Randomized Algorithm

Acknowledgement
This notes is compiled based on the following textbooks and a variety of Internet resources, such as Wikipedia for example.

- Introduction to Algorithms by Cormen, Lerserson, Rivest and Stein.
- Algorithms by Dasgupta, Papadimitriou and Vazirani
- Randomized Algorithms by Motwani and Raghavan.
- Computational Geometry, Algorithms and Applications by de Berg, Cheong, Kreveld, and Overmars.
- Probability and Computig, Randomized Algorithms and Probabilitistic Analysis by Mitzenmacher and Upfal.

1 Background

1.1 Discrete Probability

Definition 1 (Probabilistic Experiment) *There are experiments that do not yield the same results when performed repeatedly. Such experiments are called probabilistic experiments. In contast a deterministic experiment always produces the same outcome.*

Definition 2 *The sample space S of a probabilitic experiment is the set of all of its outcomes.*

Example 1 *The sample space for tossing a 6-sided die is* $\{1, 2, 3, 4, 5, 6\}$.

Consider the experiment of tossing a 6-sided die. Suppose the outcome is 2. Does the event that the outcome is even happen? The answer is yes. Now assume that the toss produces an outcome 4, does the event that the outcome is even happen? The answer is yes. Actually observe that when the outcome is from the set $\{2, 4, 6\}$, the event happens. Thus an **event** in a probabilistic experiment is actually a subet of the outcomes.

Definition 3 *A probability space for a probabilistic experiment consists of a triple $(S, \mathcal{E}, P)$:*

- *S is the sample space that includes all the possible outcomes of the experiment.*
- *$\mathcal{E}$ is a collection of events, i.e., $\mathcal{E} \subset 2^S$.*
- *$P : \mathcal{E} \to [0, 1]$ is the probability function that assigns a probability between 0 and 1 to all the events, where a probability of 1 means the event will always happen and a probability of 0 means the event will never happen.*

There are some special events of interests, and we shall define them here.

The sample space S itself is an event. Since all the outcomes belong to the sample space, no matter what outcome happens, the event S always happens. Thus the probability of the event S happening is 1. We call the event S a **certain event**. The empty set ϕ is also an event. Since ϕ doesn't contain any outcomes, it will never happen. Thus the probability of ϕ happening is 0. We call the event ϕ an **impossible event**.

Two events E and F are said to be **mutually exclusive** or **disjoint** if $E \cap F = \phi$. A set of events $E_1, E_2, ..., E_n$ are said to be mutually exclusive if every pair of them are mutually exclusive.

Axiom 1 (Axiom of Probability) *Let $(S, \mathcal{E}, P)$ be a probability space. Then:*

- $0 \leq P(E) \leq 1$ *for any event* $E \in \mathcal{E}$.
- $P(S) = 1$ *and* $P(\phi) = 0$.
- $P(E \cup F) = P(E) + P(F)$ *for* $E \cap F = \phi$.
- $\mathcal{E}$ *is closed under intersection, union, and complement. In other words, for* $E, F \in \mathcal{E}$, $E \cap F, E \cup F, \overline{E} \in \mathcal{E}$.

1.2 Conditional Probability

Definition 4 *The joint probability between two events E and F is the probability that both events occur and is equal to $P(E \cap F)$.*

Definition 5 *Two events E and F are said to be independent if $P(E \cap F) = P(E) \cdot P(F)$.*

Definition 6 *Let E and F be two events. The probability that E happens given that F happens is called the conditional probability, which is denoted by $p(E|F)$. $P(E|F) = \frac{P(E \cap F)}{P(F)}$.*

Theorem 1 (Bayes Theorem) $P(E|F) = \frac{P(F|E) \cdot P(E)}{P(F|E) \cdot P(E) + P(F|\overline{E}) \cdot P(\overline{E})}$.

Proof: Observe that from $P(E|F) = \frac{P(E \cap F)}{P(F)}$, we have $P(E \cap F) = P(E|F) \cdot P(F)$.

Thus $\frac{P(F|E) \cdot P(E)}{P(F|E) \cdot P(E) + P(F|\overline{E}) \cdot P(\overline{E})} = \frac{P(F \cap E)}{P(F \cap E) + P(F \cap \overline{E})} = \frac{P(E \cap F)}{P(F)}$. □

Example 2 *Suppose a person in 100,000 has this rare disease. There is diagnostic test. The test is correct for 99% when administered to a subject with the disease. The test is correct for 99.5% when administered to a subject without the disease. What is the probability that a given person testing positive actually has the disease.*

Solution: Let E be the event that a given person has the disease, and F be the event that a given person testing positive. We'd like to calculate $P(E|F)$.

Using Bayes Theorem, we know that $P(E|F) = \frac{P(F|E)P(E)}{P(F|E)P(E) + P(F|\overline{E})P(\overline{E})}$.

Observe that:

- $P(F|E)$ is the conditional probabiilty that a person with the desease tests positive, and is 0.99.

- $P(E)$ is the probability that a person has the rare disease, and is 0.000001.
- $P(\overline{E})$ is the probability that a person doesn't have the disease, and is $1 - P(E) = 0.999999$.
- $P(F|\overline{E})$ is the probability that a healthy person testing positive, and is $1 - 0.995 = 0.005$.

Substitute in all these numbers to Bayes Theorem, we have $P(E|F) \approx 0.002$.

□

1.3 Random Variables

Cnnsider the probability space derived from the experiment of tossing a fair coin. The sample space $S = \{Head, Tail\}$, the event set $\mathcal{E} = \{\phi, \{Head\}, \{Tail\}, S\}$, the elementary probabilties are $P(\{Head\}) = 0.5$ and $P(\{Tail\}) = 0.5$.

It is very cumbersome to use Head and Tail to represent the outcomes. What if we use a function that maps Head and Tail to some numbers. For example, let X be a function from the sample space S to real numbers. We define X as $X(Head) = 1$ and $X(Tail) = 0$. In this case, we no longer need to use Head and Tail, we can use their "surrogates" 1 and 0 to represent them. This is the concept of a **random variable (RV)**, which allows us to deal with numbers instead of a concrete probabilistic experiment. Thus a random variable is not really a variable, but rather a function whose domain is the sample space and whose range are real numbers.

Let S be a sample space and $X(\cdot)$ a random variable that maps S to the real numebrs $\mathcal{R}$. Let the image of X denoted by $Image(X)$ be defined as: $Image(X) = \{x | \exists s \in S, X(s) = x\}$. We say $Image(X)$ is the set of the possible values of the random variable X. A random variable X is discrete if $Image(X)$ is countable. A random variable X is nonnegative if all elements of $Image(X)$ are nonnegative.

For a discrete probability space, a random variable X maps the sample space to a set discrete numbers $Image(X)$. What about elementary probabilities of the invidual numbers in $Image(X)$? The elementary probabilities of the numbers in $Image(X)$ will be the same as the elementary probability of their corresponding elementary event in S. The elementary probabilities actually defines a function from $Image(X)$ to $[0, 1]$, which is called the **probability mass function** and is denoted by P_X.

Definition 7 *Let X be a random varaibles with values $\{X_1, X_2, ..., X_n\}$ and proabability mass function P_X. Let Y be a random varaibles with values $\{Y_1, Y_2, ..., Y_m\}$ and proabability mass function P_Y. The joint probability distribution for X and Y, denoted by (X, Y) is a probability distribution that gives the probability for $P(X = X_j, Y = Y_k)$.*

Theorem 2 *Let X be a random varaibles with values $\{X_1, X_2, ..., X_n\}$ and proabability mass function P_X. Let Y be a random varaibles with values $\{Y_1, Y_2, ..., Y_m\}$ and proabability mass function P_Y. Let (X, Y) be the joint random variable with probability mass funciton $P_{X,Y}$. Then $\sum_{j=1}^{n} P(X = X_j, Y = Y_k) = P_Y(Y = Y_k)$ and $\sum_{k=1}^{m} P(X = X_j, Y = Y_k) = P_X(X = X_j)$.*

Definition 8 *Two discrete random variables X and Y are said to be* **independent** *if $P_{X,Y} = P_X P_Y$, i.e., for any X_j, Y_k, $P_{X,Y}(X = X_j, Y = Y_k) = P_X(X = X_j) P_Y(Y = Y_k)$.*

1.4 Sum of Random Variables

Consider tossing two 6-sided fair dice. Let X be the random variable representing the outcome of the first die, and Y be the random varaible representing the outcome of the second die. The sum of the values of the two dice is also a random variable. If we use Z to denote the sum, then $Z = X + Y$. What is the probability distribution of Z?

Observe that the possible values of Z are $\{X + Y | X \in \{1, 2, 3, 4, 5, 6\}, Y \in \{1, 2, 3, 4, 5, 6\}\} = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$. What is the probability mass function P_Z of Z? To obtain P_Z, we must know the probability of each individual value of Z. We shall use the calculation of the probability of $Z = 9$ to illustrate the construction of P_Z. Observe that in order for Z to be 7, one of the following situations must occur: (1) $X = 3$ and $Y = 6$, (2) $X = 4$ and $Y = 5$, (3) $X = 5$, $Y = 4$, and (4) $X = 6$ and $Y = 3$. Since both dice are fair, the odds of getting any of these 4 situations is $\frac{1}{36}$. Thus the probably $P(Z = 9) = \frac{4}{36}$.

More generally, let X and Y be two random variables, where X assumes the values $\{X_1, X_2, ..., X_n\}$ with a probability mass function P_X and Y assumes the values $\{Y_1, Y_2, ..., Y_m\}$ with a probability mass function P_Y. Let Z be the random variable such that $Z = X+Y$, then $P(Z = Z_i) = \sum_{X_j+Y_k=Z_i} P(X = X_j, Y = Y_k)$.

1.5 Expectation and Variance

Definition 9 (Expectation) *Let X be a discrete random variable with values $\{X_1, X_2, ..., X_n\}$ and a probability mass funciton P. The expectation of X is defined as $E[X] = \sum_{k=1}^{n} P(X_k) \cdot X_k$.*

Theorem 3 *Let X and Y be discrete random variables. Then $E[X + Y] = E[X] + E[Y]$.*

Proof: Let X be a random varaibles with values $\{X_1, X_2, ..., X_n\}$ and proabability mass function P_X. Let Y be a random varaibles with values $\{Y_1, Y_2, ..., Y_m\}$ and proabability mass function P_Y. Let (X, Y) be the joint random variables with probability mass function $P_{X,Y}$. Let $Z = X+Y$ with values $\{Z_1, Z_2, ..., Z_l\}$ and probability mass function P_Z.

$$E[X + Y] = E[Z] = \sum_{i=1}^{l} Z_i \cdot P(Z = Z_i)$$

$$= \sum_{i=1}^{l} Z_i \sum_{X_j+Y_k=Z_i} P_{X,Y}(X = X_j, Y = Y_k)$$

$$= \sum_{j=1}^{n} \sum_{k=1}^{m} P_{X,Y}(X = X_j, Y = Y_k)(X_j + Y_k)$$

$$= \sum_{j=1}^{n} \sum_{k=1}^{m} P_{X,Y}(X = X_j, Y = Y_k)X_j + \sum_{j=1}^{n} \sum_{k=1}^{m} P_{X,Y}(X = X_j, Y = Y_k)Y_k$$

$$= \sum_{j=1}^{n} X_j \sum_{k=1}^{m} P_{X,Y}(X = X_j, Y = Y_k) + \sum_{k=1}^{m} Y_k \sum_{j=1}^{n} P_{X,Y}(X = X_j, Y = Y_k)$$

$$= \sum_{j=1}^{n} X_j P_X(X = X_j) + \sum_{k=1}^{m} Y_k P_Y(Y = Y_k)$$

$$= E[X] + E[Y]$$

□

Theorem 4 (Markov Inequality) *Let X be a discrete nonnegative random variable, and $\delta > 0$. Then $P(X \geq \delta) \leq \frac{E[x]}{\delta}$.*

Proof: Let the values of X be $0 \leq X_1 < X_2 < ... < X_{j-1} \leq \delta \leq X_j < ... < X_n$. Then

$$E[X] = P(X_1) \cdot X_1 + P(X_2) \cdot X_2 + ... + P(X_n) \cdot X_n$$

$$\geq P(X_j) \cdot X_j + P(X_{j+1}) \cdot X_{j+1} + ... + P(X_n) \cdot X_n$$

$$\geq P(X_j) \cdot \delta + P(X_{j+1}) \cdot \delta + ... + P(X_n) \cdot \delta$$

$$\geq (P(X_j) + P(X_{j+1}) + ... + P(X_n)) \cdot \delta$$

$$= P(X \geq \delta) \cdot \delta$$

Thus $P(X \geq \delta) \leq \frac{E[x]}{\delta}$. □

Definition 10 (Variance) *Let X be a discrete random variable with values $X_1, X_2, ..., X_n$. The variance of X is defined as: $Var(X) = \sum_{k=1}^{n} \left(P(X_j) \cdot (X - E[X])^2\right)$. The standard deviation of X is defined as $\sigma(X) = \sqrt{Var(X)}$.*

The expectation measures the average of a random variable, while the standard deviation measures its spread.

Theorem 5 *$Var(X) = E[X^2] - (E[X])^2$.*

Proof: $Var(X) = E[(X - E[X])^2] = E[X^2 - 2XE[X] + (E[X])^2] = E[X^2] - E[2XE[X]] + (E[X])^2 = E[X^2] - 2(E[X])^2 + (E[X])^2 = E[X^2] - (E[X])^2$. □

Theorem 6 (Chebyshev's Inequality) *$P(|X - E[X]| \geq \delta) \leq \frac{Var(X)}{\delta^2}$.*

Proof: Consider the random variable $Z = (X - E[X])^2$. Observe that $E[Z] = Var(X)$. Thus $P(|X - E[X]| \geq \delta) = P(Z \geq \delta^2) \leq \frac{E[Z]}{\delta^2} = \frac{Var(X)}{\delta^2}$. The inequality is from Markov's Inequality. □

Theorem 7 *Let X_1, X_2, ..., X_n be a set of independent random variables. Then $Var(X_1 + X_2 + ... + X_n) = Var(X_1) + Var(X_2) + ... + Var(X_n)$.*

1.6 Some Important Proabiity Distributions

Definition 11 *The Bernolli trial is a probabilitistic experiment with only two outcomes. Since there are only two outcomes, we usually refer to one of the outcome as "success" and the other as "failure". Usually we use a binary random variable X (random variable with value either 1 for success and 0 for failure) to represent a Bernolli trial, where $P(X = 1) = p$ and $P(X = 0) = 1 - p$.*

Definition 12 (Binomial Distribution) *Consider a Bernoulli trial with a probability of success equal to p and probability of failure equal to $1-p$. Let X be the random variable representing the number of success in a sequence of n independent Bernoulli trials. Then $P(X=k) = \binom{n}{k}p^k(1-p)^{n-k}$, and is denoted by $B(n,p)$. The probability mass function of X is called a binomial distribtion.*

Theorem 8 *$E[B(n,p)] = np$ and $Var(B(n,p)) = np(1-p)$.*

Proof: Let X_k be the random variable associated with the $k-$th Bernoulli trial, and $P(X_k = 1) = p$ and $P(X_k = 1) = 1-p$. Then $E[X_k] = p\cdot 1 + (1-p)\cdot 0 = p$. $Var(X_k) = p\cdot(1-p)^2 + (1-p)\cdot(0-p)^2 = p(1-p)$.

Now observe that $X = X_1 + X_2 + ... + X_n$. Using linearity of expectation, we have $E[X] = E[X_1] + E[X_2] + ... + E[X_n] = np$. From Theorem 7, we have $Var[X] = Var[X_1] + Var[X_2] + ... + var[X_n] = np(1-p)$.

□

Definition 13 (Geometric Distribution) *Consider a Bernoulli trial with a probability of success equal to p and probability of failure equal to $1-p$. Let X be the random variable representing the number of trials until the first success occurs. Then $P(X=k) = (1-p)^{k-1}p$. The probability distribution of X is called a geometric distribution.*

Theorem 9 *Let X be a random variable subject to a geometric distribution. The $E[X] = \frac{1}{p}$*

Proof: $[E[X] = (1-p)^0p\cdot 1 + (1-p)^1p\cdot 2 + (1-p)^3p\cdot 3 + ...$ Let $S_n = \sum_{j=1}^{n}(1-p)^{j-1}p$, then we have $E[X] = lim_{n\rightarrow}S_n$.

Observe that

$$S_n = (1-p)^0p\cdot 1 + (1-p)^1p\cdot 2 + (1-p)^3p\cdot 3 + ... + (1-p)^{n-1}p\cdot n$$

$$(1-p)S_n = (1-p)^1p\cdot 1 + (1-p)^2p\cdot 2 + (1-p)^3p\cdot 3 + ... + (1-p)^np\cdot n$$

$$S_n - (1-p)S_n = (1-p)^0p + (1-p)^1p + (1-p)^3p + ... + (1-p)^{n-1}p - (1-p)^np\cdot n$$

$$pS_n = ((1-p)^0 + (1-p)^1 + (1-p)^3 + ... + (1-p)^{n-1})\cdot p - (1-p)^n\cdot n$$

$$pS_n = \frac{(1-p)^n - 1}{(1-p)-1}\cdot p - (1-p)^n\cdot n$$

$$pS_n = 1 - (1-p)^n - (1-p)^n\cdot n$$

$$pS_n = 1 - (1-p)^n\cdot(n+1)$$

$$S_n = \frac{1}{p} - \frac{(1-p)^n\cdot(n+1)}{p}$$

Thus

$$E[X] = lim_{n\rightarrow\infty}\left(\frac{1}{p} - \frac{(1-p)^n\cdot(n+1)}{p}\right) = \frac{1}{p}$$

$E[X^2] = (1-p)^0p\cdot 1^2 + (1-p)^1p\cdot 2^2 + (1-p)^3p\cdot 3^2 + ...$ Let $T_n = (1-p)^0p\cdot 1^2 + (1-p)^1p\cdot 2^2 + (1-p)^3p\cdot 3^2 + ... + (1-p)^{n-1}p\cdot n^2$, then $lim_{n\rightarrow\infty}T_n = E[X^2]$. □

Definition 14 (Poisson Distribution) *A discrete random variable X is said to subjec to Poisson distribution if $P(X=k)=\frac{\lambda^k e^{-\lambda}}{k!}$ for fixed $\lambda>0$ and $k=0,1,...$*

Observe that $\sum_{k=0}^{\infty}\frac{\lambda^k e^{-\lambda}}{k!}=e^{-\lambda}\sum_{k=0}^{\infty}\frac{\lambda^k}{k!}=e^{-\lambda}e^{\lambda}=1$. Thus Poisson distribution is well defined.

The Poisson distribution describes the probability of a given number of events occurring in a fixed interval of time if these events occur with a known average rate, and independent of the time since last event.

Theorem 10 *Let X be a Poission random variable. Then $E[X]=\lambda$.*

Proof:

$E[X]=\sum_{k=0}^{\infty}\left(\frac{\lambda^k e^{-\lambda}}{k!}\cdot k\right)=\sum_{k=0}^{\infty}\left(\frac{\lambda^k e^{-\lambda}}{(k-1)!}\right)=\lambda\sum_{k=0}^{\infty}\left(\frac{\lambda^{k-1} e^{-\lambda}}{(k-1)!}\right)=\lambda\sum_{k=0}^{\infty}\left(\frac{\lambda^k e^{-\lambda}}{k!}\right)=\lambda.$ □

2 Introuduction

Definition 15 *A* **randomized algorithms** *is an algorithms that make random choices during its execution. It assumes the on-demand availability of uniform random bits.*

Generally speaking, there are two broad categories of randomized algorithms. One type of randomized algorithms is the **Las Vegas** randomized algorithms. These algorithms use randomness to achieve good expected running times, and guarantees that the output is always correct. The other type is the **Monte Carlo** randomized algorithms. These algorithms have a deterministic running time, but guarantees that the output is correct with a certain probability.

Randomized algorithms have many applications such as:

- **Beating the adversary:** Traditionally algorithmic running time analysis are pessimistic based on the assumption of an "evil" adversary. Using randomization, such as randomly perturb the input for an example, radnomized algorithm can beat adversary and achieve good expected running times.
- **Random sampling:** The idea of using a relatively small random sampling from a population as a representative of the population.
- **Hasing:** The idea of a small set of elements drawn from a larger universe can be mapped into a smaller space.
- **Existence Proofs:** To prove the existence of a mathematical object or certain properties, it suffices to show that if one randomly chooses objects from a specified class, the object chosen will have the prescribed properties with nonzero probability.
- etc.

3 Random Number Generator

Definition 16 *A sequence of numbers are said to be* **random** *if it can't be predicted.*

Given a sequence of numbers, we can test its randomess using the following methods:

- Frequency and Histogram Test: We will check and make sure the frequence of each number are the same.
- Serial Test: We will check and make sure any sequence of numbers (e.g., a sequence of 2 numbers, 5 numbers, etc) have the same frequency of occurrence.
- Gap test: We will check the inverval between recurrence of the same number.
- *etc.*

Definition 17 *A* **random number generator** *is a physical device or a computational algorithm for producing a sequence of numbers that appears to be* **random***, i.e., contains no recognizable patterns and can not be predicted.*

Examples of physical random number generators are tossing a coin, a dice, the roulette wheel, *etc.* In this section, we are more interested in algebraic random number generators, which employs a computer algorithm to produce on-demand random numbers.

3.1 Linear Congruential Generator

Definition 18 *The linear congruential generator uses the following recurrence relation to produce a sequence of pseudo random numbers:*

$$X_{n+1} = (aX_n + c) \mod m$$

where $X_0, X_1.X_2, ...$ is the pseudo-random sequence, X_0 is call the seed, a is called the multiplier, c is the called increment, and m is called the modulus.

Observe that if $X_0 = X_n$ for some n, the sequence will repeat. The smallest integer n that satisfies $X_0 = X_n$ is called the **period length** of the random number generator. Clearly $n \leq m$.

Theorem 11 *The linear congruence defined by m, a, c, and X_0 has the (maximum) period legnth m if and only if*

- *c is relatively prime to m.*
- *$a - 1$ is a multiple of p for every prime factor p of m.*
- *$a - 1$ is a muliple of 4 if m is a multiple of 4.*

4 Goldern Rule of Sampling

Suppose that we have on-demand avaiability of uniform random numbers on the interval $[0, 1]$. In other words, every number on the interval is equally likely to be chosen by our random number generator. The question we are interested in this section is how do we generate random numbers subject to other distributions, for example the Gaussian distribution $N(0, 1)$?

4.1 Continuous Random Variables, Probability Density Function, and Cumulative Distribution Function

Let X be a continuous random variable. The **cumulative distribution function (CDF)** (or simply **distribution function** of X if the function $F(x)$ that specifies the probability $P(X \leq x)$, i.e., $F(x) = P(X \leq x)$.

From the definitions of F, it is obvious that: $\lim_{x\to-\infty} F(x) = 0$ and $\lim_{x\to\infty} F(x) = 1$.

Now given the CDF $F(x)$ of X, what is the probability $P(a < x \leq b)$?

Since $P(X \leq a) + P(a < X \leq b) = P(X \leq b)$, $P(a < X \leq b) = P(X \leq b) - P(X \leq a) = F(b) - F(a)$.

The function $f(x) = \frac{dF(x)}{dx}$ is called the **probability density function** of X. (Note that we call this functino the "density" because $\frac{dF(x)}{dx} = \lim_{\Delta x\to 0} \frac{F(x+\Delta x)-F(x)}{\Delta x}$, which can be viewed as the density of the probability $P(x < X \leq x + \Delta x)$ on the interval $(x, x + \Delta x]$.)

Now assume that the probability density function exists, then:

- $\int_{-\infty}^{a} f(x)dx = \int_{-\infty}^{a} \frac{dF(x)}{dx}dx = \int_{-\infty}^{a} dF(x) = F(x)|_{-\infty}^{a} = F(a) - F(-\infty) = F(a)$.
- $\int_{a}^{b} f(x)dx = \int_{a}^{b} \frac{dF(x)}{dx}dx = \int_{a}^{b} dF(x) = F(x)|_{a}^{b} = F(b) - F(a)$.
- $\int_{-\infty}^{\infty} f(x)dx = F(\infty) - F(-\infty) = 1$

For continuous random variable X, we can define its **expectation** $E[X] = \int_{-\infty}^{\infty} xf(x)dx$, and its **variance** $Var(X) = \int_{-\infty}^{\infty} (x - E[x])^2 f(x)dx$.

We we give two important examples of continous random variables:

The first is the random variable U that is unformly distirbution on the interval $[0, 1]$, i.e., it is equally likely for U to assume any real value between 0 and 1. This implies that the probability density $f_U(x)$ of U has a constant value 1 from any value from 0 to 1, i.e., $f_U(u) = \begin{cases} 0 & u < 0 \\ 1 & 0 \leq x \leq 1 \\ 0 & u > 1 \end{cases}$

The second is the Gaussian (Normal) random variable $N(\mu, \sigma)$, whose probability density is defined as: $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$.

4.2 Golden Rule of Sampling

Let X be a continous R.V., with CDF F. We want to generate random samples of X. The Golden Rule of Sampling works as follows:

1. Generate a random number u of a R.V. U that is uniformly distributed on the interval [0,1].
2. Compute the value x such that $F(x) = u$, i.e., $\int_{-\infty}^{x} f(t)dt = u$.

3. Take x to be the random number drawn from the distribution function F.

Proof: Let F^{-1} be the **inverse transform function** of F defined as $F^{-1}(u) = \inf\{x | F(x) \leq u\}$. Note that for $u \in [0,1]$, $F(F^{-1}(u)) = u$.

Let X be the R.V. defined as $X = F^{-1}(U)$. Observe that $P(X \leq x) = P(F^{-1}(U) \leq x) = P\left(F\left(F^{-1}(U)\right) \leq F(x)\right) = P\left(U \leq F(x)\right) = \int_0^{F(x)} 1dt = F(x)$. Thus, the CDF of X is F, hence the R.V., $X = F^{-1}(U)$ is distributed according to F. □

5 Randomized Quick Sort

Algorithm: Randomized Quicksort

Input: An array A of n numbers

Output: The given numbers sorted

1. Generate a random perturbaton of the input
2. Proceed as normal quick sort, i.e., use the very first element as the pivot to partition the array and recursively sort the numbers $<$ and $>$ the pivot.

For ease of explanation, we shall assume that the input to the quicksort phase is a random perturbation σ of $x_1 < x_2 < ... < x_n$.

Observation 1 *The x_i and x_j will be compared at most once during the execution of the algorithm.*

Let $X_{i,j}$ be the indicator R.V., such that $X_{i,j} = 1$ if and only if x_i and x_j are compared during the execution of randomized quick sort. Thus, the running times of the quick sort algorithm is:

$$\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} X_{i,j}$$

The expected running time of randomized quick sort is obviously

$$E\left[\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} X_{i,j}\right]$$

$$= \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} E\left[X_{i,j}\right]$$

Observation 2 *For x_i and x_j to be compared in the randomized quick sort, the numbers $x_{i+1}, x_{i+2}, ..., x_{j-1}$ can be not be chosen as the pivot before x_i and x_j.*

Observation 3 *Since we always use the first given number as the pivot, in order for x_i and x_j to be compared, they must be before $x_{i+1}, x_{i+2}, ..., x_{j-1}$ in the random permutation.*

Since every permutation is equally likely, the odds that x_i and x_j are before $x_{i+1}, x_{i+2}, ..., x_{j-1}$ in the random permutation is $\frac{2}{j-i+1}$.

Thus, $Pr\left(X_{i,j} = 1\right) = \frac{2}{j-i+1}$ and $E\left[X_{i,j}\right] = \frac{2}{j-i+1}$. Hence,

$$E\left[\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} X_{i,j}\right]$$

$$= \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} E\left[X_{i,j}\right]$$

$$= \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{2}{j-i+1}$$

$$= \sum_{i=1}^{n-1} \sum_{k=j-i=1}^{n-i} \frac{2}{k+1}$$

$$\leq \sum_{i=1}^{n-1} \sum_{k=1}^{n} \frac{2}{k}$$

$$= \sum_{i=1}^{n-1} 2 \ln n = 2n \ln n$$

5.1 Harmonic Sequence

Definition 19 *The sequence* $1, \frac{1}{2}, \frac{1}{3}, ..., \frac{1}{n}, ...$ *is called the harmonic sequence.*

We are interested in the sum of a harmonic sequence. Let $S = \sum_{j=1} n \frac{1}{j}$, then we have:

$$S = \sum_{j=1} n \frac{1}{j} = 1 + \sum_{j=2} n \frac{1}{j} \leq 1 + \int_{x=1}^{n} \frac{1}{x} = 1 + \ln n$$

5.2 Variance of the Running Time of Randomized Quicksort

Let X be the random variable describing the running time of randomized quicksort, then:

$$Var(X) = 7n^2 - 4(n+1)^2 \sum_{j=1}^{n} \frac{1}{j^2} - 2(n+1) \sum_{j=1}^{n} \frac{1}{j} + 13n$$

and

$$\sigma(X) \approx 0.65n$$

From Chevyshev Inequality, we have: $P(|X - 2n \ln n| \geq n \ln n) \leq \frac{(0.65n)^2}{(n \ln n)^2} = \frac{0.4225}{\ln^2 n}$. For $n = 10,000$, the probability of the $X \leq n \ln n$ or $\geq 3n \ln n$ is less than 0.5%.

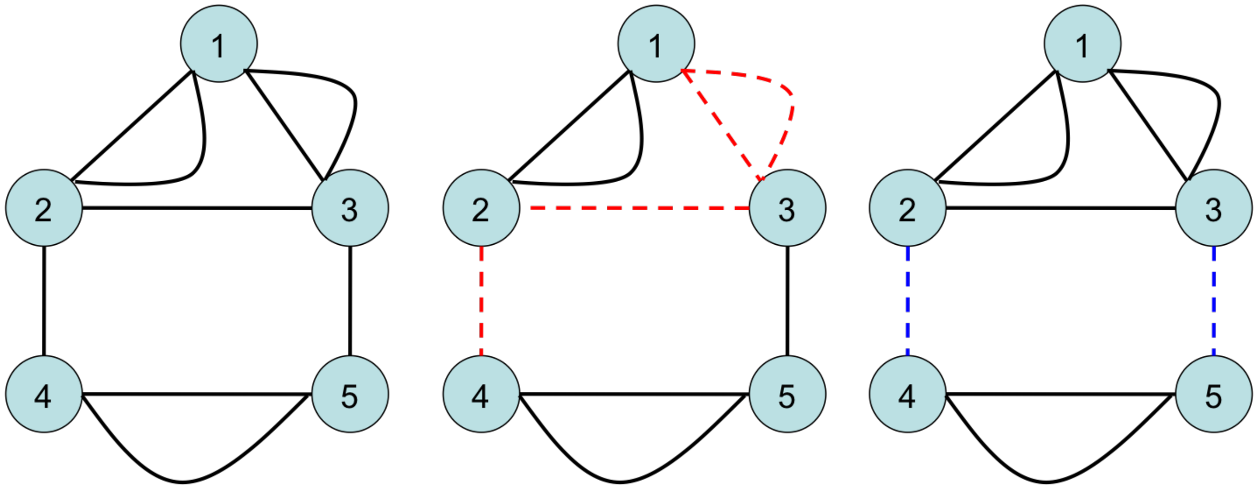

Figure 1: Illustrating cut and min-cut of a multi-graph.

6 Karger's algorithm for Minimum Cut

Definition 20 *Let G be a connected, undirected multi-graph with no self-loops with n vertices. A multi-graph may contain multiple edges between any pair of vertices. A* **cut** *in G is a set of edges whose removal result in G being disconnected. A* **min-cut** *is a cut with minimum number of edges.*

Figure 1 illustrates the concept of cut and min-cut. In Figure 1, the picture to the left is a multi-graph with no self loops; the picture in the middle shows a cut; and the picture to the right shows a min-cut.

The randomized min-cut algorithm that we will present in this section uses an operation called **edge contraction**. Let $e(v, w)$ be an edge of a graph G. By **contracting** e, we mean:

1. Combining the vertices v and w into a new vertex u.

2. Removing all the edges that are previously between v and w.

3. Connecting all other edges that are previouslyincident to v or w to the new vertex u.

Figure 2 illustrates the concept of edge contration, where an edge between vertices 1 and 3 are contracted.

Observation 4 *Let G be a multi-graph with no self-loops and G' be the graph after contracting an arbitrary edge in G. Then, a cut of G' is also a cut in G, and therefore the size of the min-cut in G' is $\geq$ the size of a min-cut in G. In other words, contraction will not reduce the size of the min-cut.*

The following is our randomized min-cut algorithm. (See Figure 3 for illustrations.

Algorithm: Randomized Min-cut

Input: An undirected multi-graph G with no self-loops.

Output: A cut, and hopefully a min-cut of G.

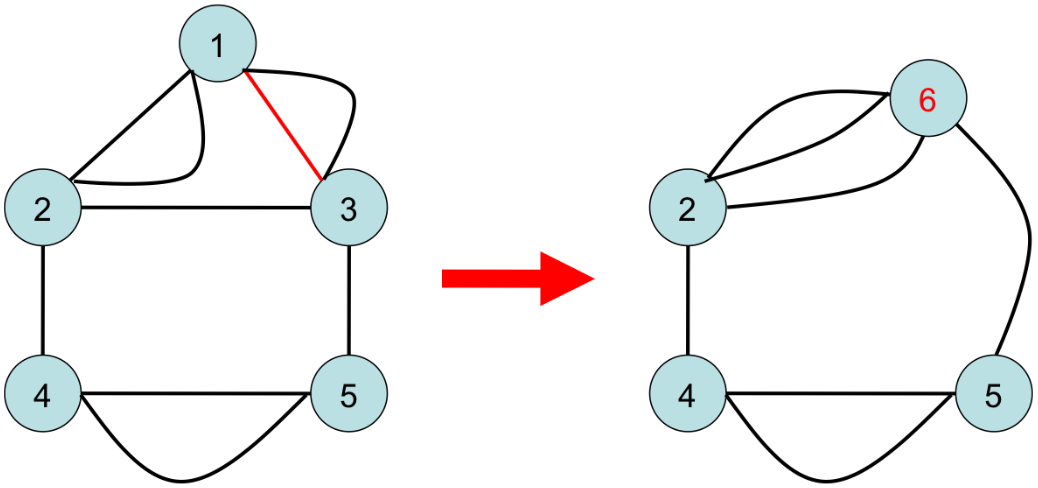

Figure 2: Illustrating edge contraction.

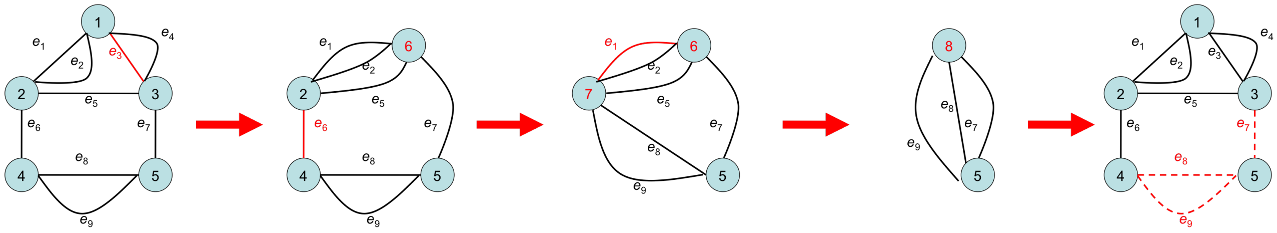

Figure 3: Illustrating the min-cut algorithm.

```
While $G$ has more than 2 vertices left do:
    Randomly pick an edge from $G$ and contract
Report all remaining edges
```

From Figure 3, it is clear that Karger's algorithm doesn't guarantee finding a min-cut. Let's analyze the probability that Karger's algorithm will find a min-cut.

For ease of presentation, let's assume that the graph G has n vertices and m edges, and size of the min-cut is k. If the algorithm finds the min-cut, this means that the min-cut will survive all the $n-2$ contrations. So, what's the probability that the min-cut survives the very first contration?

If the min-cut survives the first contration, then any one of the k edges of the min-cut can't be chosen for the first contration. Thus, the probability is $1 - \frac{k}{m}$. On the other hand, if the min-cut has a size k, every vertex in G has a degree $\geq k$. Thus $m \geq \frac{nk}{2}$. This implies

$$\frac{1}{m} \leq \frac{2}{nk}$$

Multiplying -1 on both sides, we have:

$$-\frac{1}{m} \geq -\frac{2}{nk}$$

Multiplying k on both sides, we have:

$$-\frac{k}{m} \geq -\frac{2}{n}$$

Adding 1 on both sides, we obtain:

$$1 - \frac{k}{m} \geq 1 - \frac{2}{n}$$

Thus, the probability that the min-cut survives the very first contration is $\geq 1 - \frac{2}{n}$.

Since this lower bound probabily only depends on the number of vertices, we can use it for other contrations. Hence the probability tha the min-cut survives the second contraction is $\geq 1 - \frac{2}{n-1}$, for the third contration is $\geq 1 - \frac{2}{n-2}$, ..., and for the final contration is $1 - \frac{2}{3}$. Thus the probability that the min-cut survives all $n-2$ contractions is

$$\geq \left(1 - \frac{2}{n-1}\right)\left(1 - \frac{2}{n-2}\right)\left(1 - \frac{2}{n-3}\right) ... \left(1 - \frac{2}{3}\right)$$
$$= \frac{n-2}{n} \cdot \frac{n-3}{n-1} \cdot \frac{n-4}{n-2} \cdot ... \cdot \frac{1}{3}$$
$$= \frac{2}{n(n-1)} = \frac{1}{\binom{n}{2}}$$

So, what is the probability of not finding the min-cut if we run the above algorithm $\binom{n}{2}$ times and report the minimum cut found? Note that the probability of failure for each individual run is $\leq 1 - \frac{1}{\binom{n}{2}}$, thus the algorithm fails in all $\binom{n}{2}$ runs is $\leq \left(1 - \frac{1}{\binom{n}{2}}\right)^{\binom{n}{2}}$.

Recall that for x small, $e^x \approx 1 + x$. Let $x = -\frac{1}{\binom{n}{2}}$, we have $1 - \frac{1}{\binom{n}{2}} \approx e^{-\frac{1}{\binom{n}{2}}}$. Thus

$$\left(1 - \frac{1}{\binom{n}{2}}\right)^{\binom{n}{2}} \approx \left(e^{-\frac{1}{\binom{n}{2}}}\right)^{\binom{n}{2}} \approx e^{-1}$$

If we run the algorithm $n^2 \ln n$ times, then the probability of failure will be $\left(e^{-1}\right)^{\ln n} = \frac{1}{n}$, which is arbitrarily small as $n \to \infty$.

The above Karger's algorithm is an example of a **Monte Carlo** algorithm. It has a deterministic running time, and guarantees the output is correct with a certain probability.