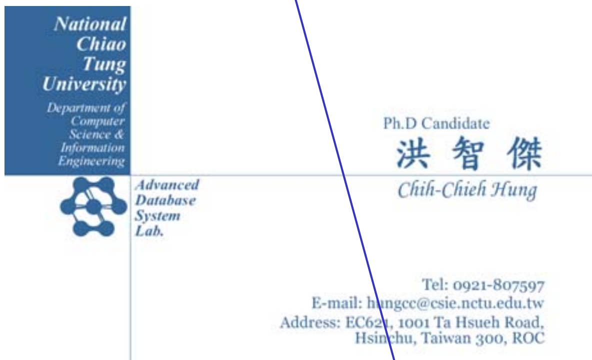
The Stable Marriage Problem

Original author: S. C. Tsai (交大蔡錫鈞教授)

Revised by Chuang-Chieh Lin and Chih-Chieh Hung

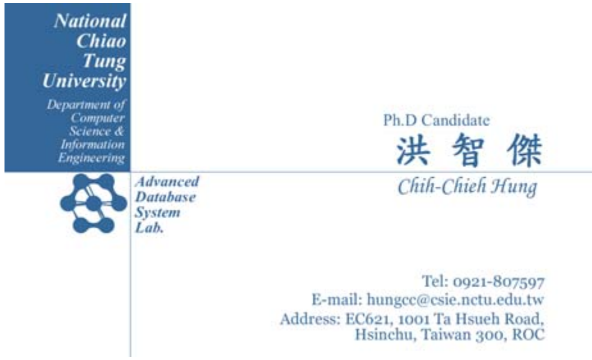
A partner from NCTU ADSL (**A**dvanced **D**atabase **S**ystem **L**ab)



- Chih-Chieh Hung (洪智傑)
  - Bachelor Degree in the Department of Applied Math, National Chung-Hsing University. (1996-2000)
  - Master Degree in the Department of Computer Science and Information Engineering, NCTU. (2003-2005)
  - Ph.D. Student (博一) in the Department of Computer Science and Information Engineering, NCTU. (since 2005)
  - Will be a Ph.D. candidate.

His master diploma.

A partner from NCTU ADSL (**A**dvanced **D**atabase **S**ystem **L**ab)



- Research topics
  - Data mining
  - Mobile Data Management
  - Data management on sensor networks
- Advisor:
  - Professor Wen-Chih Peng (彭文志)
- Personal state:
  - The master of <u>CSexam.Math</u> forum on Jupiter BBS.
  - Single, but he has a girlfriend now.

- Consider a society with n men (denoted by capital letters) and n women (denoted by lower case letters).
- A marriage M is a 1-1 correspondence between the men and women.
- Each person has a preference list of the members of the opposite sex organized in a decreasing order of desirability.

- A marriage is said to be *unstable* if there exist 2 marriage couples *X*-*x* and *Y*-*y* such that *X* desires *y* more than *x* and *y* desires *X* more than *Y*.

- The pair *X*-*y* is said to be "dissatisfied." (不滿的)

- A marriage *M* is called "*stable marriage*" if there is no dissatisfied couple.

- Assume a monogamous, hetersexual society.
- For example, $N = 4$.

  A: *abcd* *B*: *bacd* *C*: *adcb* *D*: *dcab*

  a: *ABCD* *b*: *DCBA* *c*: *ABCD* *d*: *CDAB*

- Consider a marriage *M*: *A*-*a*, *B*-*b*, *C*-*c*, *D*-*d*,
- *C*-*d* is dissatisfied. Why?

Proposal algorithm:

Assume that the men are numbered in some arbitrary order.

- The lowest numbered unmarried man *X* proposes to the most desirable woman on his list who has not already rejected him; call her *x*.

- The woman *x* will accept the proposal if she is currently unmarried, or if her current mate *Y* is less desirable to her than *X* (*Y* is jilted and reverts to the unmarried state).
- The algorithm repeats this process, terminating when every person has married.
- (This algorithm is used by hospitals in North America in the match program that assigns medical graduates to residency positions.)

Does it always terminate with a stable marriage?

- An unmatched man always has at least one woman available that he can proposition.

- At each step the proposer will eliminate one woman on his list and the total size of the lists is n^2. Thus the algorithm uses at most n^2 proposals. i.e., it always terminates.

Claim that the final marriage *M* is stable.

- Proof by contradiction:
  - Let *X*-*y* be a dissatisfied pair, where in *M* they are paired as *X*-*x*, *Y*-*y*.
  - Since *X* prefers *y* to *x*, he must have proposed to *y* before getting married to *x*.

- Since *y* either rejected *X* or accepted him only to jilt (拋棄) him later, her mates thereafter (including *Y*) must be more desirable to her than *X*.

- Therefore *y* must prefer *Y* to *X*, $\rightarrow\leftarrow$ contradicting the assumption that *y* is dissatisfied.

■

- **Goal**: Perform an average-case analysis of this (deterministic) algorithm.

- For this average-case analysis, we assume that the men's lists are chosen independently and uniformly at random; the women's lists can be arbitrary but must be fixed in advance.

- T_P denotes the number of proposal made during the execution of the Proposal Algorithm. The running time is proportional to T_P.

- But it seems difficult to analyze T_P.

- Principle of ***Deferred Decisions***:
  - The idea is to assume that the entire set of random choices is ***not*** made in advance.
  - At each step of the process, we fix only ***the random choices*** that must be revealed to the algorithm.
- We use it to simplify the average-case analysis of the Proposal Algorithm.

- Suppose that men do not know their lists to start with. Each time a man has to make a proposal, he picks a random woman from the set of women not already propositioned by him, and proceeds to propose to her.
- The only dependency that remains is that the random choice of a woman at any step depends on the set of proposals made so far by the current proposer.

- However, we can eliminate the dependency by modifying the algorithm, i.e., a man chooses a woman uniformly at random from the set of all n women, including those to whom he has already proposed.

- He *forgets* the fact that these women have already rejected him.

- Call this new version the *Amnesiac Algorithm*.

- Note that a man making a proposal to a woman who has already rejected him will be rejected again.

- Thus the output by the Amnesiac Algorithm is exactly the same as that of the original Proposal Algorithm.

- The only difference is that there are some wasted proposals in the Amnesiac Algorithm.

- Let T_A denote the number of proposals made by the Amnesiac Algorithm.

$T_P > m \Rightarrow T_A > m$, i.e., T_A stochastically dominates T_P.

That is, $\mathbf{Pr}[T_P > m] \leq \mathbf{Pr}[T_A > m]$ for all m.

- It suffices to find an upper bound to analyze the distribution T_A.
- A benefit of analyzing T_A is that we need only count that total number of proposals made, without regard to the name of the proposer at each stage.
- This is because each proposal is made uniformly and independently to one of n women.

- The algorithm terminates with a stable marriage once all women have received at least one proposal each.

- Moreover, bounding the value of T_A is a special case of the *coupon collector's problem*.

- **Theorem**: ([MR95, page 57])

  For any constant $c \in R$, and $m = n \ln n + cn$,

$$\lim_{n \to \infty} \mathbf{Pr}[T_A > m] = 1 - e^{-e^{-c}} \to 0.$$

- The Amnesiac Algorithm terminates with a stable marriage once all women have received at least one proposal each.

- Bounding the value of T_A is a special case of the *coupon collector's problem*.

The Coupon Collector's Problem

- Input: Given *n* types of coupons. At each trial a coupon is chosen at random. Each random choice of the coupons are mutually independent.

- Output: The *minimum number of trials required* to collect at least one of each type of coupon.

- You may regard this problem as "Hello Kitty Collector's Problem".

- Let X be a random variable defined to be the *number of trials* required to collect at least one of each type of coupon.

- Let $C_1, C_2, \ldots, C_X$ denote the sequence of trials, where $C_i \in \{1, \ldots, n\}$ denotes the *type* of the coupon drawn in the ith trial.

- Call the *i*th trial C_i a success if the type C_i was not drawn in any of the first $i - 1$ selections.
- Clearly, C_1 and C_X are always successes.
- We consider dividing the sequence into *epochs* (時期), where epoch *i* begins with the trial following the *i*th success and ends with the trial on which we obtain the (*i*+1)st success.

What kind of probability distribution does X_i possess?

- Define the random variable X_i, for $0 \le i \le n-1$, to be the number of trials in the ith stage (epoch), so that

$$X = \sum_{i=0}^{n-1} X_i.$$

- Let p_i denote the probability of success on any trial of the i-th stage.
  - This is the probability of drawing one of the $n-i$ remaining coupon types and so,

$$p_i = \frac{n-i}{n}.$$

Note that *binomial distribution and geometric distribution* are very, very important.

- Recall that X_i is geometrically distributed with p_i.
- So $\mathbf{E}[X_i] = 1/p_i, \quad \sigma^2_{X_i} = (1-p_i)$.
- Thus $\mathbf{E}[X] = \mathbf{E}[\sum_{i=0}^{n-1} X_i] = \sum_{i=0}^{n-1} \mathbf{E}[X_i] = \sum_{i=0}^{n-1} \frac{1}{p_i}$

$$= \sum_{i=0}^{n-1} \frac{n}{n-i} = n \sum_{i=1}^{n} \frac{1}{i} = nH_n.$$

$H_n = \ln(n) + \Theta(1)$

i.e.,
$\mathbf{E}[X] = n\ln(n) + O(n)$

- X_i's are independent, thus

$$\sigma_X^2 = \sum_{i=0}^{n-1} \sigma_{X_i}^2$$

$$= \sum_{i=0}^{n-1} \frac{ni}{(n-i)^2}$$

$$= \sum_{i'=1}^{n} \frac{n(n-i')}{i'^2}$$

$$= n^2 \boxed{\sum_{i'=1}^{n} \frac{1}{i'^2}} - nH_n.$$

$\pi^2/6$

Exercise

- Use the Chebyshev's inequality to find an upper bound on the probability that $X > \beta n \ln n$, for a constant $\beta > 1$.
  - Try to prove that

$$\mathbf{Pr}[X \geq \beta n \ln n] \leq O\left(\frac{1}{\beta^2 \ln^2 n}\right).$$

(You might need the result: $n \ln n \leq nH_n \leq n \ln n + n$.)

Remark: Chebyshev's Inequality

Let X be a random variable with expectation μ_X and standard deviation σ_X. Then for any $t \in \mathbf{R}^+$,

$$\mathbf{Pr}[|X - \mu_X| \geq t\sigma_X] \leq \frac{1}{t^2}.$$

or equivalently,

$$\mathbf{Pr}[|X - \mu_X| \geq t] \leq \frac{\sigma_X^2}{t^2}.$$

- Our next goal is to derive *sharper* estimates of the typical value of X.

- We will show that the value of X *is unlikely to deviate far from its expectations*, or, is sharply concentrated around its expected value.

- Let ξ_i^r denote the event that coupon type *i* is not collected in the first *r* trials.
- Thus $\mathbf{Pr}[\xi_i^r] = (1 - \frac{1}{n})^r \leq e^{-r/n}$.
- For $r = \beta n \ln(n)$, $e^{-r/n} = n^{-\beta}$, $\beta > 1$.

$$\mathbf{Pr}[X > r] = \mathbf{Pr}[\bigcup_{i=1}^{n} \xi_i^r]$$

$$\leq \sum_{i=1}^{n} \mathbf{Pr}[\xi_i^r] \leq \sum_{i=1}^{n} n^{-\beta} = n^{-(\beta-1)}.$$

It is still polynomially small.

- So that’s it?
- Is the analysis good enough?
- Not yet!
- Let consider the following heuristic argument which will help to establish some intuition.

Poisson Heuristic

The crucial idea!

- Let N_i^r denote the number of times the coupon of type *i* is chosen during the first *r* trials.
- ξ_i^r is the same as the event $\{N_i^r = 0\}$.
- N_i^r has the binomial distribution with parameter *r* and *p* = 1/*n*.

$$\Rightarrow \mathbf{Pr}[N_i^r = x] = \binom{r}{x} p^x (1-p)^{r-x}.$$

Recall of the Poisson distribution

- Let λ be a positive real number.
- Y: a non-negative integer r.v.

- Y has the Poisson distribution with parameter λ if for any non-negative integer y,

$$\mathbf{Pr}[Y = y] = \frac{\lambda^y e^{-\lambda}}{y!}.$$

- For proper small λ and as $r \to \infty$, the Poisson distribution with $\lambda = rp$ is **a good approximation to the binomial distribution** with parameter r and p.

- Approximate N_i^r by the Poisson distribution with parameter $\lambda = r/n$ since $p = 1/n$.

- Thus, $\mathbf{Pr}[\xi_i^r] = \mathbf{Pr}[N_i^r = 0] \approx \frac{\lambda^0 e^{-\lambda}}{0!} = e^{-r/n}$.

- **Claim**: ξ_i^r, for $1 \le i \le n$, are *almost* independent.
i.e., for any index set $\{j_1, \ldots, j_k\}$ not containing i,

$$\mathbf{Pr}[\xi_i^r \mid \bigcap_{l=1}^{k} \xi_{j_l}^r] = \mathbf{Pr}[\xi_i^r].$$

- Proof:

$$\mathbf{Pr}[\xi_i^r \mid \bigcap_{l=1}^{k} \xi_{j_l}^r] = \frac{\mathbf{Pr}[\xi_i^r \cap (\bigcap_{l=1}^{k} \xi_{j_l}^r)]}{\mathbf{Pr}[\bigcap_{l=1}^{k} \xi_{j_l}^r]} = \frac{(1 - \frac{k+1}{n})^r}{(1 - \frac{k}{n})^r}$$

$$\approx \frac{e^{-r(k+1)/n}}{e^{-rk/n}} = e^{-r/n}.$$

■

Remark: $\mathbf{Pr}[\xi_i^r] \approx e^{-r/n}$.

- Thus,

$$\mathbf{Pr}[\neg \bigcup_{i=1}^{n} \xi_i^m] = \mathbf{Pr}[\bigcap_{i=1}^{n} (\neg \xi_i^m)] \approx (1 - e^{-m/n})^n$$

$$\approx e^{-ne^{-m/n}}.$$

- Let $m = n(\ln(n) + c)$, for any constant *c*.

$$\mathbf{Pr}[X > m] = \mathbf{Pr}[\bigcup_{i=1}^{n} \xi_i^m] = 1 - \mathbf{Pr}[\neg \bigcup_{i=1}^{n} \xi_i^m]$$

$$\approx 1 - e^{-ne^{-m/n}} = 1 - e^{-e^{-c}}.$$

→ 0 for large positive *c*.

→ 1 for large negative *c*.

More Explanations for the Previous Equation:

- Since $m = n(\ln(n) + c)$, we have

$$
\begin{aligned}
&1 - e^{-ne^{-m/n}} \\
&= 1 - e^{-ne^{-\ln n - c}} \\
&= 1 - e^{-ne^{-\ln n} \cdot e^{-c}} \\
&= 1 - e^{-ne^{\ln n^{-1}} \cdot e^{-c}} \\
&= 1 - e^{-n \cdot \frac{1}{n} \cdot e^{-c}} \\
&= 1 - e^{-e^{-c}}.
\end{aligned}
$$

It is ***exponentially*** close to 0 as the value of positive *c* increases.

The Power of Poisson Heuristic

- It gives a quick back-of-the-envelope type estimation of probabilistic quantities, which *hopefully* provides some *insight* into the *true behavior* of those quantities.

- Poisson heuristic can help us do the analysis better.

But…

- However, it is not rigorous enough since it only *approximates* N_i^r.
- We can convert the previous argument into a rigorous proof using *the Boole-Bonferroni Inequalities*. (Yet the analysis will be more complex.)

- Are you ready to be rigorous?
- Tighten your seat belt!

Take a break! (感謝物理系黃教授提供)

- 「天母」地名的由來：
- 話說以前美軍曾在台北駐軍。某一日當他們行經一地時，詢問當地居民說：
  - “Where is it?”
- 當地居民看到阿豆仔，聽不懂他們講什麼，紛紛回答說：
  - 「聽無啦！」
- 美軍這時恍然大悟，從此以後就給這地方取了一個名字，叫做“Tien-Mu”.

A Rigorous Analysis

- **Theorem 1**: Let *X* be a random variable defined to be the number of trials for collecting each of the *n* types of coupons. Then, for any constant *c* and $m = n \ln n + cn$,

$$\lim_{n\to\infty} \mathbf{Pr}[X > m] = 1 - e^{-e^{-c}}.$$

- Proof: Let $P_k^n = \sum_{1 \le i_1 < i_2 < \ldots < i_k \le n} \mathbf{Pr}[\bigcap_{j=1}^{k} \xi_{i_j}^m].$

Remark: ξ_i^r denotes the event that coupon type *i* is not collected in the first *r* trials.

- Note that the event $\{X > m\} = \bigcup_{i=1}^{n} \xi_i^m$.

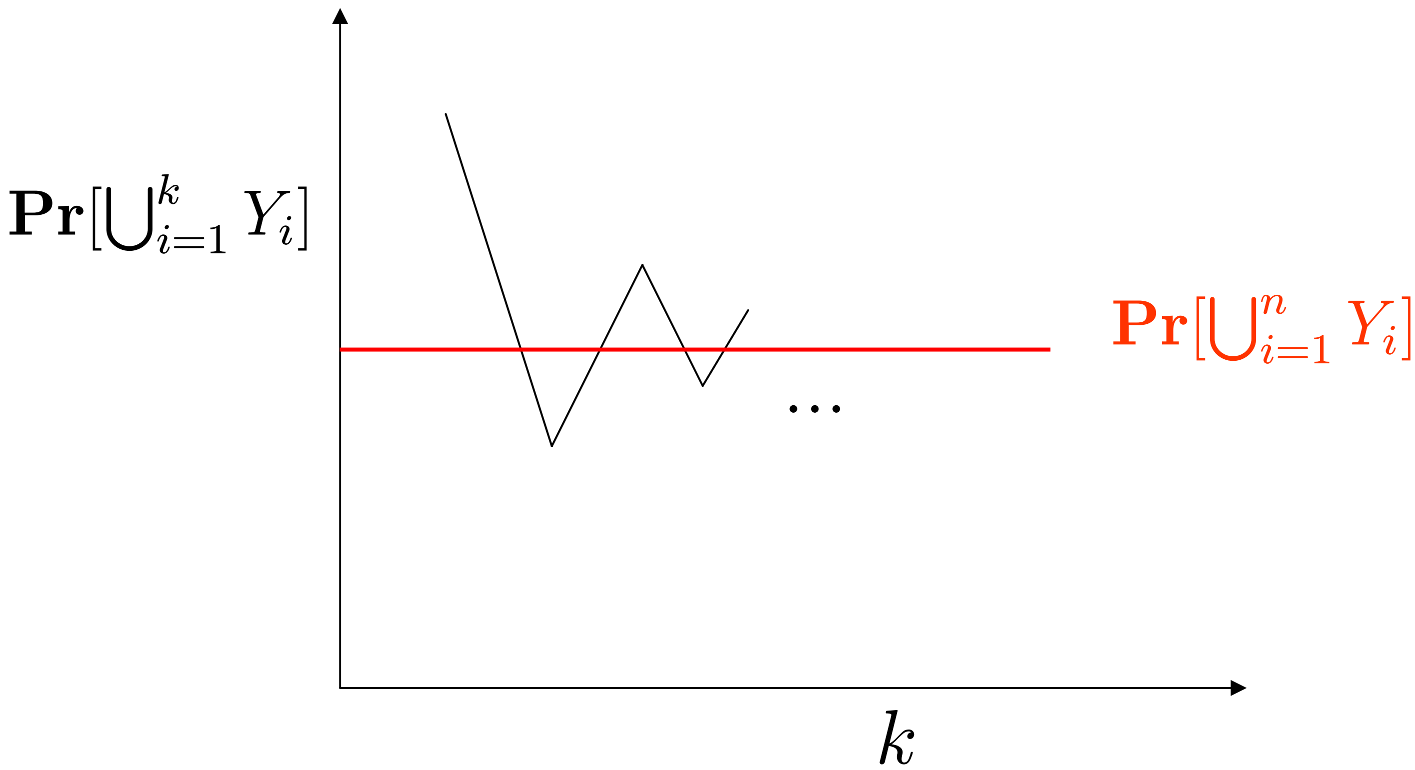
$$\mathbf{Pr}[\bigcup_i \xi_i^m] = \sum_{k=1}^{n} (-1)^{k+1} P_k^n.$$ → By the principle of Inclusion-Exclusion

- Let $S_k^n = P_1^n - P_2^n + P_3^n - \ldots + (-1)^{k+1} P_k^n$ denotes the partial sum formed by the first k terms of this series.

- We have $S^n_{2k} \le \mathbf{Pr}[\bigcup_i \xi_i^m] \le S^n_{2k+1}$ by the Boole-Bonferroni inequalities:
- $Y_1, \ldots, Y_n$: arbitrary events.

1. For odd k:
$$\mathbf{Pr}[\textstyle\bigcup_{i=1}^n Y_i] \le \sum_{j=1}^{k} (-1)^{j+1} \sum_{i_1<i_2<\ldots<i_j} \mathbf{Pr}[\bigcap_{r=1}^{j} Y_{i_r}].$$
2. For even k:
$$\mathbf{Pr}[\textstyle\bigcup_{i=1}^n Y_i] \ge \sum_{j=1}^{k} (-1)^{j+1} \sum_{i_1<i_2<\ldots<i_j} \mathbf{Pr}[\bigcap_{r=1}^{j} Y_{i_r}].$$

Illustration for the Boole-Bonferroni inequalities



- By symmetry, all the *k*-wise intersections of the events ξ_i^m are all equally likely, i.e.

$$P_k^n = \binom{n}{k}\mathbf{Pr}[\bigcap_{i=1}^k \xi_i^m].$$

- More precisely,

According to Lemma 1

$$P_k^n = \binom{n}{k}(1 - \frac{k}{n})^m \longrightarrow \boxed{e^{-ck}/k!.}$$

- For all positive integer *k*, define $P_k = e^{-ck}/k!.$

- Define the *partial sum* of P_k's as

$$S_k = \sum_{j=1}^{k} (-1)^{j+1} P_j = \sum_{j=1}^{k} (-1)^{j+1} \frac{e^{-cj}}{j!},$$

  the first k terms of the power series expansion of $f(c) = 1 - e^{-e^{-c}}$.

  Hint: Consider $g(x) = 1 - e^{-x}$ first.

- Thus $\lim_{k \to \infty} S_k = f(c)$.

- i.e., for all $\epsilon > 0$, there exists k^* such that for $k > k^*$, $|S_k - f(c)| < \epsilon$.

Remark: $S_k^n = P_1^n - P_2^n + P_3^n - \ldots + (-1)^{k+1} P_k^n.$

$$S_k = \sum_{j=1}^{k} (-1)^{j+1} P_j = \sum_{j=1}^{k} (-1)^{j+1} \frac{e^{-cj}}{j!},$$

- Since $\lim_{n \to \infty} P_k^n = P_k$, we have $\lim_{n \to \infty} S_k^n = S_k$.

- Thus for all $\epsilon > 0$ and $k > k^*$, when n is su± -ciently large, $|S_k^n - S_k| < \epsilon$.

- Thus for all $\epsilon > 0$ and $k > k^*$, and large enough n, we have $|S_k^n - S_k| < \epsilon$ and $|S_k - f(c)| < \epsilon$ which implies that

$$|S_k^n - f(c)| < 2\epsilon \text{ and } |S_{2k}^n - S_{2k+1}^n| < 4\epsilon.$$

Remark: (1) $S_{2k}^n \leq \mathbf{Pr}[\bigcup_i \xi_i^m] \leq S_{2k+1}^n$

(2) $|S_k^n - f(c)| < 2\epsilon$ and $|S_{2k}^n - S_{2k+1}^n| < 4\epsilon$,

$S_{2k}^n \downarrow f(c)$ $\quad S_{2k+1}^n \downarrow f(c)$

- Using the bracketing property of partial sum, we have that for any $\epsilon > 0$ and n su± ciently large,

$$|\mathbf{Pr}[\bigcup_i \xi_i^m] - f(c)| < 4\epsilon.$$

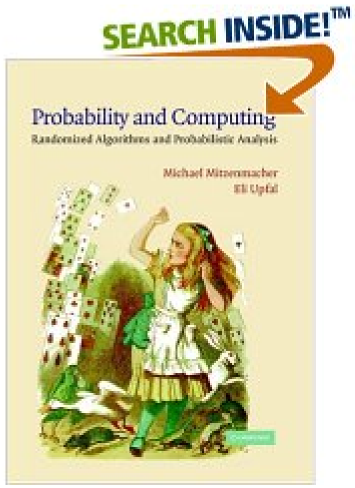
$$\Rightarrow \lim_{n\to\infty} \mathbf{Pr}[\bigcup_i \xi_i^m] = f(c) = 1 - e^{-e^{-c}}.$$

■

Thank you.

References

- [MR95] Rajeev Motwani and Prabhakar Raghavan, *Randomized algorithms*, Cambridge University Press, 1995.
- [MU05] Michael Mitzenmacher and Eli Upfal, *Probability and Computing - Randomized Algorithms and Probabilistic Analysis*, Cambridge University Press, 2005.



- **Lemma 1**: Let c be a real constant, and $m = n \ln n + cn$ for positive integer n. Then, for any fixed positive integer k,

$$\lim_{n\to\infty} \binom{n}{k}\left(1-\frac{k}{n}\right)^m = \frac{e^{-ck}}{k!}.$$

- Proof:
  - Homework:

    Prove $e^t(1-\frac{t^2}{n}) \le (1+\frac{t}{n})^n \le e^t$, for all *t*, *n* such that $n \ge 1$ and $|t| \le n$.

- By the above, we have

$$e^{-km/n}(1-\frac{k^2}{n})^{m/n} \le (1-\frac{k}{n})^m \le e^{-km/n}.$$

Remark: $m = n\ln n + cn$

- Observe that $e^{-km/n} = n^{-k}e^{-ck}$.
- Further, $\lim_{n\to\infty}(1-\frac{k^2}{n})^{m/n} = 1$ and for large n,

$$\binom{n}{k} \approx \frac{n^k}{k!}.$$

$$\therefore \lim_{n\to\infty}\binom{n}{k}(1-\frac{k}{n})^m = \lim_{n\to\infty}\frac{n^k}{k!}(1-\frac{k}{n})^m$$

$$= \lim_{n\to\infty}\frac{n^k}{k!}e^{-km/n} = \lim_{n\to\infty}\frac{n^k}{k!}n^{-k}e^{-ck} = \frac{e^{-ck}}{k!}.$$

Principle of Inclusion-Exclusion

- Let $Y_1, Y_2, \ldots Y_n$ be arbitrary events. Then

$$\mathbf{Pr}[\bigcup_{i=1}^{n} Y_i] = \sum_{i} \mathbf{Pr}[Y_i] - \sum_{i<j} \mathbf{Pr}[Y_i \cap Y_j] +$$

$$\sum_{i<j<k} \mathbf{Pr}[Y_i \cap Y_j \cap Y_k] - \ldots + (-1)^{l+1} \sum_{r=1}^{l} \mathbf{Pr}[Y_{i_r}]$$

$$+ \ldots$$

Taylor Series and Maclaurin Series

- A *Taylor series* is a series expansion of a function about a point. A one-dimensional Taylor series is an expansion of a real function $f(x)$ about a point $x = a$ is given by
$f(x) = f(a) + f'(a)(x - a) + \frac{f''(a)}{2!}(x - a)^2 + \frac{f'''(a)}{3!}(x - a)^3 + \ldots + \frac{f^{(n)}(a)}{n!}(x - a)^n + \ldots.$

- A *Maclaurin series* is a Taylor series expansion of a function about 0, i.e.,
$f(x) = f(0) + f'(0)(x) + \frac{f''(0)}{2!}(x)^2 + \frac{f'''(0)}{3!}(x)^3 + \ldots + \frac{f^{(n)}(0)}{n!}(x)^n + \ldots.$

Taylor Series and Maclaurin Series - About $f(x) = 1 - e^{-e^{-x}}$

- Using Maclaurin series, we can write e^x as
$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \ldots = \sum_{i=0}^{\infty} \frac{x^i}{i!}.$$
- So $1 - e^{-x} = x - \frac{x^2}{2!} + \frac{x^3}{3!} - \ldots = \sum_{i=1}^{\infty} (-1)^{i+1} \frac{x^i}{i!}.$
- Let $g(x) = 1 - e^{-x}$, so $g(e^{-x}) = \sum_{i=1}^{\infty} (-1)^{i+1} \frac{e^{-xi}}{i!}.$
- Let $f(x) = g(e^{-x})$, then
$$f(x) = 1 - e^{-e^{-x}} = \sum_{i=1}^{\infty} (-1)^{i+1} \frac{e^{-xi}}{i!}.$$