3.7 The Stable Marriage Problem

Situation: Given n men $A, B, \ldots$ and n women $a, b, \ldots$, each man and each woman has his or her preference list about persons of opposite sex.

- A *pairing* M: a 1-1 correspondence between the men and women.
- A pairing is said to be *unstable* if there are couples $X - x$ and $Y - y$ s.t. X prefers y to x and y prefers X to Y.
- In this case, the pair $X - y$ is called *unsatisfied* under this pairing.
- A pairing in which there are no unsatisfied couples is called a *stable pairing or stable marriage.*

Example For $n = 4$ and consider the following preference lists.

$$A : abcd \quad B : bacd \quad C : adcb \quad D : dcab$$

$$a : ABCD \quad b : DCBA \quad c : ABCD \quad d : CDAB$$

A marriage: $A - a$, $B - b$, $C - c$ and $D - d$. Stable?
A marriage: $A - a$, $B - b$, $C - d$ and $D - c$. Stable?

Question: 1. For given n men $A, B, \ldots$ and n women $a, b, \ldots$ having their preference lists, is there always a stable marriage?
2. If then, how can we find a stable marriage?

Proposal algorithm

Input: n men $A, B, \ldots$ and n women $a, b, \ldots$ having their preference lists.

Output: A stable marrige

Algorithm:

1. Give an arbitrary (but fixed) ordering of the n men, the first unmatched man X proposes to the most desirable woman on his list who has not already rejected him, call her x.

2. If the woman x is unmatched or her current mate Y is less preferable to X then x will temporarily accept X (in the latter case, poor Y is jilted and reverts to the unmatched state). Otherwise reject him.

3. Repeat this process until every person has been matched.

- Is the algorithm always terminate?
- Is the algorithm always terminated with a stable pairing?

Lemma 4 *A woman once matched will stay matched during the course of the algorithm (although her mates may change with time).*

Corollary 5 *The algorithm terminates when every woman gets at least one proposal.*

Corollary 6 *If a man proposes to the least desirable woman in his list, the algorithm stops.*

Corollary 7 *The algorithm stops in at most* $n^2 - n + 1$ *steps.*

Theorem 8 *The final pairing is stable.*

Proof. Suppose that $X - y$ is a dissatisfied pair and $X - x$, $Y - y$ are couples in the final pairing M. Since X prefers y to x, he must propose to y before getting married to x. Since y either reject X or accept him only to jilt him later. So, y must be matched with Z preferable to X in the final pairing. Contradiction!! So, the final pairing must be stable.

Average case analysis Assume that men's lists are chosen independently and uniformly at random. Women's lists can be arbitrary but must fixed in advance.

Or, equivalently

Suppose that men do not know their lists to start with. Each time, a man choose a random woman from the women not already proposed by him.

- **Clock Solitaire: An example of the Principle of deferred decisions**

In this game, there is a standard deck of 52 cards, which is assumed to be randomly shuffled.

1. The pack is then divided into 13 piles of 4 cards each, labelled $1, 2, ..., 10, J, Q, K$.
2. On the first move, we draw a card from the pile labelled K.
3. At each subsequent move, a card is drawn from the pile whose label is the face value of the card drawn at the previous move.
4. The game ends when an attempt is made to draw a card from an empty pile.
5. We win the game, all 52 cards have been drawn. Otherwise, we lose.

What is the probability of wining?

At each draw any unseen card is equally likely to appear. Thus, the process of playing this game is exactly equivalent to repeatedly drawing a card uniformly at random from a deck of 52 cards.

Lemma 9 *A winning game corresponds to the situation where the first 51 cards drawn contain exactly 3 Kings.*

Corollary 10

$$\Pr[\, wining \,] = 1/13.$$

Average case analysis Assume that men's lists are chosen independently and uniformly at random. Women's lists can be arbitrary but must fixed in advance. Or, equivalently

Suppose that men do not know their lists to start with. Each time, a man choose a random woman from the women not already proposed by him.

Amnesiac Algorithm

Each time a man makes a proposal, he chooses a woman uniformly at random from the set of all n women, including those to whom he has already proposed.

If he chooses a woman he has already proposed, then he will be rejected as she has already rejected him before. If not, the woman is chosen uniform at random from the women not already proposed by him. In other words, the final pairing would be the same as the original algorithm.

Clearly, the number T_A of steps for the amnesiac algorithm is at least as large as that of the usual proposal algorithm.

Using Corollary 5, one may show that

Theorem 11 *For any constant c, and $m = n \ln n + cn$,*

$$\Pr[T_A > m] = 1 - e^{-e^{-c}}.$$

(See the next section for the proof.)

3.8 The coupon Collector's Problem

There are n types of coupons and at each trial a coupon is chosen uniformly at random.

How many trials are needed to get all coupons?

General example of waiting for combinations of events to happen.
Expected case analysis:

3.8.1 An Elementary Analysis

For any $0 \leq i \leq n-1$,

$$X_i = \text{number of trials to get } (i+1)^{\text{th}} \text{ new coupon after getting } i \text{ coupons.}$$

Then, $X = \sum_{i=0}^{n-1} X_i$ is the random variable representing the number of trials needed to get all coupons.

- Distribution of X_i:

$$\Pr[X_i = \ell] = p_i(1 - p_i)^{\ell-1},$$

where the success probability $p_i = \frac{n-i}{n}$.

That is, X_i is geometrically distributed with parameter p_i.

In particular,

$$E[X_i] = \sum_{\ell=1}^{\infty} \ell p_i(1 - p_i)^{\ell-1} = 1/p_i = \frac{n}{n-i},$$

and hence

$$E[X] = \sum_{i=0}^{n} \frac{n}{n-i} = n \sum_{i=1}^{n} \frac{1}{i} = nH_n = n \log n + O(n).$$

For the variance of X, notice that X_i are independent. thus

$$\sigma^2(X) = \sum_{i=1}^{n} \sigma^2(X_i).$$

As

$$\sigma^2(X_i) = \sum_{\ell=1}^{\infty} \ell^2 p_i (1-p_i)^{\ell-1} - \frac{1}{p_i^2} = \frac{1-p_i}{p_i^2} = \frac{in}{(n-i)^2},$$

we have

$$\begin{aligned}\sigma^2(X) &= \sum_{i=0}^{n-1} \frac{in}{(n-i)^2} \sum_{i=1}^{n} \frac{n(n-i)}{i^2} \\ &= n^2 \sum_{i=1}^{n} \frac{1}{i^2} - nH_n \\ &= (1+o(1))n^2 \sum_{i=1}^{n} \frac{1}{i^2}.\end{aligned}$$

Therefore, the Chebyschev inequality gives

$$X = n \ln n + O(n)$$

with high probability.

3.8.2 The Poisson Approximation

- Properties of Poisson random variables

If X, Y are independent $\mathrm{Poi}(\lambda)$ and $\mathrm{Poi}(\mu)$, respectively, then $X + Y$ is a $\mathrm{Poi}(\lambda + \mu)$.

Pf.

$$
\begin{aligned}
&\Pr[X + Y = j] \\
&= \sum_{i=0}^{k} \Pr[X = i, Y = j - i] \\
&= \sum_{i=0}^{j} e^{-\lambda} \frac{\lambda^i}{i!} e^{-\mu} \frac{\mu^{j-i}}{(j-i)!} \\
&= e^{-(\lambda+\mu)} \frac{(\lambda+\mu)^j}{j!} \sum_{i=0}^{j} \frac{j!}{i!(j-i)!} \Big(\frac{\lambda}{\lambda+\mu}\Big)^i \Big(\frac{\mu}{\lambda+\mu}\Big)^{j-i} \\
&= e^{-(\lambda+\mu)} \frac{(\lambda+\mu)^j}{j!} \Big(\frac{\lambda}{\lambda+\mu} + \frac{\mu}{\lambda+\mu}\Big)^j .
\end{aligned}
$$

The ‘converse’ is also true.

The ‘converse’ is also true.

Let W be a Poi(ρ). We take W balls and color each ball red with probability p and blue with probability $1-p$, independently of the others. Let $\lambda = p\rho$ and $\mu = (1-p)\rho$. Then the numbers X, Y of red and blue balls, respectively, are independent Poi(λ) and Poi(μ), respectively.

Proof. We need to show that

$$\Pr[X = i, Y = j] = e^{-\lambda} \frac{\lambda^i}{i!} e^{-\mu} \frac{\mu^i}{j!}.$$

$$\begin{aligned}
\Pr[X = i, Y = j] &= \Pr[W = i + j] \binom{i+j}{i} p^i (1-p)^j \\
&= e^{-\rho} \frac{\rho^{i+j}}{(i+j)!} \binom{i+j}{i} p^i (1-p)^j \\
&= e^{-\rho} \frac{(p\rho)^i}{i!} \frac{((1-p)\rho)^j}{j!}.
\end{aligned}$$

Using $\rho = \lambda + \mu$ and $p\rho = \lambda, (1-p)\rho = \mu$, we have that

$$\Pr[X = i, Y = j] = e^{-\lambda-\mu} \frac{\lambda^i}{i!} \frac{\mu^j}{j!}$$

Generally,

if X_i's are independent Poi(λ_i)'s
then $\sum X_i$ is a Poi($\sum \lambda_i$).

Conversely,

Let W be a Poi(ρ).
Take W balls and color each ball i with
probability p_i, $\sum p_i = 1$, independently of the others. Denote X_i to
be the numbers of balls colored i.
Then X_i's are independent Poi(λ_i)'s, where $\lambda_i = p_i \rho$,

3.9 The Coupon Collector's Problem vs. The Occupancy Problem

Occupancy Problems:

Insert each of m balls to n
distinct bins uniformly at random.

Theorem 12 *If $m = n \ln n + cn$, then*

$$\Pr[\ \exists \text{ empty bin }] \to 1 - e^{-e^{-c}},$$

(as $n \to \infty$).

Corollary 13 *For the number of trials X for the coupon collector's problem and $m = n \ln n + cn$,*

$$\Pr[X > m] \to 1 - e^{-e^{-c}}.$$

Proof. Take a Poisson random variable M_1 with mean $m_1 = m - n^{1/2} \ln^2 n$. Notice that

$$\Pr[M_1 \geq m] \to 0.$$

We choose M_1 balls and insert each of them to the n bins uniformly at random. Then, the numbers Y_i of balls in the i^{th} bins are i.i.d Poisson $\lambda_1 := m_1/n = \ln n + c + o(1)$ random variables. Thus

$$\begin{aligned}\Pr[\ \exists \text{ empty bin }\ ] &\leq \Pr_1[\ \exists \text{ empty bin} | M_1 \leq m] \\ &\leq (1+o(1))(1 - \Pr[Y_i > 0 \ \ \forall\ i])\end{aligned}$$

For

$$\Pr[Y_i > 0 \ \ \forall\ i] = \Pr[Y_1 > 0]^n = (1 - e^{-\lambda_1})^n,$$

and

$$e^{-\lambda_1} = e^{-\ln n - c + o(1)} = \frac{(1+o(1))e^{-c}}{n},$$

we have

$$(1 - e^{-\lambda_1})^n = (1+o(1))e^{-e^{-c}}.$$

Therefore,

$$\Pr[\ \exists \text{ empty bin }\ ] \leq (1 + o(1))(1 - e^{e^{-c}}).$$

Similarly, we may take a Poisson random variable M_2 with mean $m_2 = m + n^{1/2} \ln^2 n$, and choose M_2 balls to obtain

$$\begin{aligned} \Pr[\ \exists \text{ empty bin }\ ] &\geq \Pr_2[\ \exists \text{ empty bin} | M_2 \geq m] \\ &\geq 1 - \Pr[Z_i > 0 \ \ \forall\ i] + o(1), \end{aligned}$$

for the numbers Z_i of balls in the i^{th} bins, which are i.i.d. Poisson random variables with mean $\lambda_2 = m_2/n = \ln n + c + o(1)$.

□