SPRING 2012: **COT 6936** Topics in Algorithms
Notes on Probability
Giri Narasimhan

These notes are (mostly) compiled from (a) *Probability and Computing* by Mitzenmacher and Upfal, (b) *Randomized Algorithms* by Motwani and Raghavan, (c) *Introduction to Algorithms* by Cormen, Leiserson, Rivest and Stein, (d) *The Analysis of Algorithms* by Purdom and Brown, and (e) *Algorithm Design* by Kleinberg and Tardos. **This is an evolving document that was started in Jan 2010.**

1 Useful Terms and Concepts

Sample Space It is a set of elementary events to be thought of as possible outcomes of an experiment. Probabilities are defined in terms of sample spaces.

Probability Distribution on a sample space is a mapping from events in the sample space to probabilities (i.e., real numbers from $[0, 1]$ satisfying probability axioms).

(Discrete) Random Variable (Abbreviated as r.v.) is a function from the sample space (assumed to be finite or countably infinite) to real numbers associated with outcomes of an experiment.

Bernoulli or indicator random variable Associated with events and has a value of 1 (or 0) if the event occurs (does not occur).

Probability Density Function of the random variable X is the function: $f(x) = Pr\{X = x\}$.

Expectation $E[X] = \sum_i iPr(X = i)$.

Alternative formulation If X only takes on non-negtaive integer values, then $E[X] = \sum_{i=1}^{\infty} Pr(X \geq i)$.

Binomial random variable $B(n, p)$ is the r.v. representing the number of successes in n independent experiments, each with probability p of success. $Pr(X = j) = \binom{n}{j} p^j (1-p)^{n-j}$. Expected value is np.

Geometric random variable is the r.v. representing the number of experiments before success is attained when the probability of success is p on each experiment. $Pr(X = n) = (1-p)^{n-1}p$. Expected value is $1/p$. Variance is $\frac{1-p}{p^2}$.

k-th moment of a r.v. $X = E[X^k]$.

Variance and Covariance $Var[X] = E[(X - E[X])^2] = E[X^2] - (E[X])^2$ and $Cov(X, Y) = E[(X - E[X])(Y - E[Y])]$.

Standard Deviation $\sigma[X] = \sqrt{Var[X]}$.

Monte Carlo Algorithms are randomized algorithms that might fail. They are often referred to as "always fast, not always correct".

Las Vegas Algorithms are randomized algorithms that are always correct. They are often referred to as "always correct, not always fast". A Monte Carlo algorithm can be turned into a Las Vegas algorithm by repeating it until it succeeds.

Bernoulli trials Sum of independent 0-1 iid r.v.s. Thus they are a special case of Poisson trials.

Poisson Trials Sum of independent 0-1 r.v.s. The r.v.s don't have to be iids.

Poisson distribution is often thought of as the distribution of rare events, e.g., the number of accidents or lotteries per person. It is different from the Poisson trials defined above. When we throw m balls randomly into n bins, the probability that a bin has r balls is approximately the Poisson distribution with mean m/n. It is also thought of as the limit of the Binomial distribution. A discrete Poisson r.v. X with parameter μ is given by: $Pr(X = j) = \frac{e^{-u}\mu^j}{j!}$. Expected value = μ.

2 Important Theorems

Union of Events The probability of the union of events is no more than the sum of their probabilities. The events need not be independent.

Linearity of Expectation Expectation of sum (of finite number of discrete r.v.s) is the sum of the expectations. (Note that the independence is *not* required.)

Linearity of Variances Variance of sum (of finite number of *independent* discrete r.v.s) is the sum of the variances. (Note that independence is required.)

Jensen's Inequality If f is convex (U-shaped), then $E[f(X)] \geq f(E[X])$

Markov's Inequality $Pr(X \geq a) \leq E[X]/a$, for all $a > 0$ and for all r.v.s that only assume non-negative values. A corollary: $Pr(X \geq kE[X]) \leq 1/k$, for positive integer k.

Chebyshev's Inequality uses the variance to bound the deviation from the expected value. $Pr(|X - E[X]| \geq a) \leq Var[X]/a^2$. Variants:

1. $Pr(|X - E[X]| \geq t \cdot \sigma[X]) \leq 1/t^2$
2. $Pr(|X - E[X]| \geq t \cdot E[X]) \leq \frac{Var[X]}{t^2(E[X])^2}$
3. (Using higher moments) $Pr(|X - E[X]| > t\sqrt[k]{E[(X - E[X])^k]} \leq 1/t^k$.

Chernoff's Bounds for a r.v. is obtained by applying Markov's inequality to e^{tX} for some well chosen t. For a given $\delta > 0$, these bounds give the probability that X deviates from its expectation μ by $\delta\mu$ or more. Often, Chernoff's bounds give stronger bounds than Chebyshev's inequality.

Let $X_1, \ldots, X_n$ be n independent random variables with sum X. Let $\mu \geq E[X]$. Then for any $\delta > 0$, we have

$$Pr[X > (1 + \delta)\mu] < \left[\frac{e^{\delta}}{(1 + \delta)^{(1+\delta)}}\right]^{\mu}$$

Weak Law of Large Numbers states that as a sample becomes larger and larger, the sample mean tends closer and closer to the population mean.

Balls-in-Bins Given m balls thrown into n bins (independently and uniformly at random), the probability that the maximum *load* is more than $3 \ln n / \ln \ln n$ is at most $1/n$ for n sufficiently large.

Sum of Poissons The sume of a finite number of independent Poisson r.v.s is a Poisson r.v.

Expected number of trials before success If we repeatedly perform independent trials of an experiment, each of which succeeds with probability $p > 0$, then the expected number of trials needed before the first success is $1/p$.

3 Some useful equalities and inequalities

1. If n is a positive integer, then $\sum_{k=1}^{n} \frac{1}{k} = H(n) = \ln n + O(1)$
2. $\binom{n}{i} \leq (\frac{ne}{i})^i$
3. $k! > (\frac{k}{e})^k$
4. $e^x = 1 + x + \frac{1}{2}x^2 + \frac{1}{6}x^3 + \ldots = \sum_{i=1}^{\infty} \frac{x^i}{i!}$. Consequently, if $x > 0$, then $e^x \geq 1 + x$. Also, if $x < 3$, then $e^{-x} \geq 1 - x$.
5. $\ln(1 + x) = x - \frac{1}{2}x^2\frac{1}{3}x^3 + \ldots = \sum_{i=1}^{\infty}(-1)^{i+1}\frac{x^i}{i}$. Consequently, if $0 \leq x \leq 1$, then $\ln(1 + x) \leq x$.
6. **Useful Asymptotic Bounds**
   (a) The function $\left(1 - \frac{1}{n}\right)^n$ converges monotonically from $\frac{1}{4}$ up to $\frac{1}{e}$ as n increases from 2.
   (b) The function $\left(1 - \frac{1}{n}\right)^{n-1}$ converges monotonically from $\frac{1}{2}$ up to $\frac{1}{e}$ as n increases from 2.

4 Summation from the QuickSort analysis

$$E[X] = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{2}{j - i + 1}$$

$$
\begin{aligned}
&= \sum_{i=1}^{n-1} \sum_{k=2}^{n-i+1} \frac{2}{k} \\
&= \sum_{k=2}^{n} \sum_{i=1}^{n+1-k} \frac{2}{k} \\
&= \sum_{k=2}^{n} (n+1-k)\frac{2}{k} \\
&= (n+1)2H(k) - 2(n-1) \\
&= 2n \ln n + \Theta(n)
\end{aligned}
$$

5 Summation from the Min-Cut analysis

Let E_i be the event that the edge contracted in iteration i is not in C. Let F_i be the event that no edge of C was contracted in the first i iterations.

$$Pr(E_1) = Pr(F_1) \geq 1 - \frac{k}{kn/2} = 1 - \frac{2}{n}.$$

Similarly,

$$Pr(E_i|F_{i-1}) \geq 1 - \frac{k}{k(n-i+1)/2} = 1 - \frac{2}{n-i+1}.$$

We need to compute $\Pr(F_{n-2})$. Thus,

$$
\begin{aligned}
Pr(F_{n-2}) &= Pr(E_{n-2} \cap F_{n-3}) = Pr(E_{n-2}|F_{n-3})Pr(F_{n-3}) \\
&= Pr(E_{n-2}|F_{n-3}) \cdot Pr(E_{n-3}|F_{n-4}) \ldots Pr(E_2|F_1)Pr(F_1) \\
&\geq \prod_{i=1}^{n-2} \left(1 - \frac{2}{n-i+1}\right) = \prod_{i=1}^{n-2} \frac{n-i-1}{n-I+1} \\
&= \left(\frac{n-2}{n}\right)\left(\frac{n-3}{n-1}\right) \cdots \frac{4}{6}\frac{3}{5}\frac{2}{4}\frac{1}{3} \\
&= \frac{2}{n(n-1)}.
\end{aligned}
$$

Repeating the algorithm $n(n-1)\ln n$ times, the probability that the output is not a min-cut set is bounded by

$$\left(1 - \frac{2}{n(n-1)}\right)^{n(n-1)\ln n} \leq e^{-2\ln n} = \frac{1}{n^2}$$

The inequality above arises from the fact that $1 - x \leq e^{-x}$ and replacing x by $\frac{2}{n(n-1)}$. You are probably more familiar with the following fact:

$$\lim_{x \to 0}(1-x)^{1/x} = \frac{1}{e} = e^{-1}$$

6 Analysis of the Greedy Set Cover Algorithm

In each iteration, the algorithm adds the set containing the greatest number of uncovered elements. We want to show that the above algorithm (ALG) is a $\ln \frac{n}{OPT}$–approximate algorithm. Here's the proof.

Let $K = OPT$ be the size of optimal set cover. Let E_t be the set of elements uncovered after step t, with $E_0 = E$. The optimal set cover covers every E_t with no more than K sets. ALG always picks the largest set over E_t in step $t+1$. The size of this set is at least $|E_t|/K$, which is the average size of a set in the set cover. Thus $|E_{t+1}| \leq |E_t| - |E_t|/K$, and, $|E_t| \leq |E_0|(1-1/K)^t \leq n(1-1/K)^t$. We stop when $E_t| < 1$, which happens when $(1 - \frac{1}{K})^t < \frac{1}{n}$. This is when

$$\begin{aligned} n &> \left(\frac{K}{K-1}\right)^t \\ \ln n &> t \ln\left(1 + \frac{1}{K-1}\right) \approx \frac{t}{K} \\ t &\leq K \ln n. \end{aligned}$$

7 Birthday Paradox Analysis

Probability that m balls are put in distinct bins is

$$\left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right) \ldots \left(1 - \frac{m-1}{n}\right) = \prod_{j=1}^{m-1}\left(1 - \frac{j}{n}\right).$$

When $k \ll n$, $1 - k/n \approx e^{-k/n}$, we can simplify the above as follows:

$$\prod_{j=1}^{m-1}\left(1 - \frac{j}{n}\right) = \prod_{j=1}^{m-1} e^{-j/n}$$

$$
\begin{aligned}
&= exp\Big\{ -\sum_{j=1}^{m-1} \frac{j}{n} \Big\} \\
&= e^{-m(m-1)/2n} \\
&\approx e^{-m^2/2n}
\end{aligned}
$$

For what value of m is the above probability at least $1/2$? This happens for $m \geq \sqrt{2n \ln 2}$. Thus, if there are at least 23 people in a room, then with probability at least $1/2$, there will be two people in the room with the same birthday.

8 Hashing with Chaining

Let N be the number of possible hash values. Let M be the maximum number of items that may be stored in the hash table. Let k be the number of items stored in the table currently. The probability that any item is hashed to a certain value is $1/N$.

Unsuccessful Search; Average Cost We first compute the probability that exactly i out of the k items are hashed to the same hash value. The probability that i items are hashed to a specific value is N^{-i}. The probability that $k-i$ items are not hashed to that value is $(\frac{N-1}{N})^{k-i}$. Since there are $\binom{k}{i}$ ways of choosing i items from a possible set of k items, the probability that exactly i out of the k items are hashed to the same hash value is thus:

$$p_i = \binom{k}{i}(N-1)^{k-i}N^{-k}.$$

It is also the probability that we will search a list of length i (making $i+1$ comparisons. Note that $\sum_i p_i = 1$. Therefore, the average number of comparisons made during an unsuccessful search is given by

$$
\begin{aligned}
A &= \sum_i (i+1)p_i = \sum_i \binom{k}{i}(i+1)(N-1)^{k-i}N^{-k} \\
&= \sum_i \binom{k}{i} i(N-1)^{k-i}N^{-k} + \sum_i \binom{k}{i}(N-1)^{k-i}N^{-k}
\end{aligned}
$$

$$
\begin{aligned}
&= \sum_i k\binom{k-1}{i-1}(N-1)^{k-i}N^{-k} + 1 \\
&= kN^{-k}\sum_i \binom{k-1}{i}(N-1)^{k-i-1} + 1 \\
&= kN^{-k}N^{k-1} + 1 = 1 + k/N
\end{aligned}
$$

The variance can be shown to be $\frac{k}{N}\left(1 - \frac{1}{N}\right)$. The variance can be calculated using the following relationship

$$V = \sum_i (i+1)^2 p_i - A^2.$$

When k is small compared to N, the average number of comparisons for an unsuccessful search is at most 2 and the variance is small.

Successful Search; Average Cost Since there are k items in the hash table, the probability that we have i items hashed to the same value (i.e., a specific list has i items) is $p_i = \binom{k}{i}(N-1)^{k-i}N^{-k}$ and the probability that an arbitrary item we are looking for is in a list of length i is i/k. The probability that an arbitrary item we are looking for is in a specific list is $(i/k)p_i$. The search time for the item depends on where it is on the list. The item has probability $1/i$ of being the j-th item on the list. Hence the probability q_{ij} that it is the j-th item on a list of length i is equal to $\frac{Np_i}{k}$, if $1 \leq j \leq i$ and 0 otherwise. We are now ready to compute the average as follows

$$A' = \sum_{i,j} j q_{ij} = 1 + \frac{k-1}{2N}$$

It should be no surprise that the average time is roughly half of that for an unsuccessful search.

Maximum Load The *maximum load* is the number of items in the largest list. We show the following: if n balls are thrown into n bins independently and uniformly at random, then the probability that the maximum load is more than $3 \ln n / \ln \ln n$ is at most $1/n$ for n sufficiently large.

The following is the proof of the above statememt. The probability that any bin has at least j balls is at most

$$\binom{n}{j}\left(\frac{1}{n}\right)^j \leq \frac{1}{j!} \leq \left(\frac{e}{j}\right)^j$$

The second inequality above is due to the approximation for $j!$ (see Sec 3, formula 3.

The probability that one of the n bins has at least $j = 3\ln n/\ln\ln n$ balls is bounded by

$$\begin{aligned} n\left(\frac{e}{j}\right)^j &\leq n\left(\frac{e\ln\ln n}{3\ln n}\right)^{3\ln n/\ln\ln n} \\ &\leq n\left(\frac{\ln\ln n}{3\ln n}\right)^{3\ln n/\ln\ln n} \\ &= e^{\ln n}\left(e^{\ln\ln\ln n - \ln\ln n}\right)^{3\ln n/\ln\ln n} \\ &= e^{-2\ln n + 3(\ln n)(\ln\ln\ln n)/\ln\ln n} \\ &\leq \frac{1}{n} \end{aligned}$$

for n sufficiently large. Note that when n is sufficiently large, we can make $\frac{3\ln\ln\ln n}{\ln\ln n} < 1$. Hence the result.