n}$

0.2 Coupon Collector

(A) A geometric distribution Geo(p) is a distribution by a random variable which counts the Bernoulli trials with success probability p until the first success, it has expected value 1/p. Let X_i denote the random variable of purchases until collecting i different kinds of toys, denote $Y_i = X_{i+1} - X_i$ $(0 \le i \le n-1)$. So Y_i is $Geo(\frac{n-i}{n})$ with expected value $\frac{n}{n-i}$.

$$E[X] = E[X_n - X_0] = E[\sum_{i=0}^{n-1}(X_{i+1} - X_i)] = \sum_{i=0}^{n-1} E[X_{i+1} - X_i] = \sum_{i=0}^{n-1} E[Y_i] = \sum_{i=0}^{n-1} \frac{n}{n-i}$$
$$= n\sum_{i=0}^{n-1} \frac{1}{n-i} = n\sum_{i=1}^{n} \frac{1}{i} = n * H_n \approx n * ln(n)$$

(B) Since X is positive, by Makov's inequality we have

$$Pr[X > 2\mu] \le \frac{E[X]}{2\mu} = \frac{\mu}{2\mu} = \frac{1}{2}$$

(C)

$$Pr[X < \frac{\mu}{2}] = 1 - Pr[can\ not\ collect\ n\ types\ within\ \frac{\mu}{2}\ purchases]$$
$$\leq 1 - Pr[collect\ only\ 1\ type\ within\ \frac{\mu}{2}\ purchases]$$
$$= 1 - (\frac{1}{n})^{\frac{\mu}{2}} = 1 - (\frac{1}{n})^{\frac{n*Hn}{2}}$$

Thus,

$$\lim_{n\to\infty} Pr[X < \frac{\mu}{2}] = 1$$

0.3 Lower Bound for Online Matching

(A) It turns out that $e/(e-1)$ is precisely the best competitive ratio that can be achieved by an online fractional matching algorithm. To prove this, we consider an arbitrary fractional matching algorithm ALG and evaluate its performance on a random input sequence generated as follows.
The graph G has vertex sets $L = R = [n] = \{1, ..., n\}$. We sample a uniformly random permutation π of the set [n], and we define the edge set of the graph to be
$E = \{(\pi_j, i) | j >= i\}$
The elements of R arrive in the order i=1, 2,...,n.

Observe first that there is always a perfect matching in G, consisting of edges (π_i, i) for i=1,...n. In fact, this is the unique perfect matching in G: one can easily show that every perfect matching must contain the edge (π_i, i) for all $i \in [n]$, by downward induction on i starting from i=n.

To place an upper bound on the expected size of the matching produced by ALG, we argue as follows. The expected value of x_{π_j}, i is zero if $j < i$, and it is at most $\frac{1}{n+1-i}$ if $j >= i$. To see this latter fact, note that for any two elements j, $k \in \{i, i+1, ..., n\}$, we have $\mathbf{E}\big[x_{\pi_j}, i\big] = \mathbf{E}[x_{\pi_k}, i]$ by symmetry, since the subgraph of G consisting of all edges observed up until time i has an automorphism that exchanges j and k. Since $x_{\pi_i}, i = x_{\pi_{i+1}}, i = ... = x_{\pi_n}, i$ and the sum of these numbers is at most 1, each of them is at most $\frac{1}{n+1-i}$.

(B) Now, let $k = n - [n/e]$, and observe that $\sum_{i=1}^{k} \frac{1}{n+1-i}$ is between $1 - \frac{5}{n}$ and 1. This is proven by the integral test:
$\sum_{i=1}^{k} \frac{1}{n+1-i} < \int_{n/e}^{n} \frac{dx}{x} = 1$
while
$\frac{5}{n} + \sum_{i=1}^{k} \frac{1}{n+1-i} > \frac{1}{n+5} + \frac{1}{n+5} + ... + \frac{1}{n+1} + \sum_{i=1}^{k} \frac{1}{n+1-i} > \int_{(n+6)/e}^{n+6} \frac{dx}{x} = 1$
The expect size of the fractional matching produced by ALG is bounded above by:
$\sum_{j=1}^{n} \mathbf{E}\Big[\sum_{i=1}^{j} x_{\pi_j}, i\Big] <= \sum_{j=1}^{k} \sum_{i=1}^{j} \frac{1}{n+1-i} + \sum_{j=k+1}^{n} 1$
$< \sum_{j=1}^{k} \sum_{i=1}^{j} \frac{1}{n+1-i} + \sum_{j=k+1}^{n} [\frac{5}{n} + \sum_{i=1}^{k} \frac{1}{n+1-i}]$

$< 5 + \sum_{i=1}^{k} \frac{(k+1-i)+(n-k)}{n+1-i}$
$= 5 + k < 5 + (1 + \frac{1}{e})n$.
As the expected size of the maximum matching is n, and the expected size of the fractional matching produced by ALG is bounded above by $5 + (\frac{e-1}{e})n$, we see that ALG cannot be c-competitive for any $c < \frac{e}{e-1}$.

0.4 Bin-Packing

(A) For 4 itmes a_1, a_2, a_3, a_4 come in order with weights $w_1 = 0.3, w_2 = 0.3, w_3 = 0.7, w_4 = 0.7$, First Fit uses 3 bins b_1, b_2, b_3 with
b_1 contains a_1, a_2
b_2 contains a_3
b_3 contains a_4
While optimal packing only uses 2 bins b_1, b_2 with
b_1 contains a_1, a_3
b_2 contains a_2, a_4

(B) Denote
FF: First Fit algorithm
OPT: Optimal solution

For the case that cost(FF) is even, we have cost(FF) = 2m, for some m $\geq$ 1.
And we have cost(OPT) $\geq$ (0.5)*(2m) = m. Thus, 2*cost(OPT) $\geq$ 2m = cost(FF).

For the case that cost(FF) is odd, we have cost(FF) = 2m+1, for some m $\geq$ 0.
And we have cost(OPT) $\geq$ (0.5)*(2m)+1 = m+1. Thus, 2*cost(OPT) $\geq$ 2m+2 > 2m+1 = cost(FF).

Now we show $\frac{12}{7}$ is a better bound than 2. (Xia & Tan, 2010)
Let

$$u = 31 * cost(OPT) - 18 * cost(FF)$$

be a diophantine equation relating to cost(OPT) and cost(FF), where u is an integer.
Since (7u, 12u) is a solution, any integral solution can be written as

$$\begin{cases} cost(OPT) = 7u + 18v \\ cost(FF) = 12u + 31v \end{cases}$$

Since cost(FF) $\leq \frac{17}{10} cost(OPT) + \frac{7}{10}$, it requires

$$u + 4v \leq 7$$

When u $\geq$ 4, we get v $\leq$ 0, so

$$\frac{cost(FF)}{cost(OPT)} = \frac{12u + 31v}{7u + 18v} = \frac{31}{18} - \frac{1}{18(7 + \frac{18v}{u})} \leq \frac{31}{18} - \frac{1}{18 * 7} = \frac{12}{7}$$

Bibliography

B. Xia and Z. Tan. Tighter bounds of the First Fit algorithm for the bin-packing problem. Discrete Appl. Math., 158:16681675, 2010.