March 22

Probability Space $(S, \mathbb{E}, P)$

$$\mathbb{E} = 2^{S}$$

$P : \mathbb{E} \longrightarrow [0,1]$

$0 \leq p(E) \leq 1$

Axium of Probability

$0 \leq p(Z) \leq 1$

$p(\phi) = 0 \qquad p(S) = 1$

$\boxed{P(Z \cup F) = P(Z) + P(F)}$

$Z \cup F, \; E \cap F, \; \overline{E} \in \mathbb{E}$

Joint probability $\quad P(Z \cap F)$

Conditional Probability $\quad P(Z|F) = \dfrac{P(Z \cap F)}{P(F)}$

Bayes Theorem

$$P(E|F) = \boxed{\frac{P(F|E)\cdot P(E)}{P(F|E)\cdot P(E) + P(F|\bar{E})\cdot P(\bar{E})}}$$

Pf. Observation: $P(E|F) = \dfrac{P(E\cap F)}{\boxed{P(F)}}$

$$P(E\cap F) = P(E|F)\cdot P(F)$$

Similarly

$$P(F|E)\cdot P(E) = P(E\cap F)$$

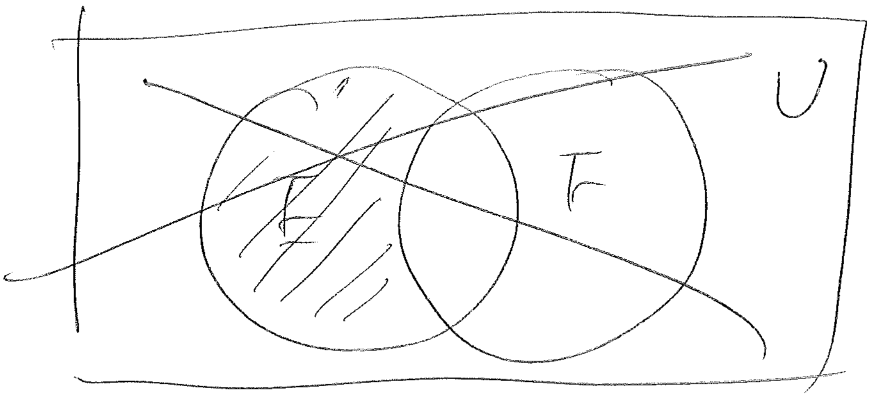
$$P(F|\bar{E})\cdot P(\bar{E}) = P(\bar{E}\cap F)$$

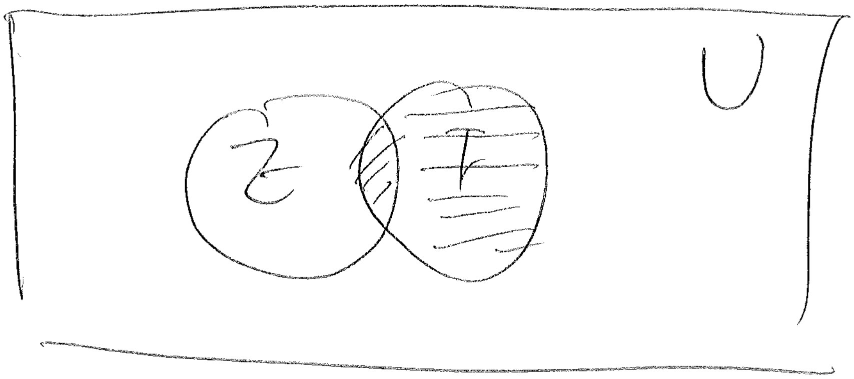
$$P(F|\bar{E}) = \frac{P(F\cap\bar{E})}{\boxed{P(\bar{E})}}$$

Back to the Theorem.

$$\frac{P(F|E)P(E)}{P(F|E)\cdot P(E) + P(F|\bar{E})\cdot P(\bar{E})} = \frac{P(E\cap F)}{P(E\cap F) + P(\bar{E}\cap F)}$$

$$= \frac{P(E\cap F)}{P(F)} = P(E|F)$$


E
F
U


E
F
U

Assume a rare disease.

The odds of having the disease is 1 out of 100,000

There is a dignostic test.

The test is correct 99% when administered to a subject with the disease.

The test is correct 99.5% when administered to a subject without the disease.

What's the probability that a given person testing positive actually has the disease?

Solution:

Let Z be the event that a given person has the disease, F be the event that a the test is positive, we want $P(Z|F)$

$$P(Z|F) = \frac{P(F|Z)\cdot P(Z)}{\boxed{P(F|Z)}\boxed{P(Z)}+\boxed{P(F|\bar{Z})}\boxed{P(\bar{Z})}}$$

$P(F|E)$: given the person has the disease

what's the probability of testing positive?

0.99

$P(F|\bar{E})$: given a health subject, what's the probabili

of testing positive?

$1 - 0.995 = 0.005$

$P(E) = 0.00009/$

$P(\bar{E}) = 1 - P(E)$

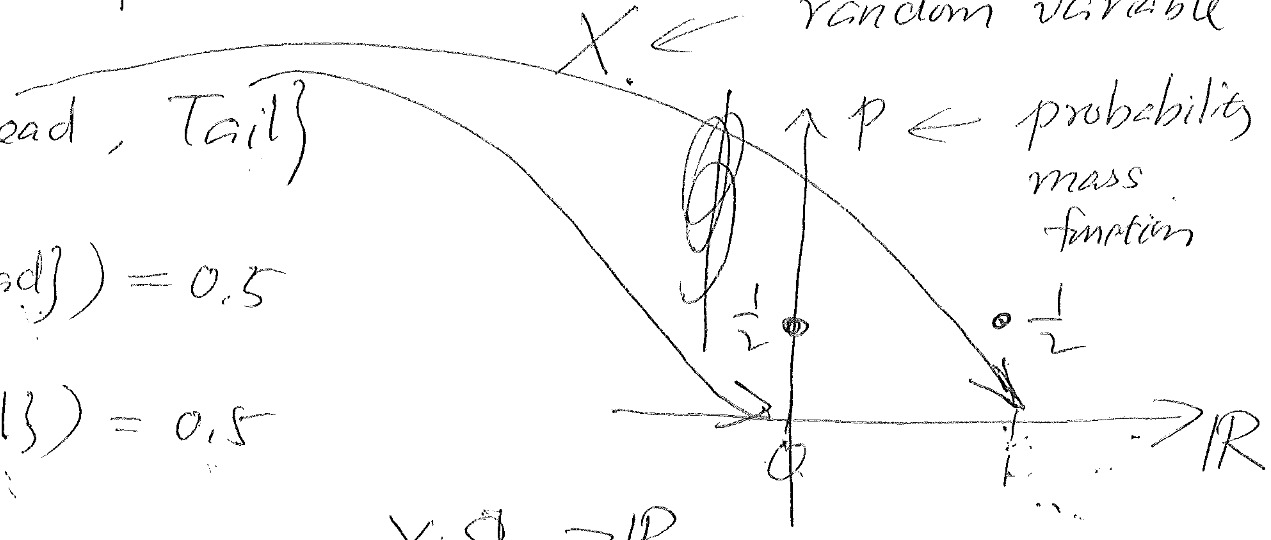
Random Variable : Not a variable but a function

Tossing a fair coin.

$S = \{\text{Head}, \text{Tail}\}$

$p(\{\text{Head}\}) = 0.5$

$p(\{\text{Tail}\}) = 0.5$



$X : S \longrightarrow \mathbb{R}$

Let X be a discrete R.V. with values $x_1, x_2, \cdots, x_n$ with probability mass function p_X.

$$P(\{x_1, \cdots, x_n\}) = \sum_{j=1}^{n} p_X(x_j) = 1$$

$$E[X] = p(x_1)x_1 + p(x_2)x_2 + \qquad + p(x_n)x_n$$

$$= \sum_{j=1}^{n} \boxed{p(x_j)}\,\boxed{x_j}$$

Variance

$$V(X) = \sum_{j=1}^{n} \left(X_j - E[X]\right)^2 p(X_j)$$

$$= E\left[(X - E[X])^2\right]$$

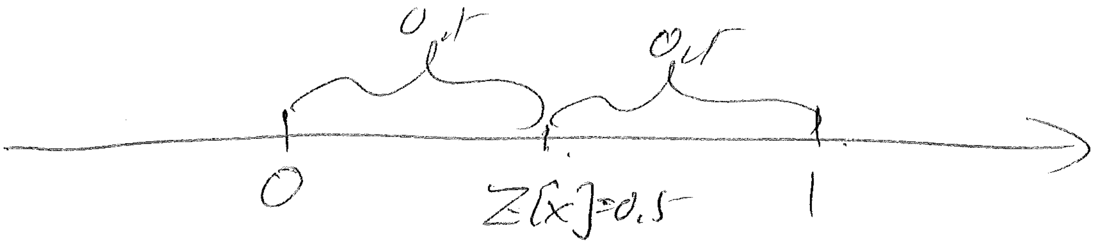
Consider a coin toss. R.V. X. $p(X=0)=0.5$, $p(X=1)=0.5$

$$E[X] = 0.5 \times 0 + 0.5 \times 1 = 0.5$$

$$Var(X) = 0.5\,(0-0.5)^2 + 0.5\,(1-0.5)^2$$

$$= \boxed{0.25}$$

$$\sigma(X) = \sqrt{Var(X)} = 0.5$$



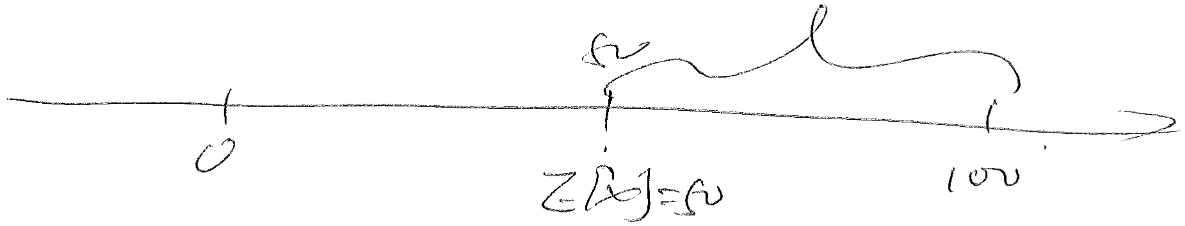
$$0.5\,(0-0.5) + 0.5\,(1-0.5)$$

$$= 0$$

$X \in \{0, 100\}$

$P(0)=0.5 \quad P(100)=0.5$

$\sigma = 50$



Markov Inequality

Let X be a discrete non-negative r.v. wich values $\{X_1, X_2, \cdots, X_n\}$,

Let $\delta > 0$, then

$$P(X \geq \delta) \leq \frac{E[X]}{\delta}$$

Pf. Assume $0 \le X_1 < X_2 < \cdots < X_n$

Then

$$X_1 < \cdots < X_j \overset{<}{\cancel{\le}} \delta \le X_{j+1} < \cdots < X_n$$

$$\boxed{E[x]} = \sum_{j=1}^{n} P(X_j) \cdot X_j$$

$$= \left(P(X_1) \cdot X_1 + P(X_2) \cdot X_2 + \cdots + P(X_j) X_j\right)$$

$$+ \left(P(X_{j+1}) X_{j+1} + \cdots + P(x_n) X_n\right)$$

$$\boxed{\ge} \; P(X_{j+1}) \boxed{X_{j+1}} + \cdots + P(x_n) \boxed{X_n}$$

$$\boxed{\ge} \; P(X_{j+1}) \delta + \cdots + P(x_n) \delta$$

$$= \left(\underbrace{P(X_{j+1}) + P(X_{j+1}) + \cdots + P(x_n)}\right) \delta$$

$$= P(\{X_{j+1}, \cdots, X_n\}) \delta$$

$$= \underline{P(X \ge \delta) \cdot \delta}$$

$$P(x \ge \delta) \cdot \delta \le E[x]$$

$$P(x \ge \delta) \le \frac{E[\cancel{X}]}{\delta}$$

The running time of randomized quick sort is $2 n \ln n$ in expectation

What's the probability that a particular run takes $\geq 20 n \ln n$ time?

Let X be the r.v. describing the running time of randomized quick sort.

X is nonnegative.

$E[X] = 2n \ln n$

$$P(X \geq 20 n \ln n) \leq \frac{2n \ln n}{20 n \ln n} = 0.1$$

Chebyshev Inequality

$$P(|X - E[X]| \ge \delta) \le \frac{Var(X)}{\delta^2}$$

Randomized Quick Sort.

Let X be the r.v. describing the run time

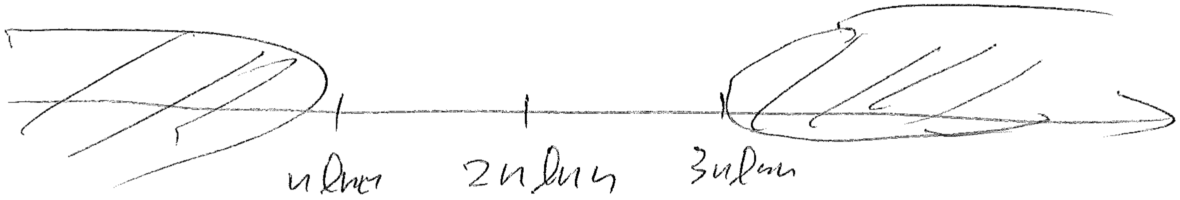
$E[X] = 2n\ln n$

~~Var(X)~~ $\sigma(X) = 0.65n$

$$P\Big(\big|X - \underset{\substack{\uparrow \\ E[X]}}{2n\ln n}\big| \ge \underset{\substack{\uparrow \\ \delta}}{n\ln n}\Big) \le \frac{(0.65n)^2}{(n\ln n)^2}$$

$$= \frac{0.4225}{(\ln n)^2}$$

For $n = 10{,}000$ this probability $\le 0.5\%$



Let X be an r.v. with values $\{X_1, X_2, \cdots X_m\}$, P_X

Y be another r.v. with values $\{Y_1, Y_2, \cdots, Y_n)$, P_Y

The joint r.v. the the r.v. (X, Y)

has values $\{X_1, \cdots, X_m\} \times \{Y_1, \cdots, Y_n)$

and a probability mass function $P_{X,Y}$

$$\sum_{j=1}^{n} P(X_i, Y_j) = P\{(X_i, Y_1), (X_i, Y_2), \cdots (X_i, Y_n)\}$$

$$= P_X(X_i) \leftarrow \text{marginal probability distr.}$$

$P\{(\text{Head}, 1)\ (\text{Head}, 2), (\text{Head}, 3), (\text{Head}, 4)$
$(\text{Head}, 5), (\text{Head}, 6)\}$

$= P(\text{Head})$

~~$P(X,Y)$~~

Let (X,Y) be a joint r.v.

Then X and Y are independent if

$$P_{X,Y} = P_X P_Y$$

$$P_{X,Y}(X_i, Y_j) = P_X(X_i) \cdot P_Y(Y_j)$$

Recall. $P(E \cap F) = P(E) \cdot P(F)$ E and F indep

The event $X = X_i$

$E = \{(X_i, Y_1) \cdots (X_i, Y_j), \cdots (X_i, Y_n)\}$

The event $Y = Y_j$

$F = \{(X_1, Y_j), (X_2, Y_j) \cdots (X_i, Y_j), \cdots (X_m, Y_j)\}$

$E \cap F = \{(X_i, Y_j)\}$

Sum of R.V.

Consider we are tossig two 6-sided fair ~~die~~ dice. independently

what's the probabiliy that sum is 11 ?

Let X be die #1 and Y be die #2

$$Z = X + Y$$

$$P(Z = 11) = ?$$

① what are the possible values of Z ?

$$\{2, 3, 4, \quad \cdots \quad, 12\}$$

$$P(Z=11) = \{\underline{(5,6)}, \quad \underline{(6,5)} \quad \}$$

$$= P(5,6) + P(6,5)$$

$$= P_X(5) \cdot P_Y(6) + P_X(6) \cdot P_Y(5)$$

$$= \frac{1}{6} \cdot \frac{1}{6} + \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{18}$$

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

$$Var(X) = E((X - E[X])^2)$$

$$= E(X^2 - 2XE[X] + (E[X])^2)$$

$$= E(X^2) - E[2 \times E[X]] + E((E[X])^2)$$

$$= E(X^2) - 2E[X] \cdot E[X] - (E[X])^2$$

$$= \underline{E[X^2)} - (E[X])^2$$

~~$X \sim 0, 1$~~

~~$E[X^2] =$~~

$$X \sim \{0, 100\}$$

$$X^2 \sim 0, \quad 10000$$

$$E[X^2] = 0.5 \times 0 + 0.5 \times 10000 = \underline{5000}.$$

$$E[X] = \underline{50}$$

$$(E[X])^2 = 2500$$

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.

Discussion questions & guide to presenting on Riccucci's *Unsung Heroes*

Instructions: Discuss your case/profile along the lines of the questions below and prepare a 6-7 minute (maximum) presentation that incorporates your answers. The times in parentheses are given to guide your presentation.

1. Introduce your leadership case profile (for presentation, not extensive discussion) (1 min.)
    a. What are the who, what, when, where, why & how of your case/profile? Give a brief introduction. Not everyone read your profile.
2. Discuss your leadership case profile in terms of the seven factors that Riccucci identifies as influencing execucratic effectiveness. Which of these factors played a significant role in the case? How so? (2-3 mins.)
    a. Factors: political skills, management and leadership skills, situational factors, experience in government, technical expertise, strategy, personality
    b. Give an example of how one of these factors worked in the case in your presentation. Share an example that you find particularly compelling. If you need to discuss a factor in relationship to others, you may.
3. What role did ethics and/or integrity play in your case? Were they integral? How so? (1 min.)
4. Is Riccucci's framework of execucratic leadership relevant to state and local level leaders? Why or why not? Explain your answer (1 min.)
5. What else do you think we should take away from this leadership profile and Riccucci's analysis? What are the lessons for emerging public sector leaders and managers (e.g. you)? (1 min.)

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.

$$\begin{aligned}
E[X+Y] = E[Z] &= \sum_{i=1}^{l} Z_i \cdot P(Z = Z_i) \\
&= \sum_{i=1}^{l} Z_i \sum_{X_j+Y_k=Z_i} P_{X,Y}(X = X_j, Y = Y_k) \\
&= \sum_{j=1}^{n}\sum_{k=1}^{m} P_{X,Y}(X = X_j, Y = Y_k)(X_j + Y_k) \\
&= \sum_{j=1}^{n}\sum_{k=1}^{m} P_{X,Y}(X = X_j, Y = Y_k)X_j + \sum_{j=1}^{n}\sum_{k=1}^{m} P_{X,Y}(X = X_j, Y = Y_k)Y_k \\
&= \sum_{j=1}^{n} X_j \sum_{k=1}^{m} P_{X,Y}(X = X_j, Y = Y_k) + \sum_{k=1}^{m} Y_k \sum_{j=1}^{n} P_{X,Y}(X = X_j, Y = Y_k) \\
&= \sum_{j=1}^{n} X_j P_X(X = X_j) + \sum_{k=1}^{m} Y_k P_Y(Y = Y_k)
\end{aligned}$$

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

□

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 $P(X = 0) = 1 - p$.*

$$T_n = (1-p)^0 p + (1-p)p\cdot 2^2 + (1-p)^2 p\cdot 3^2 + \cdot$$

$$(1-p)T_n = \qquad (1-p)p \; + (1-p)^2 p 2^2 + \cdots$$

$-$

$$pT_n = (1-p)p + (1-p)p\cdot 3 + (1-p)^2 p\cdot 5$$

$$= \sum_{j=1}^{\infty} (1-p)^{j-1} p(2j-1)$$

$$= 2\sum_{j=1}^{\infty} (1-p)^{j-1} p\, j - \sum_{j=1}^{\infty} (1-p)^{j-1} p$$

$$= 2\,\frac{1}{p} \qquad - \; 1$$

$$T_n = \frac{2}{p^2} - \frac{1}{p}$$

$$Var(x) = \cancel{(E(x))^2 - E(x)}\; E(x^2) - (E(x))^2$$

$$= \left(\frac{2}{p^2} - \frac{1}{p}\right) - \frac{1}{p^2}$$

$$= \frac{1}{p^2} - \frac{1}{p}$$

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:

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

$$E[X^2] = \sum_{k=0}^{\infty} \left(\frac{\lambda^k e^{-\lambda}}{k!} \cdot k^2\right)$$

$$= \sum_{k=0}^{\infty} \frac{\lambda^k e^{-\lambda}}{(k-1)!} \cdot k$$

$$= \sum_{k=0}^{\infty} \frac{\lambda^k e^{-\lambda}}{(k-1)!} ((k-1)+1)$$

$$= \lambda \sum_{k=0}^{\infty} \frac{\lambda^{k-1} e^{-\lambda}}{(k-1)!} (k-1) + \lambda \sum_{k=0}^{\infty} \frac{\lambda^{k-1} e^{-\lambda}}{(k-1)!}$$

$$= \lambda^2 + \lambda$$

$$Var(X) = E[X^2] - (E[X])^2 = (\lambda^2 + \lambda) - \lambda^2 = \lambda$$

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 $X_0, X_1.X_2, \ldots$ 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.*

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. □

$$
\begin{aligned}
&= \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{2}{j-i+1} \\
&= \sum_{i=1}^{n-1} \sum_{k=j-i=1}^{n-i} \frac{2}{k+1} \\
&\le \sum_{i=1}^{n-1} \sum_{k=1}^{n} \frac{2}{k} \\
&= \sum_{i=1}^{n-1} 2 \ln n = 2n \ln n
\end{aligned}
$$

5.1 Harmonic Sequence

Definition 19 *The sequence* $1, \frac{1}{2}, \frac{1}{3}, \ldots, \frac{1}{n}, \ldots$ *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:

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

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: $P(|X - 2n \ln n| \ge n \ln n) \le \frac{(0.65n)^2}{(n \ln n)^2} = \frac{0.4225}{\ln^2 n}$. For $n = 10,000$, the probability of the $X \le n \ln n$ or $\ge 3n \ln n$ is less than 0.5%.

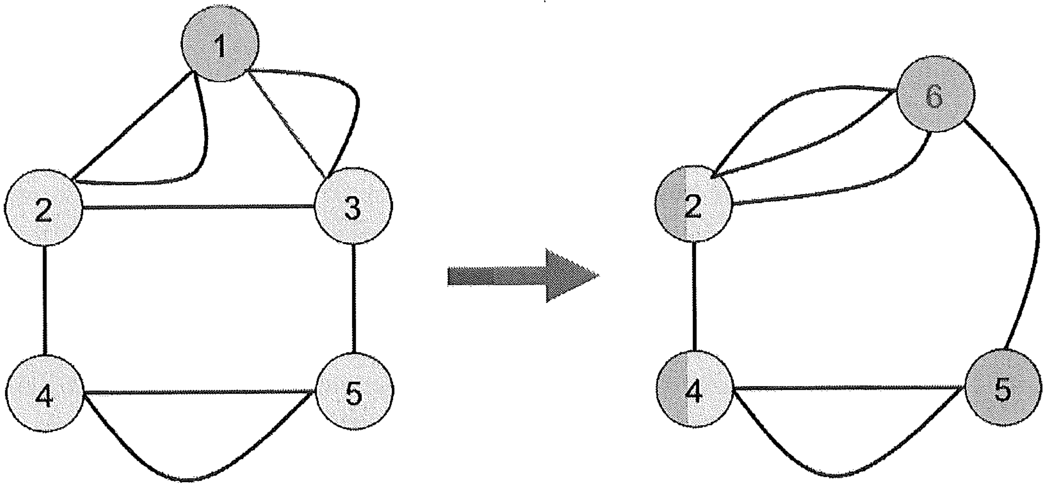


Figure 2: Illustrating edge contraction.

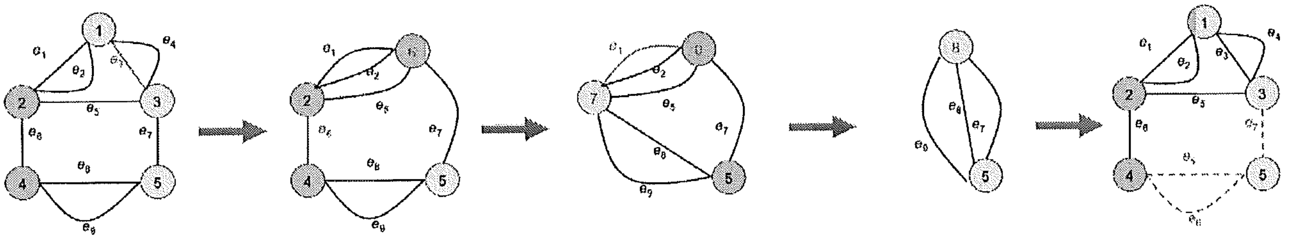


Figure 3: Illustrating the min-cut algorithm.

```
1. While $G$ has more than 2 vertices left do:
2.     Randomly pick an edge from $G$ and contract
3. 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:

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

7 Randomized Linear Programming

Problem 1 *The 2D* **linear programming** *problem is the following constrained optimization problem:*

$$\begin{array}{rl} \textit{maximize} & c_1x_1 + c_2x_2 \\ \textit{subject to:} & a_{11}x_1 + a_{12}x_2 \le b_1 \\ & a_{21}x_1 + a_{22}x_2 \le b_2 \\ & \vdots \\ & a_{n1}x_1 + a_{n2}x_2 \le b_2 \end{array}$$

Let $x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$, $c = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}$, and $A = \begin{pmatrix} a_{11} & a_{12} \\ \vdots & \vdots \\ a_{n1} & a_{n2} \end{pmatrix}$, then we have the matrix format of the LP:

$$\begin{array}{rl} \text{maximize} & c^T x \\ \text{subject to:} & Ax \le b \end{array}$$

7.1 Inner Product and Cross Product

Consider two 2D vectors $v_1 = \begin{pmatrix} r_1 \cos\alpha \\ r_1 \sin\alpha \end{pmatrix}$ and $v_2 = \begin{pmatrix} r_2 \cos\beta \\ r_2 \sin\beta \end{pmatrix}$.

Recall the **inner product** of these two vectors are defined as $r_1r_2 \cos\alpha \cos\beta + r_1r_2 sin\alpha \sin\beta$.

Since $r_1r_2 \cos\alpha\cos\beta + r_1r_2 sin\alpha \sin\beta = r_1r_2 (\cos\alpha\cos\beta + r_1r_2 sin\alpha \sin\beta) = r_1r_2 \cos(\alpha - \beta)$, geometrically, the inner product between two vectors v_1 and v_2 is simply the product of their lengths multiplied by the cosine of the angle between them. Since cosine is an even function, inner product for real vectors are commutative.

Recall the **cross product**, denoted by $v_1 \times v_2$ is the vector $|r_1r_2 \cos\alpha \sin\beta - r_1r_2 \sin\alpha \cos\beta| \vec{n}$, where $\vec{n}$ is the directional vector given by the **righthand rule**, i.e., with the index finger pointing along $\vec{v_1}$ and middle finger pointing along $\vec{v_2}$.

Since $|r_1r_2 \cos\alpha \sin\beta - r_1r_2 \sin\alpha \cos\beta| = |r_1r_2 \sin(\beta - \alpha)|$. Thus, the cross product calculates the area defined by the two vectors.

The cross product is also closely related to the **determinant** of the matrix $\left|\begin{pmatrix} r_1 \cos\alpha & r_2 \cos\beta \\ r_1 \sin\alpha & r_2 \sin\beta \end{pmatrix}\right| = r_1r_2 \sin(\beta - \alpha)$. Note that if $\beta \ge \alpha$ and $\beta - \alpha \le 180°$, then $\sin(\beta - \alpha) > 0$, and $\vec{v_2}$ is counter clockwise from $\vec{v_1}$.

7.2 Gradient Search

Recall that if $f(x)$ be a function of $n-$dimensional vectors $x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$, the gradient of f denoted by ∇f is $\begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{pmatrix}$. As an example, the gradient of the linear function $f(x) = c_1x_1 + c_2x_2$ is $\nabla f = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}$

The positive halfspace is the collection of 2D points p, such that the vector $\overrightarrow{p_1p}$ is within 90° to $\vec{n}$, i.e., the inner product $< \vec{n}, \overrightarrow{p_0p} > \geq 0$. (See Figure 7.3 (b) for illustrations.) In other words, $\left\langle \begin{pmatrix} a \\ b \end{pmatrix}, \begin{pmatrix} x - x_0 \\ y - y_0 \end{pmatrix} \right\rangle > 0$. Hence $a(x - x_0) + b(y - y_0) > 0$, i.e., $ax + by - (ax_0 + by_0) > 0$. Now if we let $c = -(ax_0 + by_0)$, then the positve halfspace are all the points satisfying the inequality $ax + by + c > 0$.

Similarly, the negative halfspace is defined by the inequality $ax + by + c < 0$.

The inequality $ax + by + c \geq 0$ is obviously the positive halfspace plus the dividing straight line $ax + by + c = 0$.

Note that since we are using vector operations, all of the above can be generalized to high dimensions. More precisely, an $n-$dimensional hyperplane is defined by the function $a_1x_1 + a_2x_2 + ... + a_nx_n = b$. The positive halfspace in $n-$dimension is defined as $a_1x_1 + a_2x_2 + ... + a_nx_n \geq b$

7.4 Geometry of LP

Consider the following example:

$$\begin{array}{rl} \max & f(x) = x_1 + x_2 \\ \textit{subject to} & 2x_1 + 3x_2 \leq 12 \\ & x_1 \leq 4 \\ & x_2 \leq 3 \\ & x_1 \geq 0 \\ & x_2 \geq 0 \end{array}$$

We know that each constraint in an LP specifies a halfplane. The **feasible domain** of the LP is in fact the intersection of all these half spaces. Since each half space is convex, the intersection between two convex sets is also convex, the feasible domain of an LP is also convex. In 2D space, if the feasible domain is bounded, then the feasble domain is a convex polygon. (See Figure 5 for illustrations.)

Now suppose we have located a feasible point x. To determine if x is optimal, we ask the question can be find another feasible point $x' = x + \Delta x$, such that $f(x') > f(x)$? If no such point x' exists, then obviously x is optimal. From the section on gradient search, we know that in order for $f(x') > f(x)$, Δx must be an **ascent** and **feasible** direction, where "ascent" implies it is within 90° to the gradient ∇f, and "feasible" implies $x + \Delta x$ still belongs to the feasible domain.

Observe that if x is in the **interior** of the feasible domain, then all 360° around x are feasible, for non-zerolinear functions $f(x)$, **feasible ascent** directions always exist. Thus the optimal solution to the LP, if exists, must lie on the boundary of the feasible domain.

Suppose x^* is an optimal solution to the LP on a boundary edge $\overline{p_1p_2}$, where p_1 and p_2 are two extreme vertices of the feasible domain. Since x^* is optimal, then ∇f must be perpendicular to the edge $\overline{p_1p_2}$. Thus, all points on this edge all optimal, and therefore, one of the extreme vertices of the feasible domain must be an optimal solution (if exsits) to the LP. This is the basis of the famous **simplex** method, which iterate through all the extreme points of the feasible domain before terminating at the optimal solution.

Let X_j be the R.V. describing the running time of the $j-$iteration. X_{j-1} has two possible values. When $v_{j-1} = v_j$, $X_j = 1$, and when $v_{j-1} \neq v_j$, $X_j = j$. The running time of the whole algorithm is $\sum_{j=1}^{n} X_j$, and the expected running time is: $E\left[\sum_{j=1}^{n} X_j\right] = \sum_{j=1}^{n} E\left[X_j\right]$.

Since $E\left[X_j\right] = j \cdot Pr\left(v_{j-1} \neq v_j\right) + 1 \cdot Pr\left(v_{j-1} = H_j\right)$, in order the figure out this expectation, we need to know the probability of $Pr\left(v_{j-1} \neq v_j\right)$.

Suppose that v_j is the lines of the two lines l_s and l_t, where l_s and l_t are the lines defining the half spaces h_s and h_t from $\{h_1, h_2, ..., h_j\}$. Notice that in order for v_j to be not equal to v_{j-1}, h_j must be either h_s or h_t. Since we randomly perturb the order of the constraints, every halfspace is equally likely to be h_j. Thus, the probability that h_j is either h_s or h_t is $\frac{2}{j}$, i.e., $Pr\left(v_{j-1} \neq v_j\right) = \frac{2}{j}$. Hence, $E\left[X_j\right] = j \cdot \frac{2}{j} + +1 \cdot \left(1 - \frac{2}{j}\right) = 2 + \frac{j-2}{j} \leq 3$. Thus $\sum_{j=1}^{n} E\left[X_j\right] = O(n)$.

Lemma 1 *If $b \equiv c \mod n$, then $m|(b-c)$.*

Lemma 2 *If $m|(b-c)$, then $b \equiv c \mod n$.*

Lemma 3 *Given integers a, b, p, if $GCD(a,p) = 1$ and $GCD(b,p) = 1$, then $GCD(ab,p) = 1$.*

Observe that we have $p(p-1)$ hash function, i.e., $|H| = p(p-1)$. We have to show that for arbitrary pair of distinct keys $x \neq y$, the number of hash functions h from H can gives a collision (i.e., h(x) = h(y)) is $\leq \frac{|H|}{n} = \frac{p(p-1)}{n}$.

Consider an arbitrary hash function $h(a,b) \in H$. Let $s = ax + b \mod p$ and $t = ay + b \mod p$. Notice that if $s = t$, then $(ax+b) \equiv (ay+b) \mod p$. In other words, $p|\left((ax+b) - (ay+b)\right)$, i.e., $p|a(x-y)$. However, since $x \neq y$ and $x - y \in \{-(M-1), -M-2, ..., -1, 1, ..., M-2, M-1\}$, hence $GCD(p, x-y) = 1$. On the other hand, $a \in \{1, 2, ..., p-1\}$, $GCD(a,p) = 1$. From one of the lemmas, we know that $GCD(a(x-y), p) = 1$. A contradiction, thus $s \neq t$.

Thus distinct (x,y) are mapped to distinct (s,t) by the function $ax + b \mod p$. If $h(x) = h(y)$, then $s \mod n = t \mod n$.

Now consider two functions $h_1(a_1,b_1), h_2(a_2,b_2) \in H$, with $h_1 \neq h_2$, i.e., $(a_1,b_1) \neq (a_2,b_2)$. Let $s_1 = a_1x + b_1$ and $t_1 = a_1y + b_1$. Let $s_2 = a_2x + b_2$ and $t_2 = a_2y + b_2$. We will show that $(s_1,t_1) \neq (s_2,t_2)$. This implies that for fixed x, y, there is a one-to-one corresponence between the two sets $\{(s,t) | s \neq t, s, t \in \{0, 1, ..., p-1\}\}$ to H. Thus, the number of hash functions from H that will cause a collision is the number of (s,t) pairs such that $s \equiv t \mod n$.

Since $(a_1,b_1) \neq (a_2,b_2)$, there are three possibilities:

(1) $a_1 = a_2$, $b_1 \neq b_2$:

$s_1 - s_2 = (a_1x + b_1) - (a_2x + b_2) \mod p = (b_1 - b_2) \mod p \neq 0$. Hence $s_1 \neq s_2$, and $(s_1,t_1) \neq (s_2,t_2)$.

(2) $a_1 \neq a_2$, $b_1 = b_2$:

$s_1 - s_2 = (a_1x + b_1) - (a_2x + b_2) \mod p = (a_1 - a_2)x \mod p$.

$t_1 - t_2 = (a_1y + b_1) - (a_2y + b_2) \mod p = (a_1 - a_2)y \mod p$.

Since $x \neq y$, either at least one of x and y is nonzero. Thus, at least one of $(a_1 - a_2)x \mod p$ and $(a_1 - a_2)y \mod p$ is non-zero. Hence either $s_1 \neq s_2$ or $t_1 \neq t_2$, and $(s_1,t_1) \neq (s_2,t_2)$.

(3) $a_1 \neq a_2$, $b_1 \neq b_2$:

We will use proof by contradiction here, and assume that $s_1 = s_2$ and $t_1 = t_2$. Thus we have:

$$\begin{cases} a_1x + b_1 \equiv a_2x + b_2 \mod p \\ a_1y + b_1 \equiv a_2y + b_2 \mod p \end{cases} \quad (1)$$

This gives:

$$\begin{cases} p|(a_1 - a_2)x + (b_1 - b_2) \\ p|(a_1 - a_2)y + (b_1 - b_2) \end{cases} \quad \text{Equation (2)}$$

This gives:

$$p|(a_1 - a_2)(x - y) \quad \text{Equation (3)}$$

9 String Matching and Rolling Hash

Problem 2 *Given a string $s[1..n]$ and a pattern $p[1..m]$, $m < n$ over some alphabet Σ, determine if the pattern p occurs in the string s.*

For example, let $s =$ "$ACGTCGTA$" and $p =$ "TCG", then the answer will be yes.

9.1 The Shifting Naive Algorithm

```
for $j = 1$ to $n$ do:
    for $k = 1$ to $m$ do:
        if $s[j + k - 1] \neq p[k]$
            break ;
        else
            return Yes and $j$
```

Obviously, the running time of the naive shifting string search algorithm is $O(mn)$. If m and n are both large, then this quadratic and is very expensive. The inner loop of the algorithm takes $O(m)$ time, is it possible to reduce the inner loop to constant time?

9.2 The Rabin-Karp Algorithm

```
$hp = hash(p[1..m])$
for $j = 1$ to $n$ do:
    $hs = hash(s[j..j + m - 1])$
        if $hs = hp$
            return Yes and $j$
```

In order to hash a string of length m, we need to convert the string to a number. The easiest way to do it is to convert the number to base$-|Simga|$ number. More specifically, let $b = |\Sigma|$, then we can convert a string of length m such as $p[1..m]$ to the interger $p[1]b^{m-1} + p[2]b^{m-2} + ... + p[m-1]b + p[m]$.

Since hashing a string of size m will take at least $O(m)$ time, the Rabin Karp algroithm doesn't appear to be any improvement. However, if one considers two successive iterations, i.e., hashing $s[j..j+m-1]$ and $s[j+1..j+m]$, the two strings only differs by one character. Is it possible to update the hash result instead of re-compute the hash from stractch? This is the idea of the rolling hashing. The goal of rolling hash is to calculate $hashs[j + 1..j + m]$ from $hashs[j..j + m - 1]$ in constant time.