Information Theory & Product Cipher

Presentation by:
V. Balasubramanian
SSN College of Engineering

Elementary probability

EXAMPLE 1.1 If we toss a coin, the result of the experiment is that it will either come up "tails," symbolized by T (or 0), or "heads," symbolized by H (or 1), i.e., one of the elements of the set $\{H, T\}$ (or $\{0, 1\}$).

EXAMPLE 1.2 If we toss a die, the result of the experiment is that it will come up with one of the numbers in the set $\{1, 2, 3, 4, 5, 6\}$.

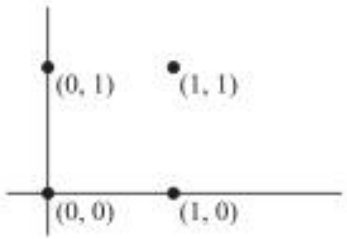
EXAMPLE 1.3 If we toss a coin twice, there are four results possible, as indicated by $\{HH, HT, TH, TT\}$, i.e., both heads, heads on first and tails on second, etc.

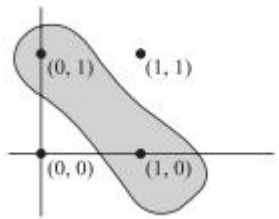
EXAMPLE 1.4 If we are making bolts with a machine, the result of the experiment is that some may be defective. Thus when a bolt is made, it will be a member of the set {defective, nondefective}.

EXAMPLE 1.5 If an experiment consists of measuring "lifetimes" of electric light bulbs produced by a company, then the result of the experiment is a time t in hours that lies in some interval—say, $0 \leq t \leq 4000$—where we assume that no bulb lasts more than 4000 hours.

Sample Space

- If we toss a die, one sample space, or set of all possible outcomes, is given by {1, 2, 3, 4, 5, 6} while another is {odd, even}. It is clear, however, that the latter would not be adequate to determine, for example, whether an outcome is divisible by 3.
- An event is a subset A of the sample space S, i.e., it is a set of possible outcomes.





Axiomatic approach

Axiom 1 For every event A in the class C,

$$P(A) \geq 0$$

Axiom 2 For the sure or certain event S in the class C,

$$P(S) = 1$$

Axiom 3 For any number of mutually exclusive events $A_1, A_2, \ldots,$ in the class C,

$$P(A_1 \cup A_2 \cup \cdots) = P(A_1) + P(A_2) + \cdots$$

In particular, for two mutually exclusive events A_1, A_2,

$$P(A_1 \cup A_2) = P(A_1) + P(A_2)$$

Axioms

Theorem 1-1 If $A_1 \subset A_2$, then $P(A_1) \leq P(A_2)$ and $P(A_2 - A_1) = P(A_2) - P(A_1)$.

Theorem 1-2 For every event A,

$$0 \leq P(A) \leq 1,$$

i.e., a probability is between 0 and 1.

Theorem 1-3 $$P(\varnothing) = 0$$

i.e., the impossible event has probability zero.

Example

EXAMPLE 1.12 A single die is tossed once. Find the probability of a 2 or 5 turning up.

The sample space is $S = \{1, 2, 3, 4, 5, 6\}$. If we assign equal probabilities to the sample points, i.e., if we assume that the die is fair, then

$$P(1) = P(2) = \cdots = P(6) = \frac{1}{6}$$

The event that either 2 or 5 turns up is indicated by $2 \cup 5$. Therefore,

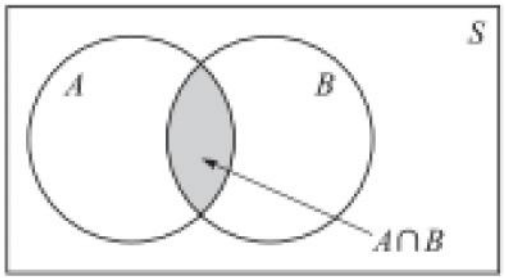
$$P(2 \cup 5) = P(2) + P(5) = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}$$

Conditional Probability

- Let *A and B be two events such that* $P(A) \geq 0$. *Denote by* $P(B|A)$ *the probability of B given that A* has occurred. Since *A is known to have occurred, it becomes the new sample space replacing the original S.*

$$P(B \mid A) \equiv \frac{P(A \cap B)}{P(A)}$$

$$P(A \cap B) \equiv P(A)\, P(B \mid A)$$



Example

EXAMPLE 1.13 Find the probability that a single toss of a die will result in a number less than 4 if (a) no other information is given and (b) it is given that the toss resulted in an odd number.

(a) Let B denote the event {less than 4}. Since B is the union of the events 1, 2, or 3 turning up, we see by Theorem 1-5 that

$$P(B) = P(1) + P(2) + P(3) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2}$$

assuming equal probabilities for the sample points.

(b) Letting A be the event {odd number}, we see that $P(A) = \frac{3}{6} = \frac{1}{2}$. Also $P(A \cap B) = \frac{2}{6} = \frac{1}{3}$. Then

$$P(B \mid A) = \frac{P(A \cap B)}{P(A)} = \frac{1/3}{1/2} = \frac{2}{3}$$

Hence, the added knowledge that the toss results in an odd number raises the probability from 1/2 to 2/3.

- If *P(B|A)= P(B), i.e., the probability of B occurring is not affected by the occurrence or non-occurrence of A,* then we say that *A and B are independent events. This is equivalent to*
- $P(A \cap B) = P(A).P(B)$

Random Variable

Definition 3.1: A ***discrete random variable***, say $\mathbf{X}$, consists of a finite set X and a ***probability distribution*** defined on X. The probability that the random variable $\mathbf{X}$ takes on the value x is denoted $\mathbf{Pr}[\mathbf{X} = x]$; sometimes we will abbreviate this to $\mathbf{Pr}[x]$ if the random variable $\mathbf{X}$ is fixed. It must be the case that $0 \leq \mathbf{Pr}[x]$ for all $x \in X$, and

$$\sum_{x \in X} \mathbf{Pr}[x] = 1.$$

Example

Example 3.1 Suppose we consider a random throw of a pair of dice. This can be modeled by a random variable **Z** defined on the set

$$Z = \{1,2,3,4,5,6\} \times \{1,2,3,4,5,6\},$$

where $\mathbf{Pr}[(i,j)] = 1/36$ for all $(i,j) \in Z$. Let's consider the sum of the two dice. Each possible sum defines an event, and the probabilities of these events can be computed using equation (3.1). For example, suppose that we want to compute the probability that the sum is 4. This corresponds to the event

$$S_4 = \{(1,3),(2,2),(3,1)\},$$

and therefore $\mathbf{Pr}[S_4] = 3/36 = 1/12$.

The probabilities of all the sums can be computed in a similar fashion. If we denote by S_j the event that the sum is j, then we obtain the following: $\mathbf{Pr}[S_2] = \mathbf{Pr}[S_{12}] = 1/36$, $\mathbf{Pr}[S_3] = \mathbf{Pr}[S_{11}] = 1/18$, $\mathbf{Pr}[S_4] = \mathbf{Pr}[S_{10}] = 1/12$, $\mathbf{Pr}[S_5] = \mathbf{Pr}[S_9] = 1/9$, $\mathbf{Pr}[S_6] = \mathbf{Pr}[S_8] = 5/36$, and $\mathbf{Pr}[S_7] = 1/6$.

Joint Probability

Definition 3.2: Suppose **X** and **Y** are random variables defined on finite sets X and Y, respectively. The ***joint probability*** $\mathbf{Pr}[x, y]$ is the probability that **X** takes on the value x and **Y** takes on the value y. The ***conditional probability*** $\mathbf{Pr}[x|y]$ denotes the probability that **X** takes on the value x given that **Y** takes on the value y. The random variables **X** and **Y** are said to be ***independent random variables*** if $\mathbf{Pr}[x, y] = \mathbf{Pr}[x]\mathbf{Pr}[y]$ for all $x \in X$ and $y \in Y$.

Bayes Theorem

$$\mathbf{Pr}[x, y] = \mathbf{Pr}[x|y]\mathbf{Pr}[y].$$

Interchanging x and y, we have that

$$\mathbf{Pr}[x, y] = \mathbf{Pr}[y|x]\mathbf{Pr}[x].$$

$$\mathrm{P_r}[x|y]\,P_r[y] = \mathrm{P_r}[\mathrm{y}|x]\mathrm{P_r}[x]$$

$$\mathbf{Pr}[x|y] = \frac{\mathbf{Pr}[x]\mathbf{Pr}[y|x]}{\mathbf{Pr}[y]}.$$

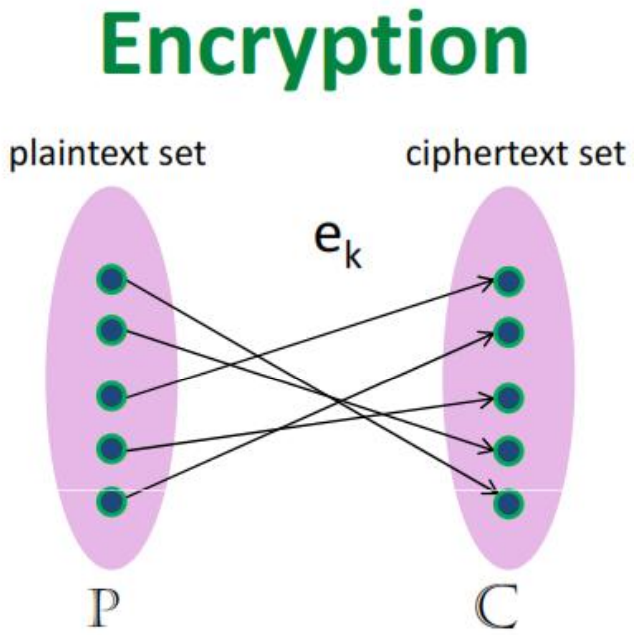
Example

- Suppose we consider a random throw of a pair of dice. Let **X be the** random variable defined on the set $X = \{2,3, \ldots 12\}$, obtained by considering the sum of two dice. Further, suppose that Y is a random variable which takes on the value D if the two dice are the same (i.e., if we throw "doubles"), and the value N, otherwise.

Solution

- $P_r[D] = \frac{1}{6}$ $\quad [(1,1),(2,2),(3,3),(4,4),(5,5),(6,6)]$
- $P_r[N] = \frac{5}{6}$
- $P_r[D|4] = \frac{P_r[D \cap 4]}{P_r[4]} = \frac{\frac{1}{36}}{\frac{3}{36}} = \frac{1}{3}$
- $P_r[4|D] = \frac{P_r[4 \cap D]}{P_r[D]} = \frac{\frac{1}{36}}{\frac{1}{6}} = \frac{1}{6}$
- $P_r(4, D) = P_r(4|D).P_r(D) = P_r(D|4).P_r(4) = 1/36$

Encryption



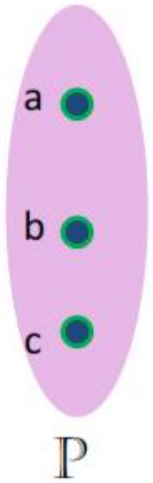
- For a given key, the encryption (e_k) defines an injective mapping between the plaintext set ($\mathbb{P}$) and ciphertext set ($\mathbb{C}$)
- We assume that the key and plaintext are independent

- Alice picks a plaintext $x \in P$ and encrypts it to obtain a ciphertext $y \in C$

Plaintext Distribution

Plaintext Distribution

- Let **X** be a discrete random variable over the set $\mathbb{P}$
- Alice chooses x from $\mathbb{P}$ based on some probability distribution
  - Let *Pr[**X** = x]* be the probability that x is chosen
  - This probability may depend on the language



Plaintext set
Pr[**X**=a] = 1/2
Pr[**X**=b] = 1/3
Pr[**X**=c] = 1/6

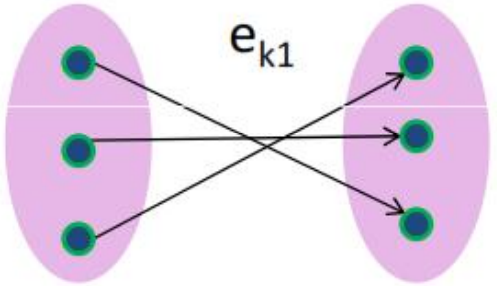
Note : Pr[a] + Pr[b] + Pr[c] = 1

Key Distribution

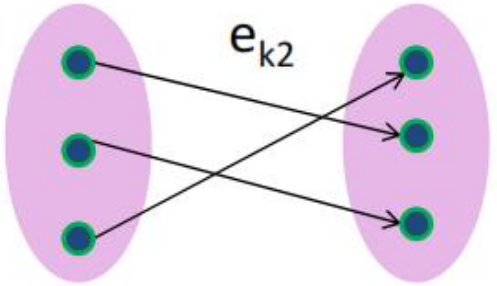
Key Distribution

- Alice & Bob agree upon a key k chosen from a key set K
- Let ***K*** be a random variable denoting this choice

keyspace
$\Pr[\mathbf{K}=k_1] = ¾$
$\Pr[\mathbf{K}=k_2] = ¼$



There are two keys in the keyset
thus there are two possible encryption mappings



Cipher Text Distribution

- Let **Y** be a discrete random variable over the set $\mathbb{C}$
- The probability of obtaining a particular ciphertext y depends on the plaintext and key probabilities

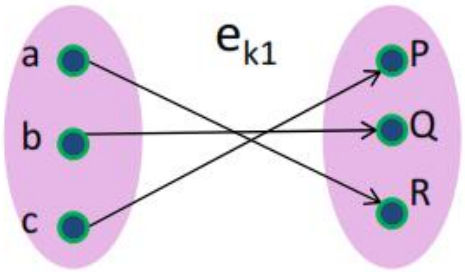
$$\Pr[Y = y] = \sum_k \Pr(k) \Pr(d_k(y))$$

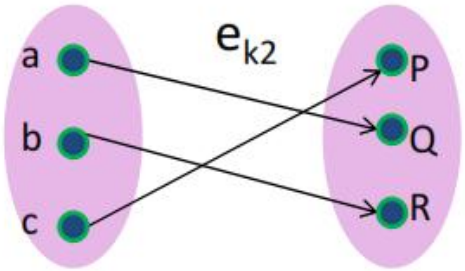
Pr[Y = P] = $\Pr(k_1) * \Pr(c) + \Pr(k_2) * \Pr(c)$
= $(3/4 * 1/6) + (1/4 * 1/6) =$ **1/6**

Pr[Y = Q] = $\Pr(k_1) * \Pr(b) + \Pr(k_2) * \Pr(a)$
= $(3/4 * 1/3) + (1/4 * 1/2) =$ **3/8**

Pr[Y = R] = $\Pr(k_1) * \Pr(a) + \Pr(k_2) * \Pr(b)$
= $(3/4 * 1/2) + (1/4 * 1/3) =$ **11/24**

Note: Pr[Y=P] + Pr[Y=Q] + Pr[Y=R] = 1





plaintext
Pr[**X**=a] = 1/2
Pr[**X**=b] = 1/3
Pr[**X**=c] = 1/6

keyspace
Pr[**K**=k_1] = ¾
Pr[**K**=k_2] = ¼

Attackers Probability

- The attacker wants to determine the plaintext ***x***
- Two scenarios
  - Attacker does not have y (a priori Probability)
    - Probability of determining ***x*** is simply ***Pr[x]***
    - Depends on plaintext distribution (eg. Language charcteristics)
  - Attacker has y (a posteriori probability)
    - Probability of determining x is simply ***Pr[x|y]***

Posteriori Probability

- How to compute the attacker's a posteriori probabilities? $\Pr[X = x \mid Y = y]$
  - Bayes' Theorem

$$\Pr[x \mid y] = \frac{\Pr[x] \times \Pr[y \mid x]}{\Pr[y]}$$

probability of the plaintext

probability of this ciphertext

?

The probability that y is obtained given x depends on the keys which provide such a mapping

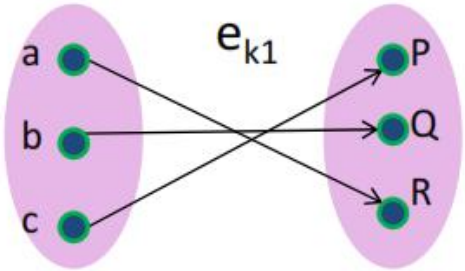
$$\Pr[y \mid x] = \sum_{\{k \,:\, d_k(y)=x\}} \Pr[k]$$

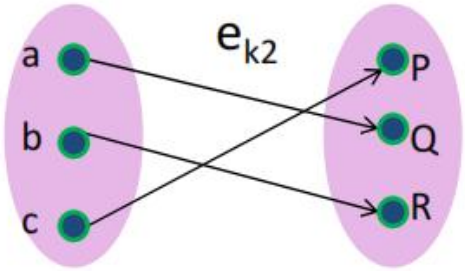
P[Y|X]

$Pr[P|a] = 0$
$Pr[P|b] = 0$
$Pr[P|c] = 1$

$Pr[Q|a] = Pr[k_2] = ¼$
$Pr[Q|b] = Pr[k_1] = ¾$
$Pr[Q|c] = 0$

$Pr[R|a] = Pr[k_1] = ¾$
$Pr[R|b] = Pr[k_2] = ¼$
$Pr[R|c] = 0$





keyspace
$Pr[\mathbf{K}=k_1] = ¾$
$Pr[\mathbf{K}=k_2] = ¼$

Computing Posteriori Probability

$$\Pr[x \mid y] = \frac{\Pr[x] \times \Pr[y \mid x]}{\Pr[y]}$$

plaintext	ciphertext	Pr[y|x]
Pr[**X**=a] = 1/2	Pr[**Y**=P] = 1/6	Pr[P|a] = 0 Pr[P|b] = 0 Pr[P|c] = 1
Pr[**X**=b] = 1/3	Pr[**Y**=Q] = 3/8	Pr[Q|a] = ¼ Pr[Q|b] = ¾ Pr[Q|c] = 0
Pr[**X**=c] = 1/6	Pr[**Y**=R] = 11/24	Pr[R|a] = ¾ Pr[R|b] = ¼ Pr[R|c] = 0

Pr[a|P] = 0 Pr[b|P] = 0 Pr[c|P] = 1

Pr[a|Q] = 1/3 Pr[b|Q] = 2/3 Pr[c|Q] = 0

Pr[a|R] = 9/11 Pr[b|R] = 2/11 Pr[c|R] = 0

If the attacker sees ciphertext ***P*** then she would know the plaintext was ***c***

If the attacker sees ciphertext ***R*** then she would know ***a*** is the most likely plaintext

Not a good encryption mechanism!!

Perfect Secrecy

- Perfect secrecy achieved when

a posteriori probabilities = a priori probabilities

$$\Pr[x \mid y] = \Pr[x]$$

i.e the attacker learns nothing from the ciphertext

Observation on Perfect secrecy

Perfect Secrecy iff

$$\Pr[Y = y \mid X = x] = \Pr[Y = y]$$

Follows from
Baye's theorem

Perfect Indistinguishability

$$\forall x_1, x_2 \in P \qquad \Pr[Y = y \mid X = x_1] = \Pr[Y = y \mid X = x_2]$$

Perfect secrecy has nothing to do with plaintext distribution.
Thus a crypto-scheme will achieve perfect secrecy irrespective of the language used in the plaintext.

Example

Example 3.3 Let $\mathcal{P} = \{a, b\}$ with $\mathbf{Pr}[a] = 1/4, \mathbf{Pr}[b] = 3/4$. Let $\mathcal{K} = \{K_1, K_2, K_3\}$ with $\mathbf{Pr}[K_1] = 1/2, \mathbf{Pr}[K_2] = \mathbf{Pr}[K_3] = 1/4$. Let $\mathcal{C} = \{1, 2, 3, 4\}$, and suppose the encryption functions are defined to be $e_{K_1}(a) = 1, e_{K_1}(b) = 2; e_{K_2}(a) = 2, e_{K_2}(b) = 3$; and $e_{K_3}(a) = 3, e_{K_3}(b) = 4$. This cryptosystem can be represented by the following ***encryption matrix***:

	a	b
K_1	1	2
K_2	2	3
K_3	3	4

Solution

- Given:
- $P_r[a] = \frac{1}{4}, P_r[b] = \frac{3}{4}$
- $P_r[k_1] = \frac{1}{2}, P_r[k_2] = \frac{1}{4}, \mathrm{P_r[k_3]} = \frac{1}{4}$

- $P_r[\mathrm{Y = y}] = \sum_k P_r(k).P_r(d_k(y))$
- $P_r[Y = 1] = P_r[k_1].P_r[a] = \frac{1}{2} * \frac{1}{4} = \frac{1}{8}$
- $P_r[Y = 2] = P_r[k_1].P_r[b] + P_r[k_2].P_r[a] = \frac{1}{2} * \frac{3}{4} + \frac{1}{4} * \frac{1}{4} = \frac{3}{8} + \frac{1}{16} = \frac{7}{16}$

Contd…

	a	b
K_1	1	2
K_2	2	3
K_3	3	4

- $P_r[Y = 3] = P_r[k_2].P_r[b] + P_r[k_3].P_r[a] = \frac{1}{4} * \frac{3}{4} + \frac{1}{4} * \frac{1}{4} = \frac{3}{16} + \frac{1}{16} = \frac{1}{4}$
- $P_r[Y = 4] = P_r[k_3].P_r[b] = \frac{1}{4} * \frac{3}{4} = \frac{3}{16}$

Conditional Probability

The probability that y is obtained given x depends on the keys which provide such a mapping

$$\Pr[y \mid x] = \sum_{\{k \,:\, d_k(y)=x\}} \Pr[k]$$

	a	b
K_1	1	2
K_2	2	3
K_3	3	4

$$P_r[1|a] = P_r[k_1] = \frac{1}{2}$$

$$P_r[2|a] = P_r[k_2] = \frac{1}{4}$$

$$P_r[3|a] = P_r[k_3] = \frac{1}{4}$$

$$P_r[4|a] = 0$$

$$P_r[1|b] = 0$$

$$P_r[2|b] = P_r[k_1] = \frac{1}{2}$$

$$P_r[3|b] = P_r[k_2] = \frac{1}{4}$$

$$P_r[4|b] = P_r[k_3] = \frac{1}{4}$$

Posteriori Probbaility

$$\Pr[x \mid y] = \frac{\Pr[x] \times \Pr[y \mid x]}{\Pr[y]}$$

$$P_r[a|1] = \frac{P_r[a]P_r[1|a]}{\mathrm{P_r}[1]} = \frac{1}{4} * \frac{1}{2} \div \frac{1}{8} = 1$$

$$P_r[a|2] = \frac{P_r[a]P_r[2|a]}{\mathrm{P_r}[2]} = \frac{1}{4} * \frac{1}{4} \div \frac{7}{16} = \frac{1}{7}$$

Contd...

$$\begin{aligned}
\mathbf{Pr}[a|1] &= 1 & \mathbf{Pr}[b|1] &= 0 \\
\mathbf{Pr}[a|2] &= \frac{1}{7} & \mathbf{Pr}[b|2] &= \frac{6}{7} \\
\mathbf{Pr}[a|3] &= \frac{1}{4} & \mathbf{Pr}[b|3] &= \frac{3}{4} \\
\mathbf{Pr}[a|4] &= 0 & \mathbf{Pr}[b|4] &= 1.
\end{aligned}$$

Perfect secrecy for y=3

Definition 3.3: A cryptosystem has ***perfect secrecy*** if $\mathbf{Pr}[x|y] = \mathbf{Pr}[x]$ for all $x \in \mathcal{P}$, $y \in \mathcal{C}$. That is, the *a posteriori* probability that the plaintext is x, given that the ciphertext y is observed, is identical to the *a priori* probability that the plaintext is x.

$$\mathbf{Pr}[a|3] = \frac{1}{4} \qquad \mathbf{Pr}[b|3] = \frac{3}{4}$$

Shift Cipher

- Plaintext set : $\mathbb{P}$ = {0,1,2,3 …, 25}
- Ciphertext set : $\mathbb{C}$ = {0,1,2,3 …, 25}
- Keyspace : $\mathbb{K}$ = {0,1,2,3 …, 25}
- Encryption Rule : $e_K(x) = (x + K) \bmod 26$,
- Decryption Rule : $d_k(x) = (x - K) \bmod 26$

where $K \in \mathbb{K}$ and $x \in \mathbb{P}$

The Twist : the key changes after every encryption

Shift Cipher

$$\mathbf{Pr}[\mathbf{y} = y] = \sum_{K \in \mathbb{Z}_{26}} \mathbf{Pr}[\mathbf{K} = K]\mathbf{Pr}[\mathbf{x} = d_K(y)]$$

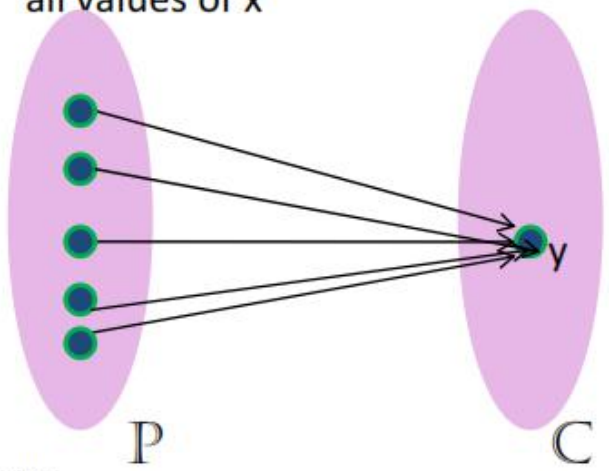
Keys chosen with uniform probability

$$= \sum_{K \in \mathbb{Z}_{26}} \frac{1}{26}\mathbf{Pr}[\mathbf{x} = y - K]$$

This is 1 because the sum is over all values of x

$$= \frac{1}{26} \sum_{K \in \mathbb{Z}_{26}} \mathbf{Pr}[\mathbf{x} = y - K].$$

$$= \frac{1}{26}$$



$$\mathbf{Pr}[y|x] = \mathbf{Pr}[\mathbf{K} = (y - x) \bmod 26]$$

$$= \frac{1}{26}$$

For every pair of y and x, there is exactly one key . Probability of that key is 1/26

Shift Cipher

$$\begin{aligned}\mathbf{Pr}[\mathbf{y}=y] &= \sum_{K\in\mathbb{Z}_{26}} \mathbf{Pr}[\mathbf{K}=K]\mathbf{Pr}[\mathbf{x}=d_K(y)] \\ &= \sum_{K\in\mathbb{Z}_{26}} \frac{1}{26}\mathbf{Pr}[\mathbf{x}=y-K] \\ &= \frac{1}{26}\sum_{K\in\mathbb{Z}_{26}} \mathbf{Pr}[\mathbf{x}=y-K]. \\ &= \frac{1}{26}\end{aligned}$$

$$\begin{aligned}\mathbf{Pr}[x|y] &= \frac{\mathbf{Pr}[x]\mathbf{Pr}[y|x]}{\mathbf{Pr}[y]} \\ &= \frac{\mathbf{Pr}[x]\frac{1}{26}}{\frac{1}{26}} \\ &= \mathbf{Pr}[x],\end{aligned}$$

$$\begin{aligned}\mathbf{Pr}[y|x] &= \mathbf{Pr}[\mathbf{K}=(y-x) \bmod 26] \\ &= \frac{1}{26}\end{aligned}$$

Shift Cipher

- The scheme becomes perfectly secure when only one character is encrypted i.e., M = C = {0, 1, . . . , 25} (in other words, a different key is chosen for each letter).
- The Shift Cipher is “unbreakable” provided that a new random key is used to encrypt every plaintext character.

Introduction

- Consider an experiment with a number of possible outcomes. Let x represent the outcome.
- Before the experiment is conducted the outcome is unknown, there is an uncertainty associated with it.
- Entropy is a measure of information content in x.
- Ex: x can take any of 8 specific values $b_2 b_1 b_0$
- If each of the 8 outcomes is equi-likely, then the probability is $\frac{1}{8}$.
- For example if partial infor is known if $b_2 = 1$ uncertainitv is reduced bv 50%.

Entropy

- We now want to look at what happens as more and more plaintexts are encrypted using the same key, and how likely a cryptanalyst will be able to carry out a successful ciphertext-only attack.
- Entropy can be thought of as a mathematical measure of information or uncertainty, and is computed as a function of a probability distribution.
- The entropy of X and is denoted by H(X)

Entropy

Let X be a discrete random variable with alphabet $\mathcal{X}$ and probability mass function $p(x)$

$$p(x) = Pr\{X = x\}, \quad x \in \mathcal{X}$$

The *entropy* of the variable X is defined by

$$H(X) = -\sum_{x \in \mathcal{X}} p(x) \log p(x)$$

The logarithm can be in any base, but normally base 2 is used. The unit of the entropy is then called *bits*. If we use base b in the logarithm, we denote the entropy by $H_b(X)$. We can easily convert between entropies in different bases

$$H_b(X) = \log_b a \cdot H_a(X)$$

By convention $0 \log 0 = 0$, since $y \log y \to 0$ as $y \to 0$.

The entropy is a measure of the information content of a random variable.

Contd...

The entropy is always non-negative

$$H(X) \geq 0$$

Proof: Since $0 \leq p(x) \leq 1$, $-\log p(x) \geq 0$

The entropy is always less than the logarithm of the alphabet size

$$H(X) \leq \log |\mathcal{X}|$$

with equality given a uniform distribution.

Probability Review

- *A random variable (event) is an experiment whose outcomes are mapped to real numbers.*
- *For our discussion we will deal with discrete-valued random variables.*
- ***Probability:*** *We denote* $p_X(x) = Pr(X = x)$.

  For a subset A, $\mathrm{p}(\mathrm{A}) = \sum_{x \in A} \mathrm{p}_X(\mathrm{x})$
- ***Joint Probability:*** *Sometimes we want to consider more than two events at the same time, in which we case we lump them together into a joint random variable, e.g.* $Z = (X,Y)$.

$$\mathrm{p}_{X,Y}((X,Y) = (x,y)) = \Pr((X = x),(Y = y))$$

- ***Independence:*** *We say that two events are independent if*

$$\mathrm{p}_{X,Y}((X,Y) = (x,y)) = \mathrm{p}_X(x)\mathrm{p}_Y(y)$$

Probability Review

- **Conditional Probability:** We will often ask questions about the probability of events Y given that we have observed X=x. In particular, we define the conditional probability of Y=y given X=x by

$$p_Y(y \mid x) = \frac{p_{XY}(x, y)}{p_X(x)}$$

- **Independence:** We immediately get $p_Y(y \mid x) = p_Y(y)$

- **Bayes's Theorem:** If $p_X(x)>0$ and $p_Y(y)>0$ then

$$p_X(x \mid y) = \frac{p_X(x)p_Y(y \mid x)}{p_Y(y)}$$

Entropy and Uncertainty

- We are concerned with how much uncertainty a random event has, but how do we define or measure uncertainty?
- We want our measure to have the following properties:

1. To each set of nonnegative numbers $p = p_1, p_2, \cdots, p_n$ with $p_1 + p_2 + \cdots + p_n = 1$, we define the uncertainty by $H(p)$.
2. $H(p)$ should be a continuous function: A slight change in p should not drastically change $H(p)$
3. $H\left(\frac{1}{n}, \cdots, \frac{1}{n}\right) \leq H\left(\frac{1}{n+1}, \cdots, \frac{1}{n+1}\right)$ for all n>0. Uncertainty increases when there are more outcomes.
4. If 0<q<1, then

$$H\left(p_1, \cdots, qp_j, (1-q)p_j, \cdots, p_n\right) = H\left(p_1, \cdots, p_n\right) + p_j H\left(q, 1-q\right)$$

Entropy

- We define the entropy of a random variable by

$$H(X) = -\sum_{x \in X} p(x) \log_2 p(x)$$

- **Example:** Consider a fair coin toss. There are two outcomes, with probability ½ each. The entropy is

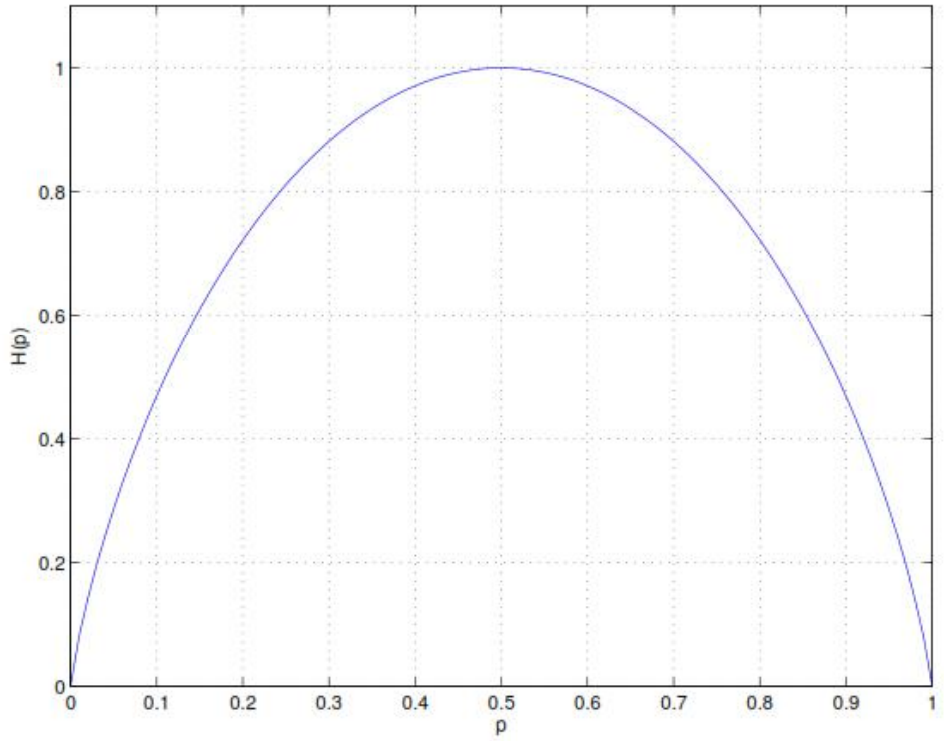
$$-\left(\frac{1}{2}\log_2\frac{1}{2} + \frac{1}{2}\log_2\frac{1}{2}\right) = 1 \text{bit}$$

- **Example:** Consider a non-fair coin toss X with probability p of getting heads and 1-p of getting tails. The entropy is

$$H(X) = -p\log_2 p - (1-p)\log_2(1-p)$$

The entropy is maximum when p= ½.

Binary Entropy



Entropy for a binary variable with symbol probabilities p and $1 - p$.
$H(p) = -p \cdot \log p - (1 - p) \cdot \log(1 - p)$

Entropy Property

- Property 1:
- The uncertainty is maximum when the outcomes are equally likely. The uniform distribution maximizes the entropy; the uniform distribution contains the largest amount of uncertainty.

Entropy

- Suppose our random variable X represents the toss of a coin. As mentioned earlier, the associated probability distribution is Pr[heads] = Pr[tails] = 1/2. It would seem reasonable to say that the information, or entropy, of a coin toss is one bit, since we could encode heads by 1 and tails by 0

Example

Example: Consider a fair die with *pmf* $p(1) = p(2) = \ldots = p(6) = 1/6$. Its entropy is

$$H(x) = -6 \cdot \frac{1}{6} \log \frac{1}{6} = \log 6$$

Example

- suppose we have a random variable X that takes on three possible values x1, x2, x3with probabilities 1/2, 1/4, 1/4

tively. Suppose we encode the three possible outcomes as follows: x_1 is encoded as 0, x_2 is encoded as 10, and x_3 is encoded as 11. Then the (weighted) average number of bits in this encoding of **X** is

$$\frac{1}{2} \times 1 + \frac{1}{4} \times 2 + \frac{1}{4} \times 2 = \frac{3}{2}.$$

Entropy Definition

Definition 3.4: Suppose $\mathbf{X}$ is a discrete random variable that takes on values from a finite set X. Then, the ***entropy*** of the random variable $\mathbf{X}$ is defined to be the quantity

$$H(\mathbf{X}) = -\sum_{x \in X} \mathbf{Pr}[x] \log_2 \mathbf{Pr}[x].$$

Example

Example 3.3 Let $\mathcal{P} = \{a, b\}$ with $\mathbf{Pr}[a] = 1/4, \mathbf{Pr}[b] = 3/4$. Let $\mathcal{K} = \{K_1, K_2, K_3\}$ with $\mathbf{Pr}[K_1] = 1/2, \mathbf{Pr}[K_2] = \mathbf{Pr}[K_3] = 1/4$. Let $\mathcal{C} = \{1, 2, 3, 4\}$, and suppose the encryption functions are defined to be $e_{K_1}(a) = 1, e_{K_1}(b) = 2$; $e_{K_2}(a) = 2, e_{K_2}(b) = 3$; and $e_{K_3}(a) = 3, e_{K_3}(b) = 4$. This cryptosystem can be represented by the following ***encryption matrix***:

	a	b
K_1	1	2
K_2	2	3
K_3	3	4

Example

$$
\begin{aligned}
H(\mathbf{P}) &= -\frac{1}{4}\log_2\frac{1}{4} - \frac{3}{4}\log_2\frac{3}{4} \\
&= -\frac{1}{4}(-2) - \frac{3}{4}(\log_2 3 - 2) \\
&= 2 - \frac{3}{4}(\log_2 3) \\
&\approx 0.81.
\end{aligned}
$$

$$H(\mathbf{K}) = 1.5 \text{ and } H(\mathbf{C}) \approx 1.85.$$

Property

Definition 3.5: A real-valued function f is a ***concave function*** on an interval I if

$$f\left(\frac{x+y}{2}\right) \geq \frac{f(x)+f(y)}{2}$$

for all $x, y \in I$. f is a ***strictly concave function*** on an interval I if

$$f\left(\frac{x+y}{2}\right) > \frac{f(x)+f(y)}{2}$$

for all $x, y \in I, x \neq y$.

Theorem

THEOREM 3.6 *Suppose* $\mathbf{X}$ *is a random variable having a probability distribution that takes on the values* $p_1, p_2, \ldots, p_n$, *where* $p_i > 0, 1 \leq i \leq n$. *Then* $H(\mathbf{X}) \leq \log_2 n$, *with equality if and only if* $p_i = 1/n, 1 \leq i \leq n$.

PROOF Applying Jensen's inequality, we have the following:

$$\begin{aligned} H(\mathbf{X}) &= -\sum_{i=1}^{n} p_i \log_2 p_i \\ &= \sum_{i=1}^{n} p_i \log_2 \frac{1}{p_i} \\ &\leq \log_2 \sum_{i=1}^{n} \left(p_i \times \frac{1}{p_i} \right) \\ &= \log_2 n. \end{aligned}$$

Further, equality occurs if and only if $p_i = 1/n, 1 \leq i \leq n$.

Joint Entropy

1.1 Joint Entropy

Definition: For two random variables X and Y, $x \in \mathcal{X}, y \in \mathcal{Y}$, *joint entropy* is defined as

$$H(X,Y) = -\sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x,y) \log p(x,y)$$

where $p(x,y) = Pr[X = x, Y = y]$ is the joint *pmf* of X and Y.

Conditional Entropy

1.2 Conditional Entropy

Definition: The *conditional entropy* of a random variable Y given $X = x$ is

$$H(Y|X = x) = -\sum_{y \in \mathcal{Y}} p(y|x) \log p(y|x)$$

When a particular value of x is not given, we must average over all possible values of X:

$$\begin{aligned} H(Y|X) &= -\sum_{x \in \mathcal{X}} p(x) \left(\sum_{y \in \mathcal{Y}} p(y|x) \log p(y|x) \right) \\ &= -\sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x, y) \log p(y|x) \end{aligned}$$

The conditional entropy of X given Y is

$$H(X|Y) = -\sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x, y) \log p(x|y)$$

In general, $H(X|Y) \neq H(Y|X)$.

Entropy

1.3 Chain Rule for Entropy

The *Chain Rule for Entropy* states that the entropy of two random variables is the entropy of one plus the conditional entropy of the other:

$$H(X,Y) = H(X) + H(Y|X) \tag{1}$$

Proof:

$$\begin{aligned} H(X,Y) &= -\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x,y)\log\Big(p(x)p(y|x)\Big) \\ &= -\sum_{x\in\mathcal{X}}\left(\sum_{y\in\mathcal{Y}} p(x,y)\right)\log p(x) - \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x,y)\log p(y|x) \\ &= -\sum_{x\in\mathcal{X}} p(x)\log p(x) - \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x,y)\log p(y|x) \\ &= H(X) + H(Y|X) \end{aligned}$$

Similarly, it can also be shown that

$$H(X,Y) = H(Y) + H(X|Y) \tag{2}$$

Example

Example: Consider the random variables $X \in \{0,1\}, Y \in \{0,1\}$, representing two coin tosses. Their joint distribution is shown in Table 1.

Y\X	0	1
0	1/2	1/4
1	1/8	1/8

X (0) = ½+1/8 =5/8
X(1) = ¼ +1/8 = 3/8

Table 1: Joint distribution of two coin tosses.

The joint entropy of X and Y is

$$H(X,Y) = -\frac{1}{2}\log\frac{1}{2} - \frac{1}{4}\log\frac{1}{4} - \frac{2}{8}\log\frac{1}{8} = 1.75$$

Note that if all probabilities were equal, we would have

$H(X,Y) = \log 4 = 2$ bits, which is the maximum entropy.

Chain Rule for Entropy: $H(X,Y) = H(Y) + H(X|Y)$.

$$H(X|Y) = 1.75 - 0.8113 \approx 0.9387$$

The individual entropies are

$$H(X) = -\frac{5}{8}\log\frac{5}{8} - \frac{3}{8}\log\frac{3}{8} \approx 0.9544$$

$$H(Y) = -\frac{3}{4}\log\frac{3}{4} - \frac{1}{4}\log\frac{1}{4} \approx 0.8113$$

Property 2

- If X and Y are independent variables, then
- $H(X,Y) = H(X) + H(Y)$

Property 3

- The joint entropy of a set of variables is less than or equal to the sum of the individual entropies of the variables in the set.
- $H(X,Y) \leq \mathrm{H(X)} + \mathrm{H(Y)}$

Conditional Entropy

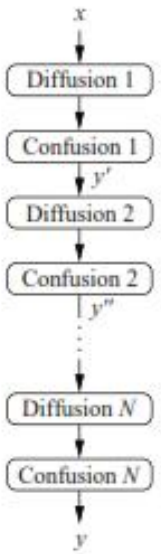
- $H(X,Y) = H(X) + H(Y|X)$

Product Cryptosystem

1. **Confusion** is an encryption operation where the relationship between key and ciphertext is obscured. Today, a common element for achieving confusion is substitution, which is found in both DES and AES.
2. **Diffusion** is an encryption operation where the influence of one plaintext symbol is spread over many ciphertext symbols with the goal of hiding statistical properties of the plaintext. A simple diffusion element is the bit permutation, which is used frequently within DES. AES uses the more advanced Mixcolumn operation.

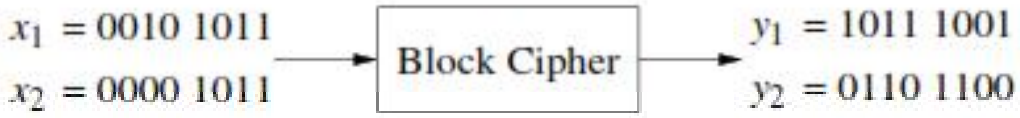
Product Ciphers

- The idea of concatenating several encryption operation was also proposed by Shannon. Such ciphers are known as product ciphers.



Example

Example 3.1. Let's assume a small block cipher with a block length of 8 bits. Encryption of two plaintexts x_1 and x_2, which differ only by one bit, should roughly result in something as shown in Fig. 3.2.



Perfect Secrecy

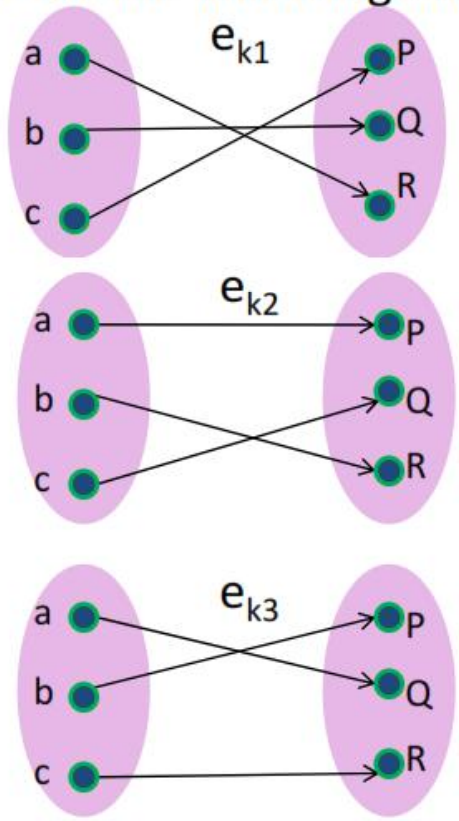
Definition 3.3: A cryptosystem has ***perfect secrecy*** if $\mathbf{Pr}[x|y] = \mathbf{Pr}[x]$ for all $x \in \mathcal{P}$, $y \in \mathcal{C}$. That is, the *a posteriori* probability that the plaintext is x, given that the ciphertext y is observed, is identical to the *a priori* probability that the plaintext is x.

Example 6

- Find the a posteriori probabilities for the following scheme
- Verify that it is perfectly secret.

plaintext
Pr[**X**=a] = 1/2
Pr[**X**=b] = 1/3
Pr[**X**=c] = 1/6

keyspace
Pr[**K**=k_1] = 1/3
Pr[**K**=k_2] = 1/3
Pr[K=k_3] = 1/3



Solution

- Given

plaintext
Pr[**X**=a] = 1/2
Pr[**X**=b] = 1/3
Pr[**X**=c] = 1/6

keyspace
Pr[**K**=k_1] = 1/3
Pr[**K**=k_2] = 1/3
Pr[K=k_3] = 1/3

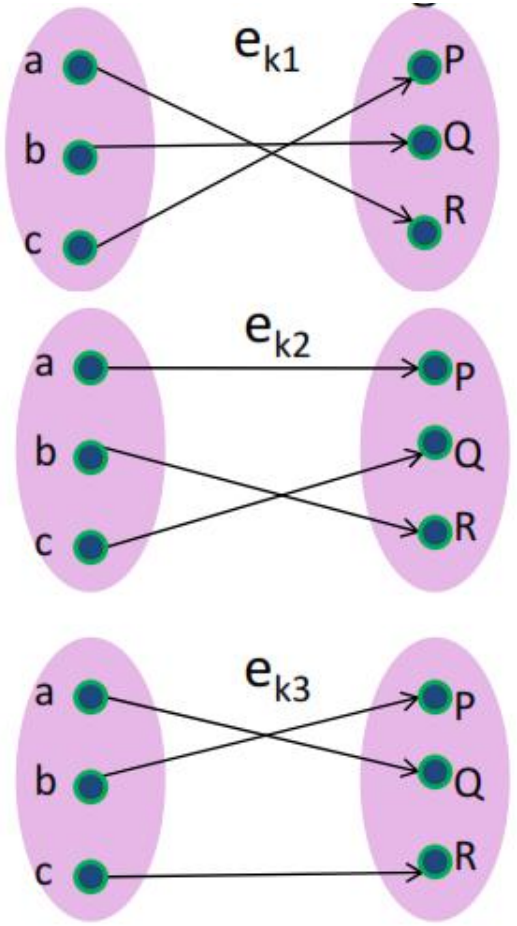
Cipher Text Distribution

$$P_r[Y = y] = \sum_k P_r(k).P_r(d_k(y))$$

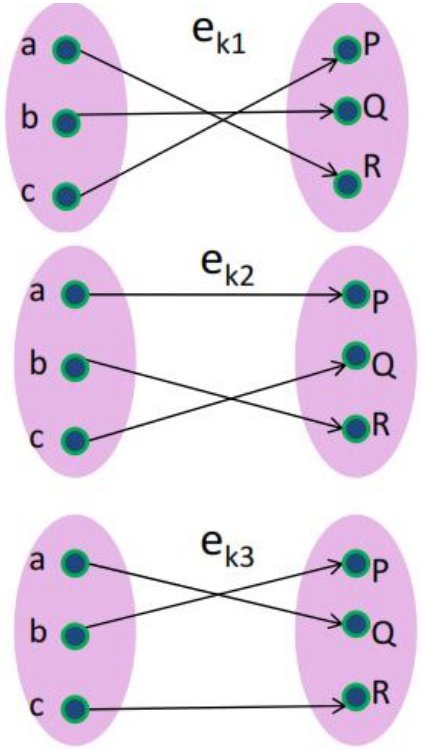
$$P_r[Y = P] = P_r(k_1).P_r(c) + P_r(k_2).P_r(a) + P_r(k_3).P_r(b) = \frac{1}{3} * \frac{1}{6} + \frac{1}{3} * \frac{1}{2} + \frac{1}{3} * \frac{1}{3} = \frac{1}{18} + \frac{1}{6} + \frac{1}{9} = \frac{1+3+2}{18} = \frac{1}{3}$$

$$P_r[Y = Q] = P_r(k_1).P_r(b) + P_r(k_2).P_r(c) + P_r(k_3).P_r(a) = \frac{1}{3} * \frac{1}{3} + \frac{1}{3} * \frac{1}{6} + \frac{1}{3} * \frac{1}{2} = \frac{1}{9} + \frac{1}{18} + \frac{1}{6} = \frac{2+1+3}{18} = \frac{1}{3}$$

$$P_r[Y = R] = P_r(k_1).P_r(a) + P_r(k_2).P_r(b) + P_r(k_3).P_r(c) = \frac{1}{3} * \frac{1}{2} + \frac{1}{3} * \frac{1}{3} + \frac{1}{3} * \frac{1}{6} = \frac{1}{6} + \frac{1}{9} + \frac{1}{18} = \frac{3+2+1}{18} = \frac{1}{3}$$



P(Y|X)



The probability that y is obtained given x depends on the keys which provide such a mapping

$$\Pr[y \mid x] = \sum_{\{k\,:\,d_k(y)=x\}} \Pr[k]$$

$$P(P|a) = P_r(k_2) = \frac{1}{3}\ P(P|b) = P_r(k_3) = \frac{1}{3}\ P(P|c) = P_r(k_1) = \frac{1}{3}$$
$$P(Q|a) = P_r(k_3) = \frac{1}{3}\ P(Q|b) = P_r(k_1) = \frac{1}{3}\ P(Q|c) = P_r(k_2) = \frac{1}{3}$$
$$P(R|a) = P_r(k_1) = \frac{1}{3}\ P(R|b) = P_r(k_2) = \frac{1}{3}\ P(R|c) = P_r(k_3) = \frac{1}{3}$$

Computing Posteriori Probability

$$\Pr[x \mid y] = \frac{\Pr[x] \times \Pr[y \mid x]}{\Pr[y]}$$

$$P_r[a|P] = P_r[a] * \frac{P_r[P|a]}{P_r[P]} = \frac{1}{2} * \frac{\frac{1}{3}}{\frac{1}{3}} = \frac{1}{2} = \mathrm{P_r[a]} = \frac{1}{2}$$

$$P_r[a|Q] = P_r[a] * \frac{P_r[Q|a]}{P_r[Q]} = \frac{1}{2} * \frac{\frac{1}{3}}{\frac{1}{3}} = \frac{1}{2} = \mathrm{P_r[a]} = \frac{1}{2}$$

$$P_r[a|R] = P_r[a] * \frac{P_r[R|a]}{P_r[R]} = \frac{1}{2} * \frac{\frac{1}{3}}{\frac{1}{3}} = \frac{1}{2} = \mathrm{P_r[a]} = \frac{1}{2}$$

Contd...

$$P_r[b|P] = P_r[b] * \frac{P_r[P|b]}{P_r[P]} = \frac{1}{3} * \frac{\frac{1}{3}}{\frac{1}{3}} = \frac{1}{3} = \mathrm{P_r[b]} = \frac{1}{3}$$

$$P_r[b|Q] = P_r[b] * \frac{P_r[Q|b]}{P_r[Q]} = \frac{1}{3} * \frac{\frac{1}{3}}{\frac{1}{3}} = \frac{1}{3} = \mathrm{P_r[b]} = \frac{1}{3}$$

$$P_r[b|R] = P_r[b] * \frac{P_r[R|b]}{P_r[R]} = \frac{1}{3} * \frac{\frac{1}{3}}{\frac{1}{3}} = \frac{1}{3} = \mathrm{P_r[b]} = \frac{1}{3}$$

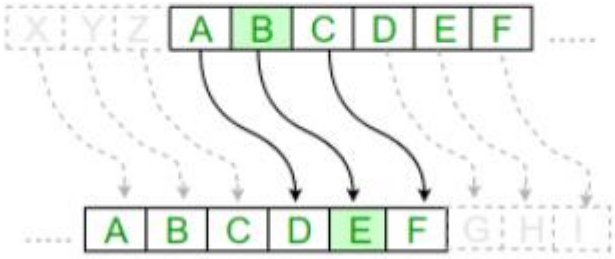
Caesar Cipher

$$E_n(x) = (x + n) mod\ 26$$

(Encryption Phase with shift n)

$$D_n(x) = (x - n) mod\ 26$$

(Decryption Phase with shift n)



Algorithm

Algorithm for Caesar Cipher:

Input:

1. A String of lower case letters, called Text.
2. An Integer between 0-25 denoting the required shift.

Procedure:

- Traverse the given text one character at a time .
- For each character, transform the given character as per the rule, depending on whether we're encrypting or decrypting the text.
- Return the new string generated.

Algorithm

1. *A. Encryption Procedure for Caesar Cipher:*

    1. Read the plain text message
    2. Read the key value (displacement)
    3. To generate the cipher text , replace each letter of plaintext by a letter at the position specified by the key value down the alphabetical stream.
    4. Display the cipher text.

Decryption Procedure for Caesar Cipher:

1. Use the cipher text as input
2. Use the same key value as displacement
3. To retrieve the plaintext text from cipher text, replace a letter of cipher text by the letter at the position specified by the key value in the reverse alphabetical stream.
4. Display the plain text.

```java
class Main {
  // Encrypts text using a shift od s
  public static StringBuffer encrypt(String text, int s) {
    StringBuffer result = new StringBuffer();

    for (int i = 0; i < text.length(); i++) {
      if (Character.isUpperCase(text.charAt(i))) {
        char ch = (char) (((int) text.charAt(i) + s - 65) % 26 + 65);
        result.append(ch);
      } else {
        char ch = (char) (((int) text.charAt(i) + s - 97) % 26 + 97);
        result.append(ch);
      }
    }
    return result;
  }

  // Driver code
  public static void main(String[] args) {
    String text = "ATTACKATONCE";
    int s = 4;
    System.out.println("Text  : " + text);
    System.out.println("Shift : " + s);
    System.out.println("Cipher: " + encrypt(text, s));
  }
}
```

Sample output

```
$ java -classpath .:/run_dir/junit-4.12.jar:target/dependency/* Main
Text   : ATTACKATONCE
Shift  : 4
Cipher: EXXEGOEXSRGI
$
```

Playfair

The Playfair Cipher Encryption Algorithm:

The Algorithm consistes of 2 steps:

1. **Generate the key Square(5×5):**
   - The key square is a 5×5 grid of alphabets that acts as the key for encrypting the plaintext. Each of the 25 alphabets must be unique and one letter of the alphabet (usually J) is omitted from the table (as the table can hold only 25 alphabets). If the plaintext contains J, then it is replaced by I.
   - The initial alphabets in the key square are the unique alphabets of the key in the order in which they appear followed by the remaining letters of the alphabet in order.

```
The key is "monarchy"
Thus the initial entires are
'm', 'o', 'n', 'a', 'r', 'c', 'h', 'y'
followed by remaining characters of
a-z(except 'j') in that order.
```

M	O	N	A	R
C	H	Y	B	D
E	F	G	I	K
L	P	Q	S	T
U	V	W	X	Z

Rules

- **If both the letters are in the same column**: Take the letter below each one (going back to the top if at the bottom).

  For example:

```
Diagraph: "me"
Encrypted Text: cl
Encryption:
  m -> c
  e -> l
```

M	O	N	A	R
C	H	Y	B	D
E	F	G	I	K
L	P	Q	S	T
U	V	W	X	Z

Rules

- If both the letters are in the same row: Take the letter to the right of each one (going back to the leftmost if at the rightmost position).For example:

```
Diagraph: "st"
Encrypted Text: tl
Encryption:
  s -> t
  t -> l
```

M	O	N	A	R
C	H	Y	B	D
E	F	G	I	K
L	P	Q	S	T
U	V	W	X	Z

Rules

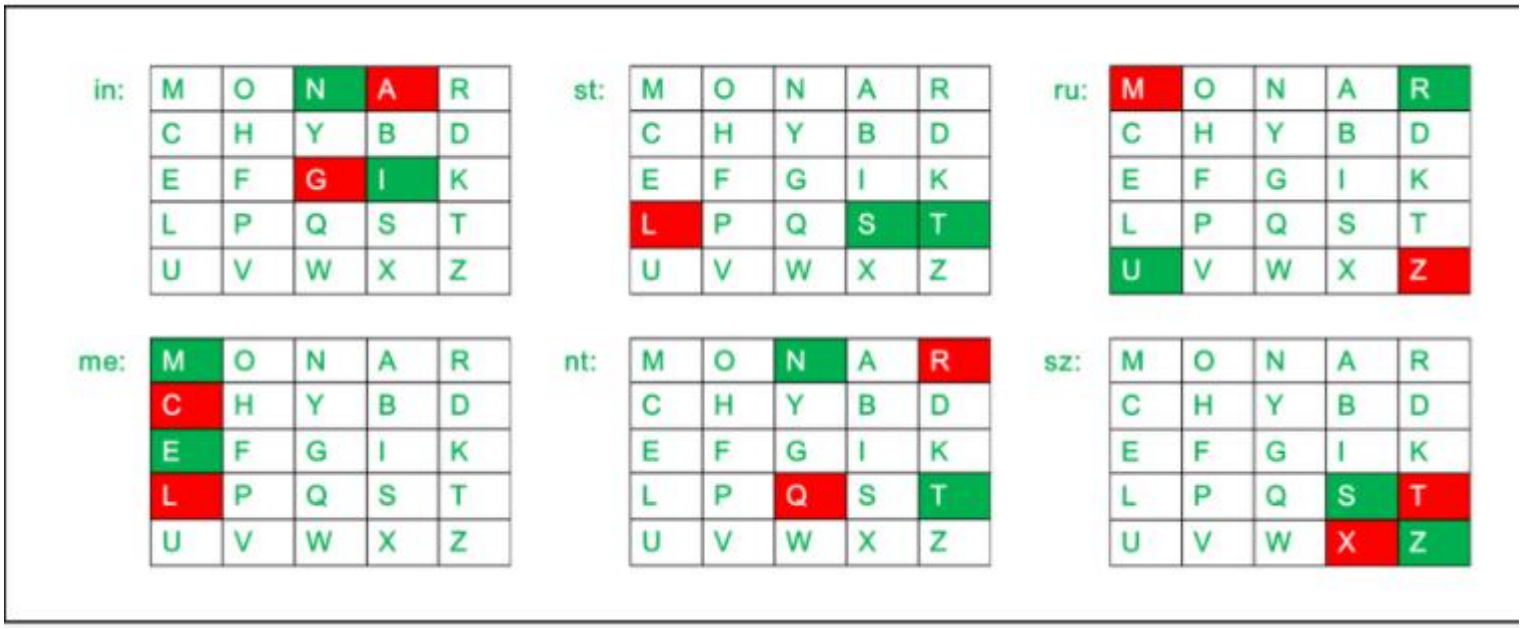
If neither of the above rules is true: Form a rectangle with the two letters and take the letters on the horizontal opposite corner of the rectangle.

For example:

```
Diagraph: "nt"
Encrypted Text: rq
Encryption:
  n -> r
  t -> q
```

M	O	N	A	R
C	H	Y	B	D
E	F	G	I	K
L	P	Q	S	T
U	V	W	X	Z

Cipher text


in:
M O N A R
C H Y B D
E F G I K
L P Q S T
U V W X Z
st:
M O N A R
C H Y B D
E F G I K
L P Q S T
U V W X Z
ru:
M O N A R
C H Y B D
E F G I K
L P Q S T
U V W X Z
me:
M O N A R
C H Y B D
E F G I K
L P Q S T
U V W X Z
nt:
M O N A R
C H Y B D
E F G I K
L P Q S T
U V W X Z
sz:
M O N A R
C H Y B D
E F G I K
L P Q S T
U V W X Z

Playfair

1. *B. Encryption Procedure for Playfair Cipher:*
2. Read the plain text message
3. Read the key value (a string without any repetition letters)
4. Construct a 5 X 5 matrix and fill in the key text in row wise manner.
5. Fill in the remaining cells of the matrix with the rest of the alphabets sans the letters of the key.
6. Split the plain text into two letter words without repetition.
7. If a pair has a repeated letter, insert filler like 'X'
8. If both letters fall in the same row, replace each with letter to right (wrapping back to start from end)
9. If both letters fall in the same column, replace each with the letter below it (wrapping to top from bottom)
10. Otherwise each letter is replaced by the letter in the same row and in the column of the other letter of the pair
11. Display the cipher text.

Decryption

Decryption Procedure for Playfair Cipher:

1. Use the cipher text as input
2. Use the same key value
3. To retrieve the plaintext text from cipher text, split the plain text into two letter words without repetition.
4. Repeat the steps 7 to 9 of encryption to generate the plaintext
5. Display the plain text.

Playfair

```
public class PlayfairCipher {
    private static char[][] charTable;
    private static Point[] positions;

    public static void main(String[] args) {
        Scanner sc = new Scanner(System.in);

        String key = prompt("Enter an encryption key (min length 6): ", sc, 6);
        String txt = prompt("Enter the message: ", sc, 1);
        String jti = prompt("Replace J with I? y/n: ", sc, 1);

        boolean changeJtoI = jti.equalsIgnoreCase("y");

        createTable(key, changeJtoI);

        String enc = encode(prepareText(txt, changeJtoI));

        System.out.printf("%nEncoded message: %n%s%n", enc);
        System.out.printf("%nDecoded message: %n%s%n", decode(enc));
    }
```

Encode / Decode

```
private static String encode(String s) {
    StringBuilder sb = new StringBuilder(s);

    for (int i = 0; i < sb.length(); i += 2) {

        if (i == sb.length() - 1)
            sb.append(sb.length() % 2 == 1 ? 'X' : "");

        else if (sb.charAt(i) == sb.charAt(i + 1))
            sb.insert(i + 1, 'X');
    }
    return codec(sb, 1);
}

private static String decode(String s) {
    return codec(new StringBuilder(s), 4);
}
```

```java
private static String codec(StringBuilder text, int direction) {
    int len = text.length();
    for (int i = 0; i < len; i += 2) {
        char a = text.charAt(i);
        char b = text.charAt(i + 1);

        int row1 = positions[a - 'A'].y;
        int row2 = positions[b - 'A'].y;
        int col1 = positions[a - 'A'].x;
        int col2 = positions[b - 'A'].x;

        if (row1 == row2) {
            col1 = (col1 + direction) % 5;
            col2 = (col2 + direction) % 5;

        } else if (col1 == col2) {
            row1 = (row1 + direction) % 5;
            row2 = (row2 + direction) % 5;

        } else {
            int tmp = col1;
            col1 = col2;
            col2 = tmp;
        }

        text.setCharAt(i, charTable[row1][col1]);
        text.setCharAt(i + 1, charTable[row2][col2]);
    }
    return text.toString();
}
```