

المقدمة عن مادة نظرية المعلومات

في ذهننا تصور عام مبهم إلى حد ما عن ما تمثله المعلومات، نعتقد عموماً أن المعلومات هي شيء من المعرفة أو الأفكار التي يكتسبها الإنسان أو يحملها في ذهنه أو موجودة في كتاب أو موجودة في صورة، لذلك يقال الصورة تساوي أو تمثل ألف كلمة بمعنى أن هناك كم من الأفكار أو من الأشياء المحمولة في الكتب، أو في أذهان البشر، أو في الحاسوب، أو في الصور...

لكن كيف نكم هذه المعلومات؟ أو كيف نُقدّر هذه المعلومات المحمولة في جهاز معين أو في الصورة أو في الكتاب؟ هل هو عدد الصفحات؟ بالتأكيد لا.

لأنه قد يكون الكتاب هذا بسيط جداً لديه مثلاً فكرة واحدة ويعيد تكرارها دائماً حتى لو مثلاً أعاد تكرارها بطرق مختلفة. في هذه الحالة عندنا فكرة وعندنا عدد الطرق التي تم التعبير بها عن تلك الفكرة، إذن الفكرة موجودة ومعلومات أخرى إضافية أيضاً موجودة لكن تبقى محدودة.

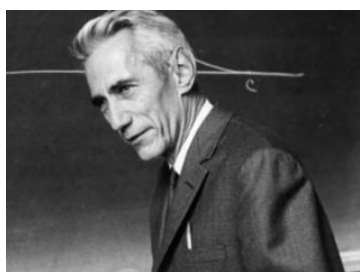


مثلاً لو تم كتابة 11111111 هكذا 10 مرات أو تم كتابة (1 x 10) هل هذه الكتابة المختصرة تحمل معلومات أقل من كتابة 11111111 هكذا 10 أو 20 أو 50 مرة؟ أكيد لا

تم التعبير عن نفس المعلومات ولكن بطريقة مختصرة فكونها طريقة مختصرة لا يغير من كمية المعلومة الموجودة فيها.

ماهو سبب أهمية ذلك مؤخراً؟

وذلك لأنه أصبح لدينا نقل للمعلومات.



في أواخر الأربعينيات، في 1948 نشر عالم رياضيات كان يعمل في نقل المعلومات في أحد المختبرات لشركة بيل الأمريكية المعروفة وهو كلود شانون بحثاً مهم جداً وضع فيه طريقة رياضية لتقدير كمية المعلومات الموجودة في أي قطعة معينة وعملية تشفيرها من أجل نقلها من مكان إلى مكان.

كانت الفكرة: إذا كانت لديك معلومات واردة نقلها إلى المستلم، فيجب أن يتم التأكد من أن هذه المعلومات تم تكميمها بطريقة صحيحة ثم تم نقلها بطريقة صحيحة ولم يكن هناك ضياع بالمعلومات بسبب الضجيج. ثم عندما وصلت المعلومات إلى المستلم هو تعرف على كمية المعلومات بطريقة كاملة وصحيحة.

لذلك كانت أهم مشكلة تعاني منها عملية نقل المعلومات

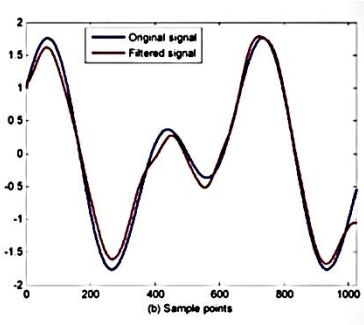
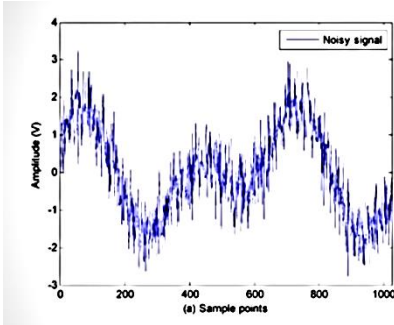
1- كيف نجعل هذه المعلومة مكتمة بشكل مدقق ونعرف كم لدينا من المعلومات، وكيف سيتم إرسالها.

رابط المحاضرة 1 على القناة: <https://youtu.be/VbPbjaNWiYs>

2- كيف يتم التخلص من الضجيج في عملية النقل.

لذلك مثلا عندما يتم التكلم بالهاتف او نقل الصورة او مشاهدة التلفزيون ... أحيانا تأتي الصورة مشوشة او تنقطع الإشارة او ينقطع البث... الخ.

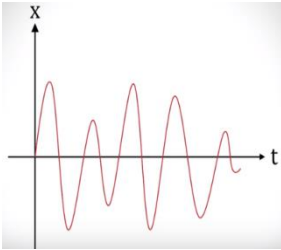
السبب: لأن هناك ضجيج والمعلومة لم يتم تشفيرها بشكل سليم ولم يتم نقلها بطريقة سليمة.



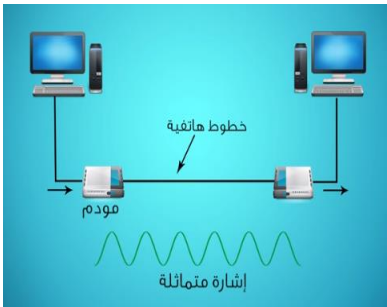
هذه المسألة (نقل المعلومات) كانت عن

طريق ما يسمى بنقل المعلومات بالتماثل (analog). معنى ذلك:

لو كان لدينا جملة نريد إرسالها إلى شخص ما، فانه سيتم إرسالها كلمة كلمة بالتعاقب فمثلا يتم إرسال الجملة: "انا اريد منك ان نرسل لي كذا ..."



كل كلمة ترسل بطريقة تماثلية (analog) أي تمثل كل كلمة بجهد كهربائي معين إذ أن لكل كلمة لها جهد محدد مختلف عن الكلمة الأخرى موجود ضمن قاموس معين، فاذا وصلت الجهود الكهربائية المتتالية فيقول المستلم مثلا إن الجهد الأول يمثل في سجلي كلمة "انا"، والجهد الثاني يمثل كذا في سجلي... الخ. أي بعملية تماثلية يعيد الإشارة أو المعلومة المستلمة.



وقد تحدث تأثيرات في الطريق أو أخطاء في الإرسال مثلا قد يرسل جهد اقل من المفترض فستختلف الكلمة لدى المستلم فتغير المعنى لان الكلمة مشوشة.

قال شانون: يجب التخلص من العملية التماثلية وهو ادخل التشفير الرقمي اذ قال: "نمثل المعلومات باقل كمية معلومة يمكن تمثيلها وهذه الكمية نسميها (Binary Digit Bit)".

ما هو اقل عدد من bits (كمية معلومة) يمكن أن يتم إرسالها؟ قال : هي 0 أو 1 . لماذا؟

إذ أن اقل معلومة يمكن إرسالها أن نقول نعم أو لا أو صحيح أو خطأ.

رابط المحاضرة 1 على القناة: <https://youtu.be/VbPbjaNWiYs>

فقال : يجب إرسال كل المعلومات مشفرة بالأصفر و الواحد، كل كلمة تتكون من مجموعة الأصفار والأحاد بحيث نضبط المعلومات بشكل دقيق.

The alphabet in binary

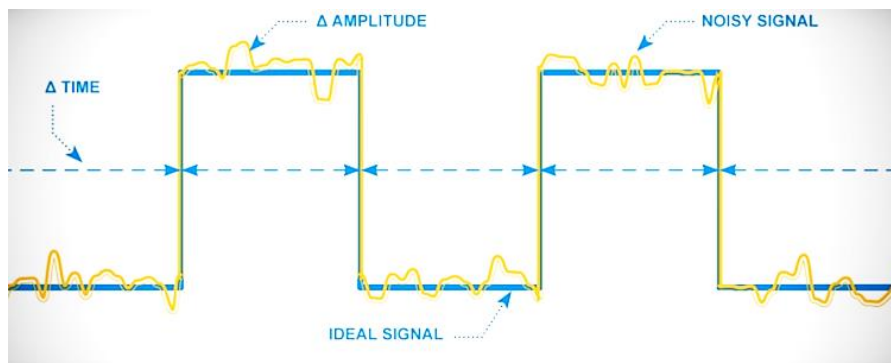
Binary	Alphabet
01100001	a
01100010	b
01100011	c
01100100	d
01100101	e
01100110	f
01100111	g
01101000	h
01101001	i

وقال: ان جهد الصفر جهد منخفض والواحد جهد مرتفع.



إذا أرسلنا جهد اقل او اكثر من الجهد العادي لكلمة بالطريقة التماثلية (analog)، فسيحتار المستلم لان الجهد لا يوجد في الجدول او في قاموسه.

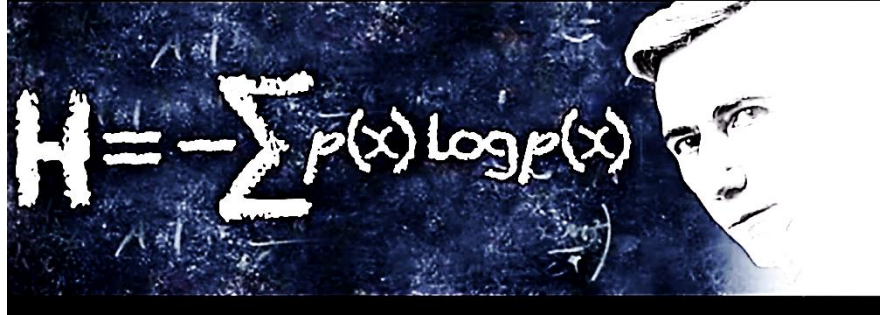
بينما بطريقة شانون : فان الجهد المنخفض مهما كان مقدار انخفاضه معناه صفر، والجهد المرتفع مهما كان ارتفاعه (أي صار عليه ضجيج في الطريق) يعني واحد. هذه العملية تسمح لنا لنقل المعلومات بشكل أدق.



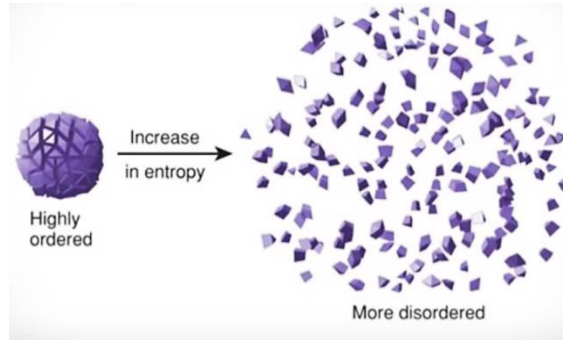
ثم اتى شانون بعملية حساب كمية المعلومة الموجودة في قطعة معينة .

رابط المحاضرة 1 على القناة: <https://youtu.be/VbPbjaNWiYs>

اذ جاء بمعادلته وهي اهم معادلات القرن العشرين هي معادلة شانون انتروبي تحسب لنا الانتروبي وهي تشبه الانتروبي في الفيزياء اذ تمثل كمية اللانتظام (عدم الانتظام).



مثلا يوجد كم من المادة لو تم تشكيلها بشكل مكعب ذلك يعني هناك انتظام هناك توزيع للجزيئات منتظم فسيكون الانتروبي لها قليل بكثير من عند ترك الجزيئات تتبعثر أي زيادة الا انتظام اي زاد الانتروبي.



فقال شانون: يجب أن يتم حساب كمية المعلومات الموجودة في قطعة معينة مثلا جملة او نص او صورة او فيديو بهذه الطريقة (حسب القانون اعلاه).

حيث تمثل قيمة x التي تظهر في المعادلة هي الحرف الذي يراد ارساله.

مثال 1: اذا كانت لدينا معلومة نريد نقلها وهي: نقل معلومة رمي قطعة نقود اذا سقطت على اي الوجهين؟

فقطعة النقود ممكن أن تسقط على الوجه الأول او الثاني، وان الاحتمال الوجه الاول يساوي احتمال الوجه الثاني. اي الجواب متساوي اي الاحتمال نصف (0.5) لاول و (0.5) للثاني.

وان قيمة X هي 0 او واحد (او + او -) وان $p(x)$ هي احتمال حصول الاول واحتمال حصول الثاني وهو متساوي للحالتين ونطبق حاصل ضرب الاحتمالية في لو غارتم الاحتمالية لكلا الحالتين، ستجد أن كم المعلومات المطلوب لنقل نتيجة هذا الرمية هي 1 اي bit واحد. ممكن ان يصل المستلم: "0" يعرف ان قطعة النقود سقطت على الوجه الأول واذا أرسلت: "1" يعرف أنها سقطت على الوجه الثاني.

رابط المحاضرة 1 على القناة: <https://youtu.be/VbPbjaNWiYs>

مثال 2: لو كان لدينا كلمة تحتوي مثلاً 8 حروف. في هذه الحالة فإن x تمثل كل حرف في الكلمة واحتمال ورود هذا الحرف في الكلمة هو مثلاً في العربية 28 حرف فاحتمالية الحرف الواحد مبدئياً هو 1 من 28 حرف. وحسب المعادلة أعلاه سنحتاج الى 5 bit تقريباً لنقل هذه المعلومة.

في حالة ان احتمال ظهور الحرف ليست متساوية، فيجب ان تؤخذ احتمالية كل الحروف بحيث يؤخذ كل حرف x وتحسب احتمال وروده (p) في أي كلمة في أي لغة ويتم ادخله في الوغارتم حسب المعادلة وتحسب الكمية كاملة للحصول على bits التي نحتاجها لنقل المعلومات.

إذا أصبح بالإمكان نقل كلمات بطريقة مضغوطة وبشكل متوالي وبطريقة سليمة وان التشويش الذي ممكن ان تتعرض له المعلومات أثناء نقلها عبر قنوات الاتصال يكون اقل بكثير.

من هذا جاءت كل تطبيقات نظرية المعلومات التي تقدم بها شانون وكانت قفزة عملاقة سمحت بالانتقال الرقمي للمعلومات وسمحت بالتشفير ومجالات أخرى لا حصر لها

*د.نضال مقسوم, " نظرية المعلومات -قفزة-عبقريّة - " ,جامعة الشارقة.

سيتم في هذا الفصل دراسة مادة نظرية المعلومات وضغط البيانات وحسب المفردات التالية:

1. Introduction
2. Information, Information theory.
3. Measure of Information.
4. Self-Information, Types of Probability, Mutual Information, Information Rate.
5. Marginal Entropy, Joint Entropy, Conditional Entropy, Mutual information.
6. Communication Channel & Channel Capacity (symmetric channels, Properties of channel capacity) with examples

7. Source Coding (Entropy Encoding): Shannon-Fano-Elias coding
8. Huffman codes
9. Arithmetic coding,
10. Preview of the channel coding theorem,
11. Hamming code,

Introduction

Information

We intuitively know what information is. We constantly receive and send information in the form of text, sound, and images. We also feel that information is an elusive non mathematical quantity that cannot be precisely defined, captured, or measured.

نحن نعرف بشكل حدسي ما هي المعلومات. نتلقى المعلومات باستمرار ونرسلها على شكل نص وصوت وصور. كما أننا نشعر أن المعلومات هي كمية غير محسوبة لا يمكن تحديدها بدقة أو التقاطها أو قياسها.

Uncertainty and Information

Uncertainty refers to epistemic situations involving imperfect or unknown information. It applies to predictions of future events.

يشير عدم اليقين إلى الحالات المعرفية التي تشمل معلومات غير كاملة أو غير معروفة. مثل تنبؤات الأحداث المستقبلية.

Quantifying information is based on the observation that the **information** content of a message is equivalent to the amount of **surprise** in the message.

يتم تحديد كمية المعلومات اعتماداً على ملاحظة أن محتوى المعلومات في الرسالة يعادل كمية المفاجأة في الرسالة.

If I tell you something that you already know: (for example, “you and I work in university of Mosul”),

I **haven't given** you any information.

If I tell you something new: (for example, “we both received a raise”),

I **have given** you **some** information.

If I tell you something that really surprises you:(for example, “only I received a raise”),

I **have given** you **more information**, regardless of the number of words I have used, and of how you feel about my information.

إذا أخبرتك بشيء تعرفه بالفعل (على سبيل المثال ، "أنت وأنا أعمل في جامعة الموصل") ، فلن أعطيك أي معلومات. إذا أخبرتك بشيء جديد (على سبيل المثال ، "تلقينا كلانا زيادة") ، فقد قدمت لك بعض المعلومات. إذا أخبرتك بشيء يفاجئك حقًا (على سبيل المثال ، "تلقيت زيادة فقط") ، فقد قدمت لك المزيد من المعلومات ، بغض النظر عن عدد الكلمات التي استخدمتها ، وكيف تشعر حيال معلوماتي.

The three sentences carry different amounts of information:

A. (Tomorrow, the sun will rise from the East).

In fact, the first sentences hardly carry any information. Everybody knows that the sun rises in the east.

B. (The phone will ring in the next one hour).

This sentence is carrying **more information** than sentence (A). The phone may ring, or it may not. There is a finite probability that the phone will ring in the next one hour.

C. (It will snow in Delhi this winter).

It has never snowed in Delhi, and the probability of snowfall is very low.

The amount of information carried by the sentences listed above have something to do with the probability of occurrence of the events started in the sentences. And we observe an inverse relationship. Sentence (A), which talks about an event which has a probability of occurrence very close to 1 carries almost no information. Sentence (C), which has very low probability of occurrence, it has carried a lot of information.

the length of the sentence has nothing to do with the amount of information it carries. Sentence (A) is the longest but carries the minimum information. We will now develop a mathematical measure of information:

Information is certain knowledge about certain things, which may or may not be conceived by an observer.

المعلومات هي معرفة معينة عن أشياء معينة ، والتي يمكن أو لا يمكن تصورها من قبل المراقب.

- Example: music, story, news, text, sound, and images, etc.
- Information is not always meaningful! ذو معنى !
- The Information within a source is the uncertainty of the source.
- The length of the sentence has nothing to do with the amount of information it carries.



Any information source:

- analog
- digital,

Shannon wants to find a way for “**reliably**” transmitting data throughout the channel at “**maximal**” possible rate.

اراد شانون إيجاد طريقة لنقل البيانات بشكل "موثوق به" عبر القناة بأعلى نسبة محتملة.

The general communication system is modelled by:

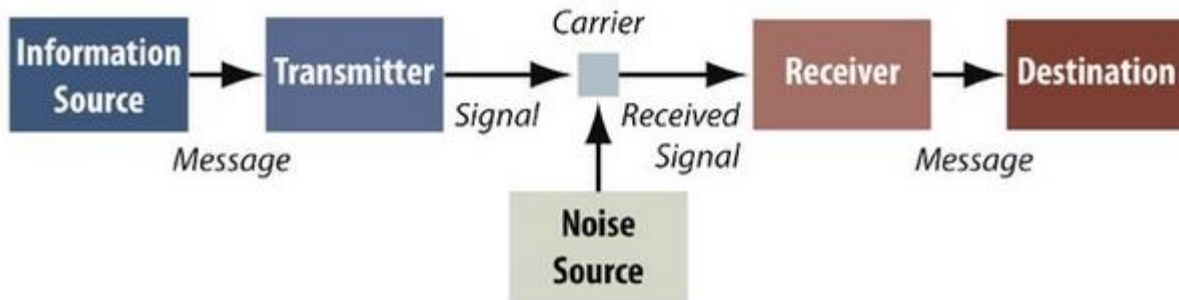


Figure 1: Shannon's schematic diagram of a general communication system.

Any communication system can be divided to five parts.

Information Source, Transmitter, Channel, Receiver, and Destination.

- (a) Information Source: This is the origin or source of information which is be transmitted.

Message: a stream of symbols taking their values in a predefined alphabet

Example 1:

Source: book

Message: a text

Symbols: letters alphabet= (a,..., z)

Example 2:

Source: a camera

Message: a picture

Symbols: RGB coefficients Alphabet= (0...255)

- (b) Transmitter: Transmitter converts the Message (from Source) to Signal (that can be sent over the channel). In case of Telephone, the Handset converts the sound pressure at the speaker to Electric Current that can be sent over wire(s)

- (c) Channel(carrier): A Channel is essentially any kind of Medium that carries the signal from Transmitter to the receiver. Some common examples include wires, co-axial cables, beam of light (optical fibers), Hard Disk, etc..
- (d) Receiver: The Job of Receiver is just opposite to that of the Transmitter. It has to convert the signal delivered by the Channel/Medium to Message. Similar to the case of transmitter in Telephone systems, the receiver (earpiece) converts Electric Current to Sound Pressure.
- (e) Destination: Destination of a communication system is the final user for whom the entire system is designed for. Generally, the Destination requires the faithful reproduction of information produced by the Information Source.

Figure 1 depicts the block diagram/ schematic of a generic Communication System. The Block 'Noise Source' is crucial, because a channel is not always 'Ideal', many physical phenomenon add up contributing to damage caused to the signal as it goes from transmitter to the receiver.

In the content of communications, information theory deals with mathematical modelling and analysis of a communication system rather than with physical sources and physical channels

Information theory

The importance of information theory is that it quantifies information. It shows how to measure information, so that we can answer the question "how much information is included in this piece of data?" with a precise number!

Information theory answers two fundamental questions in communication theory:

1. What is the ultimate data compression? (answer: the entropy H), and
2. What is the ultimate transmission rate of communication? (answer: the channel capacity C). For this reason some consider information theory to be a subset of communication theory .

رابط المحاضرة 2 <https://youtu.be/mZ5VCauZtrM>

نظرية المعلومات تجيب عن سؤالين أساسيين في نظرية الاتصالات: ما هو ضغط البيانات النهائي (الإجابة: الإنتروبيا H) ، وما هو معدل الإرسال النهائي للاتصال (الإجابة: سعة القناة C). ولهذا السبب يعتبر البعض نظرية المعلومات مجموعة فرعية من نظرية الاتصال.

Entropy is a measure of uncertainty or randomness (number of bits per symbol).

$$H(X) = - \sum_{i=1}^n p_i \log_2 p_i$$

Capacity: is defined as the ability of a channel to convey information.

A result that emerges from information theory is that if the entropy of the source is less than the capacity of the channel, then error-free communication over the channel can be achieved.

Figure 2 illustrates the relationship of information theory to other fields. As the figure suggests, information theory intersects physics (statistical mechanics), mathematics (probability theory), electrical engineering (communication theory), and computer science (algorithmic complexity).

يوضح الشكل 2 علاقة نظرية المعلومات بالحقول الأخرى. وكما تقترح النظرية ، فإن نظرية المعلومات تتقاطع مع الفيزياء (الميكانيكا الإحصائية) ، والرياضيات (نظرية الاحتمالات) ، والهندسة الكهربائية (نظرية الاتصالات) ، وعلوم الكمبيوتر (التعقيد الخوارزمي).

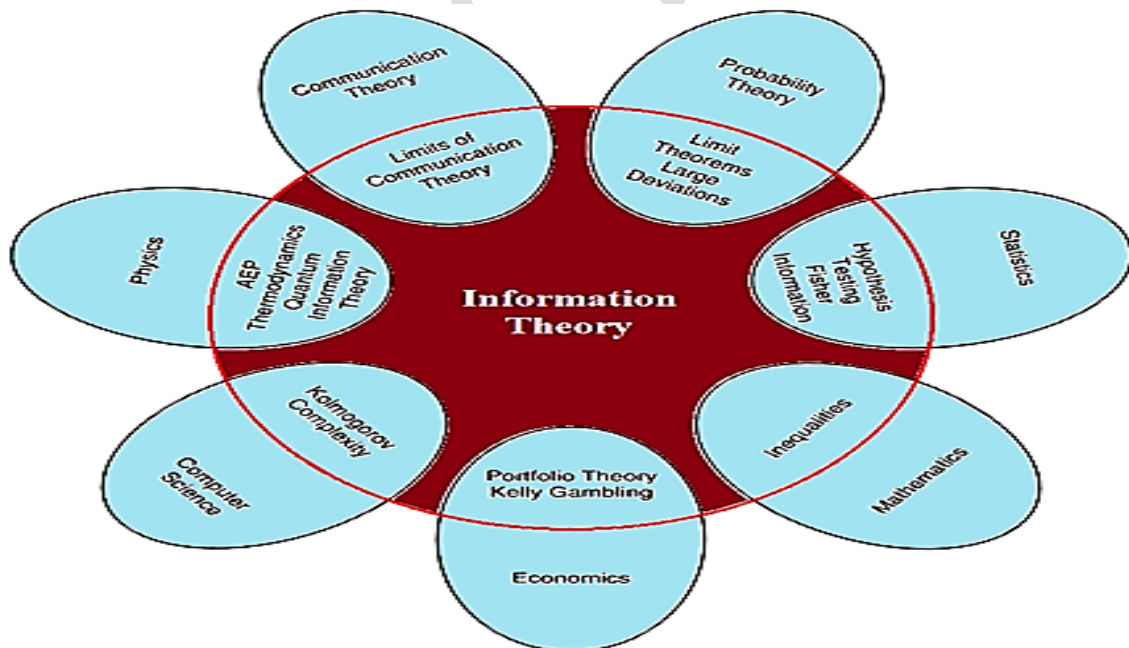


FIGURE 2 Relationship of information theory to other fields.

Summary:

Information Theory tells us...

- What exactly information is. ما هي المعلومات بالضبط
- How they are measured and presented. كيف يتم قياسها وعرضها
- Implications, limitations, and applications. الآثار والقيود والتطبيقات



Shannon Theory :the original 1948 Shannon Theory contains:

1. Measurement of Information
2. Source Coding Theory
3. Channel Coding Theory

Summary:

$$P(event) = \frac{\text{number of outcomes in event}}{\text{number of outcomes in sample space}}$$

- $0 \leq p \leq 1$
- $p=0$ impossible event
- $P=1$ certain (sure) event
- $\sum_{i=1}^n P(x_i) = 1$ in any particular situation (non equal probability)
- $P(X_1) = P(X_2) = \dots = P(X_N) = \frac{1}{N}$ in equal probability
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ Compound probability
- In Mutually exclusive events $P(A \cap B) = 0$
- $P(A \text{ and } B) = P(A) \cdot P(B)$ Independent and Dependent Events
- $P(A|B) = P(A \text{ and } B) / P(B)$ Conditional probability
- $P(A^c) = 1 - P(A)$ Complement of an event
- The probability is similar meaning of percentage
- The probability very low then the information is high.

Example 1:

A single coin is tossed 5 times. What is the probability of getting at least one head?

Solution:

Consider solving this using complement.

Probability of getting no head = $P(\text{all tails}) = (1/2)^5 = 1/32$

$P(\text{at least one head}) = 1 - P(\text{all tails}) = 1 - 1/32 = 31/32$.



Example 2:

The percentage of success is 50%, so the probability is $P=0.5$, Where, the probability (p) has a value between 0 and 1,

Example 3:

A coin is tossed **three** times. Use this information to solve these problems.

- 1) Number of variables. And type of variables.
- 2) List the **sample space**.
- 3) Find the probability of tossing **heads** exactly **twice**.
- 4) Find the probability of tossing **tails** at **least twice**.



Solution:

- 1) Number of variables=3 . And type of variables are 2 = {H,T}
- 2) For 3 tosses of the coin all the possible outcomes are
 $S = \{HHH, THH, HTH, HHT, THT, TTH, HTT, TTT\}$
 $n(S) = 8$ (n is the number of possible outcomes in sample space)
- 3) Firstly find the event A which represents the subset of S and any outcome that has exactly **two H's**.
 Event A = {THH, HTH, HHT} , $n(A) = 3$

$$P(A) = \frac{n(A)}{n(S)} = \frac{3}{8} = 0.375 * 100 = 37.5 \%$$
- 4) The outcomes that have at least two tails in them are
 Event B = {THT, TTH, HTT, TTT}
 $n(B) = 4$

$$p(B) = \frac{n(B)}{n(S)} = \frac{4}{8} = 0.5 = 50\%$$

Example 4:

- Message is { \$\$*#*\$*#\$\$}, Find
- 1) Number of variables.
 - 2) Types of variables.
 - 3.) The probability of each variable in message.

Solution:

- 1) Number of variables=10
- 2) And type of variables=3 { \$, #, * }
- 3)

Variable(x)	No.(x)	P(x)
\$	5	$P(\$) = \frac{5}{10} = 0.5$
*	3	$P(*) = \frac{3}{10} = 0.3$
#	2	$P(\#) = \frac{2}{10} = 0.2$

Example 5:

Message of 3 variable $P(x_1)=P(x_2)=0.3$. Find probability of third variable.

Solution.

$$\begin{aligned}P(x_1)+p(x_2)+P(x_3) &= 1 \\P(x_3) &= 1 - 0.3 - 0.3 \\&= 0.4\end{aligned}$$

Example 6:

binary Message $p(0)=0.6$. If number of variable in message=10 000 variables. Find number of 0&1.

Solution:

$$\begin{aligned}\text{Binary Message: Two variable} &= \{0,1\} \\ \text{If } P(0) &= 0.6 \text{ then } P(1) = 1 - P(0) = 0.4 \\ \text{No. of variable (0)} &= \text{total variable} * P(0) \\ &= 10000 * 0.6 \\ &= 6000 \text{ variable} \\ \text{No. of variable (1)} &= \text{total variable} * P(1) \\ &= 10000 * 0.4 \\ &= 4000 \text{ variable}\end{aligned}$$

Example 7:

Message with 5 variable , equiprobability . Find probability of every variable.

Solution.

$$N=5, \quad P(x_1)=P(x_2)=P(X_3)=P(X_4)=P(X_5) = \frac{1}{5}$$

<https://youtu.be/e7ZKM2WA9ul>

رابط المحاضرة 3:

The Measure of Information (self-information)

There are two results, so the result of any toss is initially uncertain. We have to actually throw the coin in order to resolve the uncertainty. The result is heads or tails, which can also be expressed as a yes or no, or as a 0 or 1; a bit.



A **single bit resolves the uncertainty in the toss of a coin**. What makes this example important is the fact that it can easily be generalized. Many real-life problems can be resolved, and their solutions expressed, by means of several bits. The principle of doing so is to find the minimum number of yes/no questions that must be answered in order to arrive at the result. Since the answer to a yes/no question can be expressed with one bit, the number of questions will equal the number of bits it takes **to express the information contained in the result**.

- A deck of **64 playing cards**.

For simplicity let's ignore their traditional names and numbers and simply number them 1 to 64. Consider the event of person A drawing one card and person B having



B



A

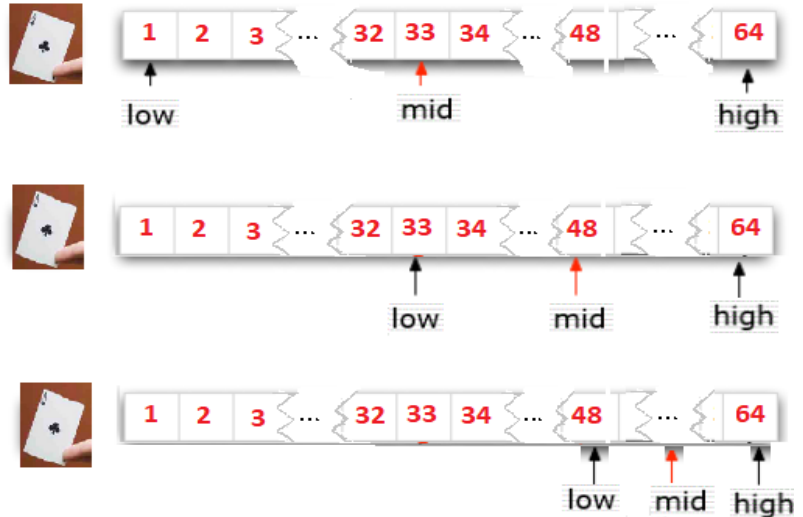
to guess what it was. The guess is **a number between 1 and 64**. What is the minimum number of yes/no questions that are necessary to guess the card?

Those who are familiar with the **technique of binary search** know the answer. Using this technique, B should divide the interval 1–64 in two, and should start by asking “is the result between 1 and 32?” If the answer is no, then the result is in the interval 33 to

<https://youtu.be/e7ZKM2WA9ul>

رابط المحاضرة: 3:

64. This interval is then divided by two and B's next question should be "is the result between 33 and 48?" This process continues until the interval selected by B reduces to a single number.



It does not take much to see that exactly six questions are necessary to get at the result. This is because 6 is the number of times 64 can be divided in half.

Mathematically,

this is equivalent to writing $6 = \log_2 64$.

This is why the logarithm is the mathematical function that quantifies information.

Another approach to the same problem is to ask the question; given a nonnegative integer N, how many digits does it take to express it?

The answer, of course, depends on N. The greater N, the more digits are needed.

The first 100 nonnegative integers (0 to 99) can be expressed by two decimal digits. The first 1000 such integers can be expressed by three digits.

then The number of digits required to represent N equals approximately $\log N$. The base of the logarithm is the same as the base of the digits. For decimal digits, use base 10; for binary digits (bits) use base 2.

the number of digits it takes to express N is proportional to the information content of N, then again the logarithm is the function that gives us a measure of the information.

Mathematical Measure of Information (self-information)

We would like to develop a usable measure of the information we get from observing the occurrence of an event having probability p . Our first reduction will be to ignore any particular features of the event, and only observe whether or not it happened. Thus we will think of an event as the observance of a symbol whose probability of occurring is p . we will thus be defining the information in terms of the probability p . A specific special case of interest is probabilities (i.e., real numbers between 0 and 1), we will want our information measure $I(p)$ to have several properties: -

- Information is a non-negative quantity: $I(p) \geq 0$.
- If an event has probability 1, we get no information from the occurrence of the event: $I(1) = 0$.
- If two independent events occur (whose joint probability is the product of their individual probabilities), then the information we get from observing the events is the sum of the two information: $I(p_1 * p_2) = I(p_1) + I(p_2)$:(This is the critical property . . .)

<https://youtu.be/e7ZKM2WA9uI>

رابط المحاضرة 3:

Summary:

Measurement of Information(self-information)

- All events are probabilistic !
- Shannon's first question is

“How to measure information in terms of bits?”


- Using Probability Theory, Shannon showed that there is only one way to measure information in terms of number of bits:

$$H(X) = H(p_1, \dots, p_n) = - \sum_{i=1}^n p_i \log_2 p_i$$

Shannon entropy


called the **entropy function**

- Shannon entropy is a measure of uncertainty.
- A key measure in information theory is "entropy". Entropy quantifies the amount of uncertainty involved in the value of a random variable or the outcome of a random process.

 = ? bits

 = ? bits

Or Lottery!?

 = ? bits

رابط المحاضرة 3: <https://youtu.be/e7ZKM2WA9ul>

نظرية المعلومات information theory :: هي أحد تخصصات وفروع الرياضيات التطبيقية الذي يتضمن تحويل البيانات الى كمية بهدف تمكين تخزينها في وسط ما أو نقلها عبر قناة اتصال ما بأكبر قدر ممكن.

Entropy : قياس المعلومات يعرف عادة بانتروبية المعلومات وهو عبارة عن عدد من البتات (متوسط عدد النبضات الثنائية) اللازم للتخزين أو الاتصال.

Information(self-information)

• Summarizing:

$$\therefore I(p) = \log_b(1/p) = -\log_b(p),$$

$$\therefore I(p) \geq 0$$

$$\therefore I(p_1 * p_2) = I(p_1) + I(p_2)$$

$$\therefore I(1) = 0$$

$$\blacksquare I(p^2) = I(p * p) = I(p) + I(p) = 2 * I(p)$$

$$\blacksquare I(p^n) = n * I(p)$$

$$\blacksquare I(p^{n/m}) = \frac{n}{m} * I(p)$$

$$\blacksquare \text{ for } 0 < p \leq 1, \text{ and } a > 0 \text{ a real number:}$$

$$I(p^a) = a * I(p)$$

NOTE: The standard convention of information theory is to take $b=2$, the corresponding unit of information is termed the bit, a contraction for binary digit. We will illustrate the definition further with a few examples.

<https://youtu.be/e7ZKM2WA9ul>

رابط المحاضرة 3:

Example:

Message of 3 variables, $P(x_1)=P(x_2)=0.2$. Calculate:

- 1) $I(x_2)$ & $I(x_3)$ 2) Entropy

Solution:

$$1) \quad \sum_{i=1}^3 P(x_i) = 1$$

$$P(x_3) = 1 - P(x_1) - P(x_2) = 0.6$$

VAR.	$P(X_i)$
x1	0.2
x2	0.2
x3	0.6

$$I(x_2) = -\log P(x_2) = -\log (0.2) = 2.321 \text{ bit}$$

$$I(x_3) = -\log P(x_3) = -\log (0.6) = 0.737 \text{ bit}$$

$$2) \quad H(X) = -\sum_x p(x) \log_2 p(x)$$

$$H(x) = - [0.2 \log 0.2 + 0.2 \log 0.2 + 0.6 \log 0.6]$$

$$= 1.37 \text{ bit /symbol}$$

Example:

Equal probability message of 10 variable. Find Entropy.

Solution.

$$N=10, \quad P(x_1)=P(x_2) \dots = P(x_{10}) = \frac{1}{10}$$

$$H(x) = \log_2 10 = 3.3 \text{ bit /symbol}$$

<https://youtu.be/e7ZKM2WA9uI>

رابط المحاضرة: 3

Example:

Suppose that we have a horse race with eight horses taking part. Assume that the probabilities of winning for the eight horses are $(\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \frac{1}{64}, \frac{1}{64}, \frac{1}{64}, \frac{1}{64})$. We can calculate the entropy of the horse race as:

$$H(x) = -\frac{1}{2} \log_2 \frac{1}{2} - \frac{1}{4} \log_2 \frac{1}{4} - \frac{1}{8} \log_2 \frac{1}{8} - \frac{1}{16} \log_2 \frac{1}{16} - 4 \frac{1}{64} \log_2 \frac{1}{64} \\ = 2 \text{ bits/symbol}$$

Examples:

- (1) Letters of alphabet. If the letters are assumed equiprobable, then the probability of a given letter is $(\frac{1}{26})$ and $I = \log_2 26 = 4.7$ bits.
- (2) Decimal Numbers. Assuming that numbers from 0 to 9 are equiprobable, then $I = \log_2 10 = 3.32$ bits.
- (3) In a binary system producing 0s and 1s with equal probability where $P(0) = P(1) = 1/2$, so the information per digit is $I = \log_2 2 = 1$ bit.
- (4) An already known event. If $P=1$ then $I = -\log_2 1 = 0$.
- (5) Sequence of binary digit. Suppose the sequence 1011 of four digits is sent. If 0s and 1s are equiprobable, then 4 bits are received. Alternatively, the probability of obtaining this sequence is $(\frac{1}{2})^4 = (\frac{1}{16})$, so the information is $\log 16 = 4$ bits again (this confirms again that the definition gives the correct result regarding the length of the message).
- (6) A sequence of decimal numbers (0-9), suppose the sequence (26314) of five numerals is sent, if the numbers (0-9) are equiprobable, then the probability of obtaining this sequence is $(\frac{1}{10})^5 = (\frac{1}{100000}) = 0.00001$
 $I = \log_2 10^5 = 5 \log_2 10 = 5 * 3.3 = 16.6 \text{ bits}$

<https://youtu.be/e7ZKM2WA9ul>

رابط المحاضرة 3:

Summary

- identifying the outcome of a fair coin flip (with two equally likely outcomes) provides less information (lower entropy) than specifying the outcome from a roll of a die (with six equally likely outcomes).



- In equal probability: $P(x_i) = \frac{1}{N}$

$$\text{Then } H(x) = - \left[\frac{1}{N} \log_2 \frac{1}{N} \right] * N$$

$$= - \log_2 \frac{1}{N}$$

$$H(x) = \log_2 N$$

- The entropy of data depends on the individual probabilities P_i .**

Least probability \Longrightarrow More information

Most probable \Longrightarrow Less information

The probability very low then the information is high.

Self-Information, Types of Probability, Mutual Information, Information Rate

Exercises

1- Self-information

1. A source produces one of four possible symbols during each interval having probabilities $p(x_1) = \frac{1}{2}$, $p(x_2) = \frac{1}{4}$, $p(x_3) = \frac{1}{8}$, obtain (or find) the information content (محتوى المعلومات) of each of these symbols.

Solution: we know that the information content $I(x_i)$ of a symbol x_i is given by

$$I(x_i) = \log_2 \frac{1}{p(x_i)}$$

Thus, we can write

$$I(x_1) = \log_2 \frac{1}{\frac{1}{2}} = \log_2(2) = 1 \text{ bit}$$

$$I(x_2) = \log_2 \frac{1}{\frac{1}{4}} = \log_2(2^2) = 2 \text{ bits}$$

$$I(x_3) = \log_2 \frac{1}{\frac{1}{8}} = \log_2(2^3) = 3 \text{ bits}$$

$$I(x_4) = \log_2 \frac{1}{\frac{1}{8}} = \log_2(2^3) = 3 \text{ bits}$$

2. Calculate the amount of information if it is given that $P(x_i) = \frac{1}{4}$.

Solution: we know that amount of information $I(x_i)$ of a discrete symbol x_i is given by,

$$I(x_i) = \log_2 \frac{1}{P(x_i)}$$

The above expression may be written as under:

$$I(x_i) = \frac{\log_{10} \frac{1}{p(x_i)}}{\log_{10} 2}$$

Substituting given value of $p(x_i)$ in above expression, we obtain

$$I(x_i) = \frac{\log_{10} 4}{\log_{10} 2} = 2 \text{ bits}$$

3. The binary symbols `0` and `1` are transmitted with probabilities $\frac{1}{4}$ and $\frac{3}{4}$ respectively. Find the corresponding self-information.

Solution:

Self-information in a `0` = $I_0 = \log_2 \frac{1}{P(x_i)} = \log_2 \frac{1}{1/4} = 2 \text{ bits}$

Self-information in a `1` = $I_1 = \log_2 \frac{1}{P(x_i)} = \log_2 \frac{1}{3/4} = 0.415 \text{ bits}$

4. The text {00000mmnm000iiij}

- a- Find probability of every sample.
 b- Find information of m,n ($I(m), I(n)$).

a- Sol.

<u>Var.</u>	<u>No.</u>	<u>P(x)</u>
<u>0</u>	<u>8</u>	<u>8/16</u>
<u>M</u>	<u>3</u>	<u>3/16</u>
<u>N</u>	<u>1</u>	<u>1/16</u>
<u>I</u>	<u>3</u>	<u>3/16</u>
<u>j</u>	<u>1</u>	<u>1/16</u>

b- Sol.

$I(m) = -\log_2 p(m) = -\log_2 3/16 =$ bits

$I(n) = -\log_2 p(n) = -\log_2 1/16 =$ bits

2- Probability Types:

Probabilities may be either: marginal, joint or conditional.

- 1- **Marginal probability**: the probability of an event occurring ($p(A)$), it may be thought of as an unconditional probability. It is not conditioned on another event.
 احتمال وقوع حدث ($p(A)$) ، يمكن اعتباره احتمالية غير مشروطة. لا يشترط على حدث آخر.
- 2- **Joint probability**: $p(A \text{ and } B)$. The probability of event A **and** event B occurring. It is the probability of the intersection of two or more events. The probability of the intersection of A and B may be written $p(A \cap B)$.
 احتمالية وقوع الحدث A والحدث B . هو احتمال تقاطع حدثين أو أكثر.
- 3- **Conditional probability**: $p(Y|X)$ is the probability of event Y occurring, given that event X occurs

هو احتمال وقوع الحدث Y ، شرط وقوع الحدث X

Example:

A survey was carried out with 500 persons in to determine people's favorite sports. The options were Football, Rugby (a team game played with an oval ball that may be kicked, carried, and passed from hand to hand.) and the rest was grouped together in Other; The results of the test are displayed in Figure 1.

	Male	Female	Total
Football	120	75	195
Rugby	100	25	125
Other	50	130	180
	270	230	500

Figure 1: The Results of the test

Figure 1 is not quite a probability distribution, but if we want to get the probability distribution, we can divide each number in Figure 1 by 500 and the result will be the image in Figure 2.

Joint Probability

The Joint probability is a statistical measure that is used to calculate the probability of two events occurring together at the same time — $P(A \text{ and } B)$ or $P(A,B)$. For example, using Figure 2 we can see that the joint probability of someone being a male and liking football is 0.24.

	Male	Female	Total
Football	0.24	0.15	0.39
Rugby	0.2	0.05	0.25
Other	0.1	0.26	0.36
	0.54	0.46	1

Figure 2: Probability Distribution

	Male	Female	Total
Football	0.24	0.15	0.39
Rugby	0.2	0.05	0.25
Other	0.1	0.26	0.36
	0.54	0.46	1

Figure 3: The Joint Probability Distribution.

Note: The cells highlighted in Figure 3 (the Joint Probability Distribution) must sum to 1 because everyone in the distribution must be in one of the cells.

The Joint probability is symmetrical meaning that $P(\text{Male and Football}) = P(\text{Football and Male})$ and we can also use it to find other types of distributions, the marginal distribution and the conditional distribution.

Marginal Distribution

In probability theory and statistics, the marginal distribution of a subset of a collection of random variables is the probability distribution of the variables contained in the subset. It gives the probabilities of various values of the variables in the subset without reference to the values of the other variables. The marginal probability is the probability of an event irrespective of the outcome of another variable — $P(A)$ or $P(B)$.

	Male	Female	Total
Football	0.24	0.15	0.39
Rugby	0.2	0.05	0.25
Other	0.1	0.26	0.36
	0.54	0.46	1

Figure 4: The Marginal Distribution

Conditional Probability

The conditional probability concept is one of the most fundamental in probability theory. It defines the probability of one event occurring given that another event has occurred

$$P(A|B) = P(A, B) / P(B)$$

Figure 5: Expression of the Conditional Probability

Figure 2; If we want to calculate the probability that a person would like Rugby given that they are a female, we must take the joint probability that the person is female and likes rugby ($P(\text{Female and Rugby})$) and divide it by the probability of the condition. In this case, the probability is that the person is a female ($P(\text{Female})$) which we can work out from the margin to be 0.46 hence we get 0.11 (2 decimal places).

Let's write that up neater:

$$P(\text{Female, Rugby}) = 0.05$$

$$P(\text{Female}) = 0.46$$

$$P(\text{Rugby} | \text{Female}) = 0.05 / 0.46 = 0.11 \text{ (to 2 decimal places).}$$

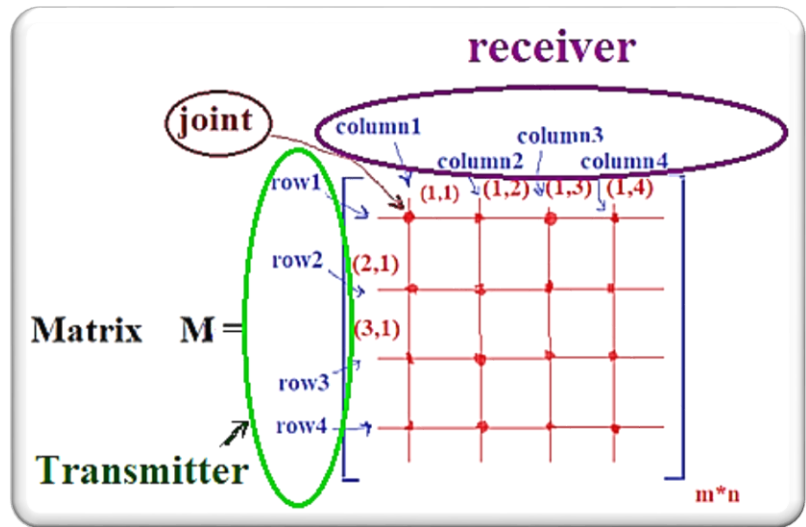
If we continued to fill in the probability of preferring a sport given the observant is a female then we would have a Conditional Probability Distribution.

Representation in communication system

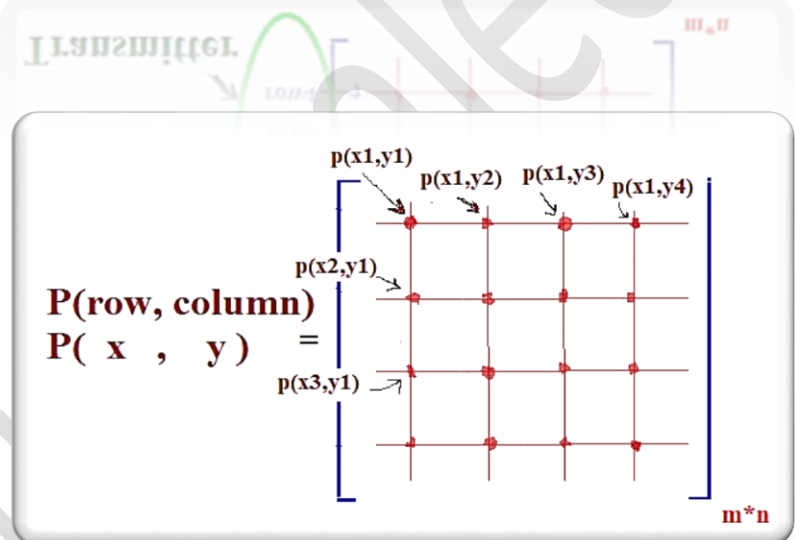
- Joint probability
 (two variable)

Columns: represent receiver

Rows: transmitter



Joint probability $p(x,y)$:
 connection between two points
 probability represented by ' , '.



- Marginal probability:

$$P(x_i) = \sum_{j=1}^m p(x_i, y_j)$$

$$P(x_1) = \sum_{j=1}^4 p(x_1, y_j)$$

$$P(x_2) = \sum_{j=1}^4 p(x_2, y_j)$$

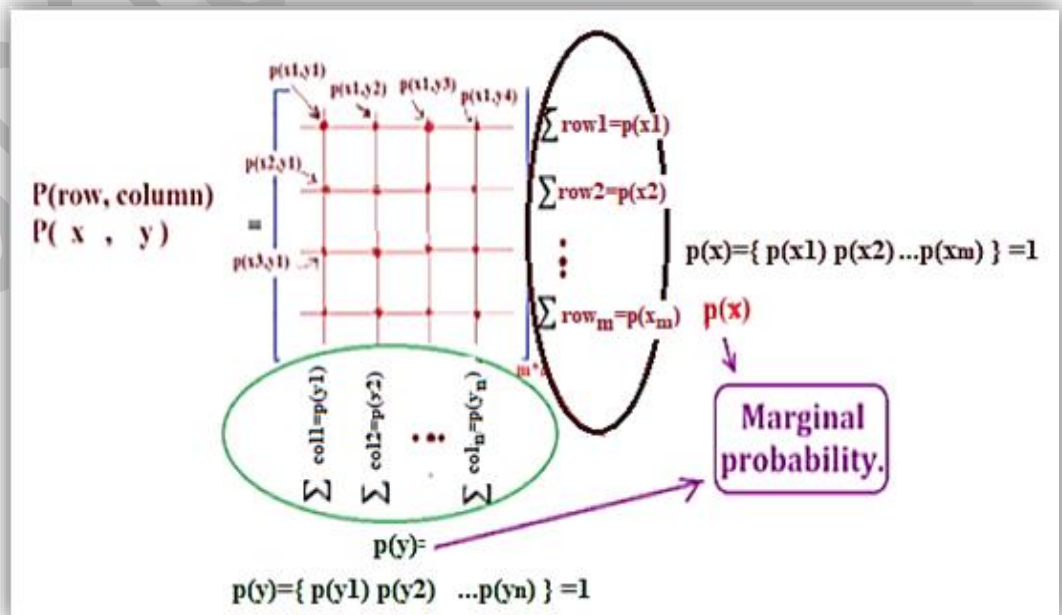
$$P(x_m) = \sum_{j=1}^4 p(x_m, y_j)$$

$$P(y_i) = \sum_{i=1}^n p(x_i, y_j)$$

$$P(y_1) = \sum_{i=1}^5 p(x_i, y_1)$$

$$P(y_2) = \sum_{i=1}^5 p(x_i, y_2)$$

$$P(y_n) = \sum_{i=1}^5 p(x_i, y_n)$$



Example: if you have the joint probability matrix shown below, find the transmitted and received prob. (marginal prob.).

$$p(x,y) = \begin{matrix} & \begin{matrix} y1 & y2 \end{matrix} \\ \begin{matrix} x1 \\ x2 \\ x3 \end{matrix} & \begin{bmatrix} 0.1 & 0.25 \\ 0 & 0.2 \\ 0.25 & 0.2 \end{bmatrix} \end{matrix}$$

Solution:

$P(x_i) = \sum_{j=1}^2 p(x_i, y_j)$	$P(y_i) = \sum_{i=1}^3 p(x_i, y_j)$
$P(x1) = 0.1 + 0.25 = 0.35$	$P(y1) = 0.1 + 0 + 0.25 = 0.35$
$P(x2) = 0 + 0.2 = 0.2$	$P(y2) = 0.25 + 0.2 + 0.2 = 0.65$
$P(x3) = 0.25 + 0.2 = 0.45$	

Marginal probability.

$$P(x) = [0.35 \quad 0.2 \quad 0.45] \text{ . note } P(x) = 0.35 + 0.2 + 0.45 = 1$$

$$P(y) = [0.35 \quad 0.65] \text{ . note } P(y) = 0.35 + 0.65 = 1$$

• Conditional probability.

$p(y|x)$: y is unknown, x is known

dependent events $p(y|x) = \frac{p(x \cap y)}{p(x)} = \frac{p(x,y)}{p(x)}$, $p(x|y) = \frac{p(x \cap y)}{p(y)} = \frac{p(x,y)}{p(y)}$

independent event

- $P(y,x) = P(y) * P(x)$
- $p(y|x) = p(y)$
- $p(x|y) = p(x)$

Example : Roll a dice once , event A=6 appear, event B an even number appears, find the condition probability.

Solution:

Sample space= {1,2,3,4,5,6}

A={6} B={ 2,4,6}

$$p(B|A) = \frac{p(A \cap B)}{p(A)} = \frac{p(A,B)}{p(A)}$$

$$A \cap B = \{6\} \quad p(A \cap B) = 1/6$$

$$p(A)=1/6 \quad p(B)=3/6$$

$$p(B|A) = \frac{p(A \cap B)}{p(A)} = \frac{1/6}{1/6} = 1$$

$$p(A|B) = \frac{p(A \cap B)}{p(B)} = \frac{1/6}{3/6} = 1/3$$

Example: if you have the joint probability matrix shown below, find the condition probability

	y1	y2
x1	0.1	0.25
x2	0	0.2
x3	0.25	0.2

Solution:

$$p(y|x) = \frac{p(x,y)}{p(x)} \quad , \quad p(x|y) = \frac{p(x,y)}{p(y)}$$

Marginal probability. $\left\{ \begin{array}{l} P(x) = [0.35 \quad 0.2 \quad 0.45] \text{ . note } P(x) = 0.35 + 0.2 + 0.45 = 1 \\ P(y) = [0.35 \quad 0.65] \text{ . note } P(y) = 0.35 + 0.65 = 1 \end{array} \right.$

$$P(x) = [0.35 \quad 0.2 \quad 0.45] \quad p(y|x) = p(x,y)/p(x) = \begin{bmatrix} 0.1/0.35 & 0.25/0.35 \\ 0/0.2 & 0.2/0.2 \\ 0.25/0.45 & 0.2/0.45 \end{bmatrix} = \begin{bmatrix} 0.286 & 0.714 \\ 0 & 1 \\ 0.556 & 0.444 \end{bmatrix}$$

$$P(y) = [0.35 \quad 0.65] \quad p(x|y) = p(x,y)/p(y) = \begin{bmatrix} 0.1/0.35 & 0.25/0.65 \\ 0/0.35 & 0.2/0.65 \\ 0.25/0.35 & 0.2/0.65 \end{bmatrix} = \begin{bmatrix} 0.286 & 0.384 \\ 0 & 0.307 \\ 0.714 & 0.307 \end{bmatrix}$$

Mutual information المتبادلة او المنقولة

It is defined as the amount of information transferred where x_i is transmitted and y_i is received.

Definition The mutual information between two discrete random variables X, Y jointly distributed according to $p(x, y)$ is given by

$$I(x_i; y_i) = \sum_{x_i} \sum_{y_i} p(x_i, y_i) \log_2 \frac{p(x_i, y_i)}{p(x_i) \cdot p(y_i)}$$

Channel Capacity

define the channel capacity, C , as the maximum mutual information with respect to the input distribution, P_X ,

$$C = \max_{p_X} I(X; Y).$$

The Information Rate

The information rate is represented by R and it is given as,

$$\text{Information Rate : } R = r H \quad \text{OR}$$

$$R(X) = \frac{H(X)}{\tau}, \quad \tau = \sum_{i=1}^n P(X_i) \tau(X_i) \quad \leftarrow \text{time}$$

Here R is the information rate.

H is the Entropy or average information

And r is the rate at which messages are generated.

Information rate R is represented in average number of bits of information per second.

It is calculated as follows:

$$R = \left(r \text{ in } \frac{\text{messages}}{\text{second}} \right) * \left(H \text{ in } \frac{\text{bits}}{\text{messages}} \right)$$

$$= \text{bits / second}$$

Example:

APCM source transmits **four samples (messages)** with a **rate 2B**(bandwidth) **samples / second**. The probabilities of occurrence of these 4 samples (messages) are **p1 = p4 = 1/8 and p2 = p3 = 3/8**.

Find out the information rate of the source.

Solution:

To calculate the Entropy (H): We have four messages with probabilities p1 = p4 = 1/8 and p2 = p3 = 3/8. Average information H (or entropy) is given by equation as;

$$H = p_1 \log_2 \left(\frac{1}{p_1} \right) + p_2 \log_2 \left(\frac{1}{p_2} \right) + p_3 \log_2 \left(\frac{1}{p_3} \right) + p_4 \log_2 \left(\frac{1}{p_4} \right)$$

$$= \frac{1}{8} \log_2 8 + \frac{3}{8} \log_2 \frac{8}{3} + \frac{3}{8} \log_2 \frac{8}{3} + \frac{1}{8} \log_2 8$$

$$H = 1.8 \text{ bits / message}$$

Message rate r = 2B samples / second

Hence, **R = r H**

Putting values of r and H in the above equation ,

$$R = 2B \text{ messages / second} * 1.8 \text{ bits / message}$$

$$= 3.6 B \text{ bits/second}$$

Example :

In the transmission scheme of example 1, calculate the information rate if all messages are equally likely.

Solution:

We know that there are four messages. Since they are equally likely, their probabilities will be equal to 1/4, i.e., $P_1 = p_2 = p_3 = p_4 = 1/4$ Average information per message (Entropy) is given by equation as,

$$\begin{aligned} H &= p_1 \log_2\left(\frac{1}{p_1}\right) + p_2 \log_2\left(\frac{1}{p_2}\right) + p_3 \log_2\left(\frac{1}{p_3}\right) + p_4 \log_2\left(\frac{1}{p_4}\right) \\ &= 4p \log_2\left(\frac{1}{p}\right) \quad \text{since } p_1 = p_2 = p_3 = p_4 \\ &= \log_2 4 \\ &= 2 \text{ bits/message} \end{aligned}$$

The information rate is given by equation as, $R = rH$

Here $r = 2B$ messages/ sec. as obtained in example 1.

$$\begin{aligned} R &= 2B \text{ messages / sec.} * 2 \text{ bits / message} \\ &= 4B \text{ bits / sec.} \end{aligned}$$

Example: Having the text (A A A B A C B C C B B A A A D B B B C D). If $\tau(A) = \tau(B) = \tau(C) = 0.1 \mu\text{sec}$ and $\tau(D) = 0.2 \mu\text{sec}$. Calculate average Source Entropy Rate $R(x)$.

Solution:

$$\begin{aligned} P(A) &= \frac{7}{20}, P(B) = \frac{7}{20}, P(C) = \frac{4}{20}, P(D) = \frac{2}{20} \\ H(x) &= - \frac{\left(\frac{7}{20} \times \ln\left(\frac{7}{20}\right) \times 2 + \frac{1}{5} \times \ln\left(\frac{1}{5}\right) + \frac{1}{10} \times \ln\left(\frac{1}{10}\right)\right)}{\ln(2)} = 1.85677 \text{ bits/} \\ &\quad \text{symbol} \end{aligned}$$

$$R(X) = \frac{H(X)}{\tau}$$

$$\overline{\tau} = \sum_i \tau_i * P(x_i) = \left(\frac{7}{20} + \frac{7}{20} + \frac{1}{5}\right) * 0.1 \times 10^{-6} + \frac{1}{10} \times 0.2 \times 10^{-6}$$
$$= 0.11 * 10^{-6} \text{ sec}$$

$$R(x) = \frac{1.8567}{0.11 \times 10^{-6}} = \text{bits/sec.}$$

Marginal Entropy, Joint Entropy, Conditional Entropy, Mutual information

Entropy

Entropy measures the amount of information in a random variable or the length of the message required to transmit the outcome

Definition: Entropy

If X is a discrete random variable and $p(x)$ is the value of its probability distribution at x , then the entropy of X is:

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

- Entropy is measured in bits (the log is \log_2);
- intuitively, it measures amount of information (or uncertainty) in random variable;
- it can also be interpreted as the length of message to transmit an outcome of the random variable;
- note that $H(X) \geq 0$ by definition.

Joint Entropy

joint entropy is the amount of information in two (or more) random variables;

Definition: Joint Entropy

If X and Y are discrete random variables and $P(x, y)$ is the value of their joint probability distribution at (x, y) , then the joint entropy of X and Y is:

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

The joint entropy represents the amount of information needed on average to specify the value of two discrete random variables.

Example $p(x,y)=$

	y1	y2	y3	y4
x1	0.25	0	0	0
x2	0.1	0.3	0	0
x3	0	0.05	0.1	0
x4	0	0	0.05	0.1
x5	0	0	0.05	0

Find: 1- Marginal Entropy. ($H(x)$, $H(y)$)
2- Joint Entropy. ($H(x,y)$)

Solution:

1- Marginal Entropy. ($H(x)$, $H(y)$)

$P(x_i)=\sum_{j=1}^4 p(x_i, y_j)$	$P(y_j)=\sum_{i=1}^5 p(x_i, y_j)$
$P(x_1)=0.25$	$P(y_1)=0.25+0.1=0.35$
$P(x_2)=0.1+0.3=0.4$	$P(y_2)=0.3+0.05=0.35$
$P(x_3)=0.05+0.1=0.15$	$P(y_3)=0.1+0.05+0.05=0.2$
$P(x_4)=0.05+0.1=0.15$	$P(y_4)=0.1$
$P(x_5)=0.05$	

$$P(x) = [0.25 \quad 0.4 \quad 0.15 \quad 0.15 \quad 0.05]$$

$$P(y) = [0.35 \quad 0.35 \quad 0.2 \quad 0.1]$$

Marginal probability.

Marginal Entropy.

$$H(x) = -\sum_{i=1}^5 p(x_i) \log_2 p(x_i)$$

$$= -[0.25 \log_2 0.25 + 0.4 \log_2 0.4 + 0.15 \log_2 0.15 + 0.15 \log_2 0.15 + 0.05 \log_2 0.05]$$

$$= 2.066 \text{ bits/symbol}$$

$$H(y) = -\sum_{j=1}^4 p(y_j) \log_2 p(y_j)$$

$$= -[0.35 \log_2 0.35 + 0.35 \log_2 0.35 + 0.2 \log_2 0.2 + 0.1 \log_2 0.1]$$

$$= 1.056 \text{ bits/symbol}$$

2- Joint Entropy

$$H(x, y) = -\sum_{i=1}^5 \sum_{j=1}^4 p(x_i, y_j) \log_2 p(x_i, y_j)$$

$$= -[0.25 \log_2 0.25 + 0.1 \log_2 0.1 + 0.3 \log_2 0.3 + 0.05 \log_2 0.05 + 0.1 \log_2 0.1 + 0.05 \log_2 0.05 + 0.1 \log_2 0.1 + 0.05 \log_2 0.05]$$

$$= -[0.25 \times (-2) + 0.1 \times (-3.321) + 0.3 \times (-1.7319) + 0.05 \times (-4.321) + 0.1 \times (-3.321) + 0.05 \times (-4.321) + 0.1 \times (-3.321) + 0.05 \times (-4.321)] = 2.665 \text{ bits/symbol}$$

Conditional Entropy

Conditional entropy is amount of information in one random variable given we already know the other.

Definition: Conditional Entropy

If X and Y are discrete random variables and $p(x, y)$ and $P(y|x)$ are the values of their joint and conditional probability distributions, then:

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

is the conditional entropy of Y given X .

The conditional entropy indicates how much extra information you still need to supply on average to communicate Y given that the other party knows X .

Conditional Entropy

For probability distributions we defined:

$$P(y|x) = \frac{P(x, y)}{P(x)}$$

A similar theorem holds for entropy:

Theorem: Conditional Entropy

If X and Y are discrete random variables with joint entropy $H(X, Y)$ and the marginal entropy of X is $H(X)$, then:

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

Division instead of subtraction as entropy is defined on logarithms.

Definition: Chain Rule For Entropy

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

Remark: Note that $H(Y|X) \neq H(X|Y)$. However,
 $H(X) - H(X|Y) = H(Y) - H(Y|X),$

Mutual Information

Definition Consider two random variables X and Y with a joint probability mass function $p(x, y)$ and marginal probability mass functions $p(x)$ and $p(y)$. The *mutual information* $I(X; Y)$ is the relative entropy between

the joint distribution and the product distribution $p(x)p(y)$:

$$I(X; Y) = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x, y) \log \frac{p(x, y)}{p(x)p(y)}$$

– What happens to information transferred across channel

$$\begin{aligned} \underbrace{I(X; Y)}_{\text{av. conveyed information}} &= \underbrace{H(X)}_{\text{source entropy}} - \underbrace{H(X|Y)}_{\text{av. information lost}} \\ \underbrace{I(X; Y)}_{\text{av. conveyed information}} &= \underbrace{H(Y)}_{\text{destination entropy}} - \underbrace{H(Y|X)}_{\text{error entropy}} \end{aligned}$$

The mutual information between two variables is 0 if and only if the two variables are statistically independent.

Find: 3 - Conditional Entropy. ($H(y|x)$, $H(x|y)$)

4- Mutual information. ($I(x;y)$)

solution

Conditional Entropy. ($H(y|x)$, $H(x|y)$)

$P(x) = [0.25 \ 0.4 \ 0.15 \ 0.15 \ 0.05]$ من الفرع السابق

$$p(y|x) = \frac{p(x,y)}{p(x)}$$

$P(x)$				
0.25	1	0	0	0
0.4	1/4	3/4	0	0
0.15	0	1/3	2/3	0
0.15	0	0	1/3	2/3
0.05	0	0	1	0

Example

نفس المثال السابق

	y1	y2	y3	y4
x1	0.25	0	0	0
x2	0.1	0.3	0	0
x3	0	0.05	0.1	0
x4	0	0	0.05	0.1
x5	0	0	0.05	0

$$p(y1|x1) = \frac{p(x1,y1)}{p(x1)} = \frac{0.25}{0.25} = 1$$

$$p(y3|x3) = \frac{p(x3,y3)}{p(x3)} = \frac{0.1}{0.15}$$

Conditional Entropy

→ $H(y|x) = - \sum_{i=1}^5 \sum_{j=1}^4 p(x_i, y_j) \log_2 p(y_j|x_i)$ OR $H(y|x) = H(x, y) - H(x)$

$$= - [0.25 \log_2 1 + 0.1 \log_2 1/4 + 0.3 \log_2 3/4 + 0.05 \log_2 1/3 + 0.1 \log_2 2/3 + 0.05 \log_2 1/3 + 0.1 \log_2 2/3 + 0.05 \log_2 1]$$

= 0.6 bits/symbol

$P(y) = [0.35 \ 0.35 \ 0.2 \ 0.1]$ من الفرع السابق

$$p(x|y) = \frac{p(x,y)}{p(y)}$$

	5/7	0	0	0
	2/7	6/7	0	0
	0	1/7	1/2	0
	0	0	1/4	1
	0	0	1/4	0

$$p(x1|y1) = \frac{p(x1,y1)}{p(y1)} = \frac{0.25}{0.35}$$

$$p(x2|y1) = \frac{p(x2,y1)}{p(y1)} = \frac{0.1}{0.35}$$

$$p(x2|y2) = \frac{p(x2,y2)}{p(y2)} = \frac{0.03}{0.35}$$

$$p(x5|y3) = \frac{p(x5,y3)}{p(y3)} = \frac{0.05}{0.2}$$

→ $H(x|y) = - \sum_{i=1}^5 \sum_{j=1}^4 p(x_i, y_j) \log_2 p(x_i|y_j)$ OR $H(x|y) = H(x, y) - H(y)$

$$= - [0.25 \log_2 5/7 + 0.1 \log_2 2/7 + 0.3 \log_2 6/7 + 0.05 \log_2 1/7 + 0.1 \log_2 1/2 + 0.05 \log_2 1/4 + 0.1 \log_2 1 + 0.05 \log_2 1/4]$$

= 0.809 bits/symbol

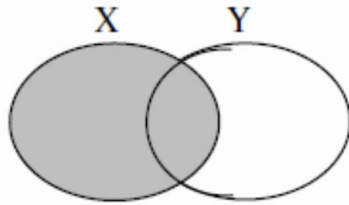
- **Mutual Entropy (Mutual Information)**

$$I(x;y) = H(x) - H(x|y)$$

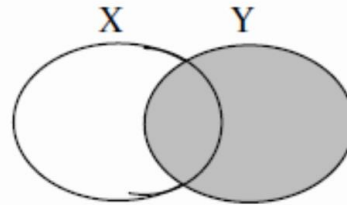
$$= 2.066 - 0.809$$

$$I(x;y) = 1.157 \text{ bits/symbol}$$

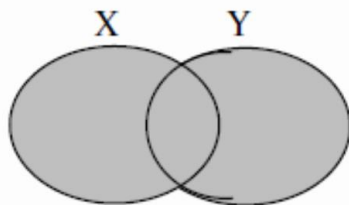
Venn Diagram of Representation of the channel.



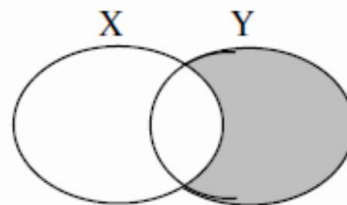
$H(X)$
source entropy



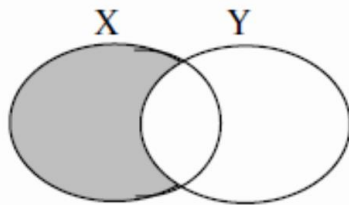
$H(Y)$
destination entropy



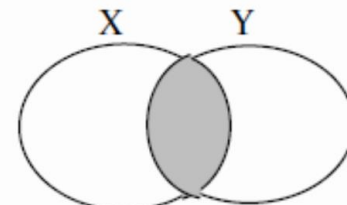
$H(X,Y)$
system entropy



$H(Y/X)$
error entropy



$H(X/Y)$
information lost



$I(X,Y)$
conveyed information

المحاضرة 5

ملاحظة :

$H(y|x)$

تمثل noise

entropy

اي مقدار

noise التي

يولدها

channel

و

$H(x|y)$

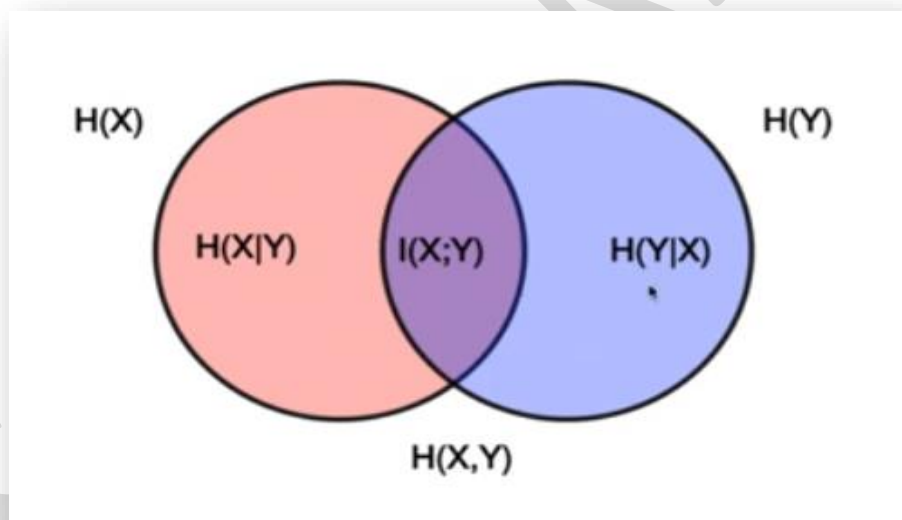
تمثل loss

entropy

اي الخسائر

information

اثناء النقل .



- **X and Y are independent:** It's safe then to say that X and Y don't have a mutual information, because $H(X, Y) = H(X) + H(Y)$

- **X and Y are dependent:** since $H(X, Y) < H(X) + H(Y)$, then there's a sort of redundant information that is counted once we're computing their information together, this information is called mutual information and yes it's the difference between the sum of information and the information of both. We can also interpret it in different way depending on each of the following equalities:

$$\begin{aligned} I(X; Y) &= H(X) + H(Y) - H(X, Y) \\ &= H(X) + H(Y) - H(Y | X) - H(X) \\ &= H(X) + H(Y) - H(X | Y) - H(Y) \\ &= H(Y) - H(Y | X) = H(X) - H(X | Y) \\ &= H(X, Y) - H(Y | X) - H(X | Y) \\ &= \sum_x \sum_y p(x, y) \log_2 \left(\frac{p(x, y)}{p(x)p(y)} \right) \end{aligned}$$

Communication Channel & Channel Capacity

Outline:

Definition of Communication Channel

Discrete Memoryless Channel (DMC)

Special Channels

- Lossless channel
- Deterministic channel
- Noiseless Channel (Ideal channel)

Symmetric Channels

- Binary Symmetric Channel (BSC)
- Ternary Symmetric Channel (TSC)

A Communication Channel: is the path or medium through which the symbols flow to the receiver.

قناة الاتصال هي المسار أو الوسيط الذي تتدفق من خلاله الرموز إلى المستقبل.

A. Discrete Memoryless Channels

A discrete memoryless channel (DMC) is a statistical model with an input X and an output Y (Fig. 1). During each unit of the time (signaling interval), the channel accepts an input symbol from X , and in response it generates an output symbol from Y . The channel is "discrete" when the alphabets of X and Y are both finite. It is

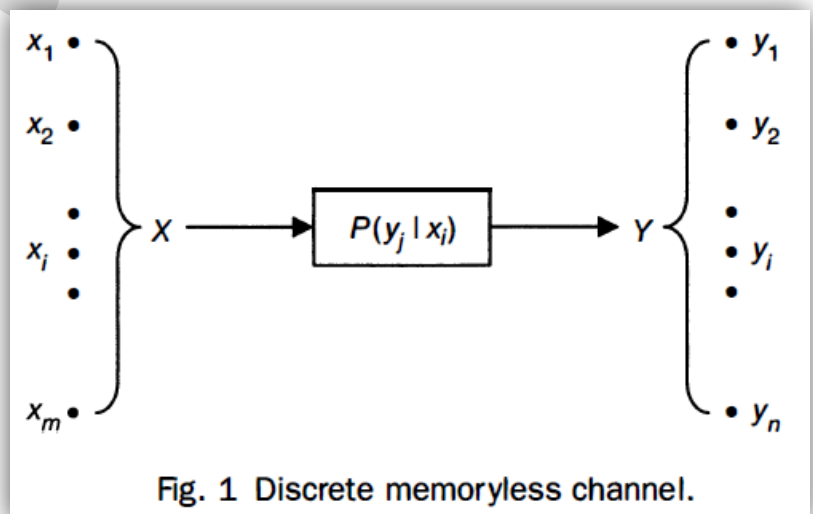


Fig. 1 Discrete memoryless channel.

المحاضرة 6.

"**memoryless**" when the *current output depends* on only the current *input and not on any of the previous inputs*.

(DMC) **Discrete Memoryless Channel**: هي نموذج إحصائي بمدخل X ومخرج Y (الشكل 1).

تقبل القناة رمز إدخال من X وتنشئ رمز إخراج من Y خلال كل وحدة من الزمن (الفاصل الزمني للإشارة). وتكون القناة "discrete" إذ أن X و Y محدودة. وهي "بلا ذاكرة" حيث أن الإخراج الحالي يعتمد على الإدخال الحالي فقط وليس على أي من المدخلات السابقة.

A diagram of a DMC with m inputs and n outputs is illustrated in Fig. 1. The input X consists of input symbols x_1, x_2, \dots, x_m . The a priori probabilities of these source symbols $P(x)$ are assumed to be known. The output Y consists of output symbols y_1, y_2, \dots, y_n . *Each possible input-to-output path is indicated along with a conditional probability $P(y_j/x_i)$, where $P(y_j | x_i)$ is the **conditional probability** of obtaining output y_j given that the input is x_i , and is called **a channel transition probability**.*

B. Channel Matrix:

A channel is completely specified by the complete set of transition probabilities. Accordingly, the channel of Fig.1 is often specified by **the matrix of transition probabilities $[P(Y | X)]$** , given by prob.1.

$$[P(Y | X)] = \begin{bmatrix} P(y_1 | x_1) & P(y_2 | x_1) & \dots & P(y_n | x_1) \\ P(y_1 | x_2) & P(y_2 | x_2) & \dots & P(y_n | x_2) \\ \dots & \dots & \dots & \dots \\ P(y_1 | x_m) & P(y_2 | x_m) & \dots & P(y_n | x_m) \end{bmatrix}$$

The matrix $[P(Y | X)]$ is called the **channel matrix**.

Since each input to the channel results in some output, each row of the channel matrix must sum to unity; that is,

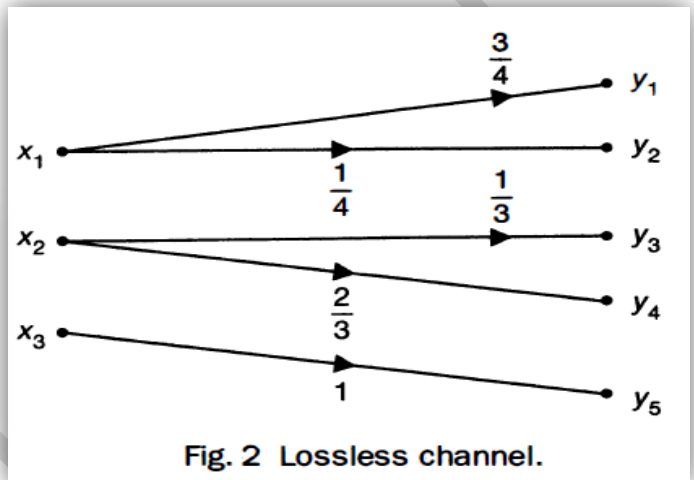
$$\sum_{j=1}^n P(y_j | x_i) = 1 \quad \text{for all } i \quad (1)$$

C. Special Channels (different Types of Channels):

1. Lossless Channel:

A channel described by a channel matrix with only **one nonzero** element in each **column** is called a lossless channel. An example of a lossless channel is shown in Fig.2, and the corresponding channel matrix is shown in prob. (2).

$$[P(Y|X)] = \begin{bmatrix} \frac{3}{4} & \frac{1}{4} & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{3} & \frac{2}{3} & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad (2)$$



For a lossless channel, $H(X|Y) = 0$ (Prob. 1) and

$$I(X; Y) = H(X)$$

Thus, the mutual information (information transfer) is equal to the input (source) entropy, and no source information is lost in transmission. Consequently, the channel capacity per symbol is

$$C_s = \max_{\{P(x_i)\}} H(X) = \log_2 m$$

2. Deterministic Channel:

A channel described by a channel matrix with **only one nonzero element** in each **row** is called a deterministic channel. An example of a deterministic channel is shown in Fig. 3, and the corresponding channel matrix is shown in prob. (3).

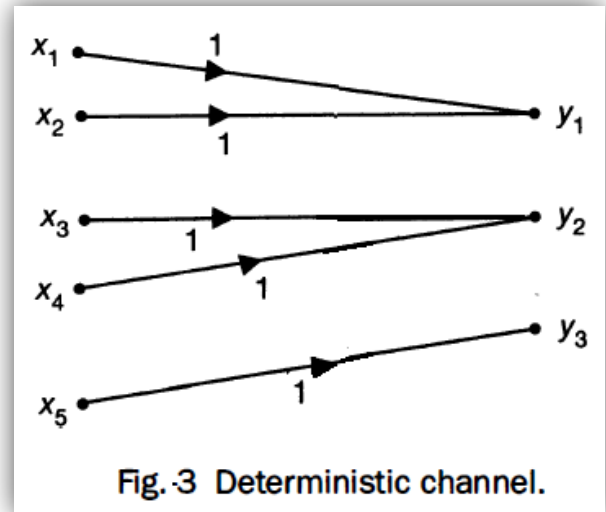


Fig.-3 Deterministic channel.

Note that since each **row** has only **one nonzero element**, this element must be unity by prob. (1). Thus, when a given source symbol is sent in the deterministic channel, it is clear which output symbol will be received.

$$[P(Y|X)] = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (3)$$

For a deterministic channel, $H(Y|X) = 0$ for all input distributions $P(x_i)$, and

$$I(X; Y) = H(Y)$$

Thus, the information transfer is equal to the output entropy. The **channel capacity** per symbol is

$$C_s = \max_{\{P(x_i)\}} H(Y) = \log_2 n$$

where n is the number of symbols in Y .

3. Noiseless (ideal) Channel:

A channel is called noiseless if it is **both lossless and deterministic**. A noiseless channel is shown in Fig. 4.

The channel matrix has only **one element** in each **row** and in each **column**, and this element is **unity**. Note that the input and output alphabets are of the **same size**; that is, **$m = n$** for the noiseless channel.

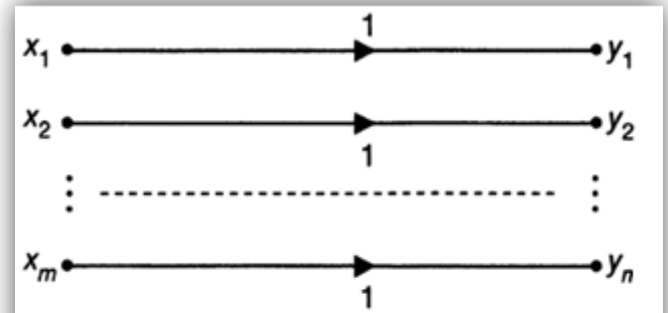


Fig. 4 Noiseless channel.

$$[P(Y|X)] = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (4)$$

Since a noiseless channel is both lossless and deterministic, we have

$$I(X; Y) = H(X) = H(Y)$$

and the **channel capacity** per symbol is

$$C_s = \log_2 m = \log_2 n$$

D. Symmetric Channel:

A communication channel is considered as symmetric only when it satisfies the following two conditions:

- 1- **Equal** number of symbols in **X** & **Y**, i.e. the conditional probability **P(Y|X)** is a **square** matrix.
- 2- **Each row** in the conditional probability matrix **P(Y|X)** comes from other row in the matrix after **changing** (shift) **the positions of its elements**.

$$P(Y|X) = \begin{bmatrix} 0.9 & 0.05 & 0.05 \\ 0.05 & 0.9 & 0.05 \\ 0.05 & 0.05 & 0.9 \end{bmatrix}$$

1. Binary Symmetric Channel:

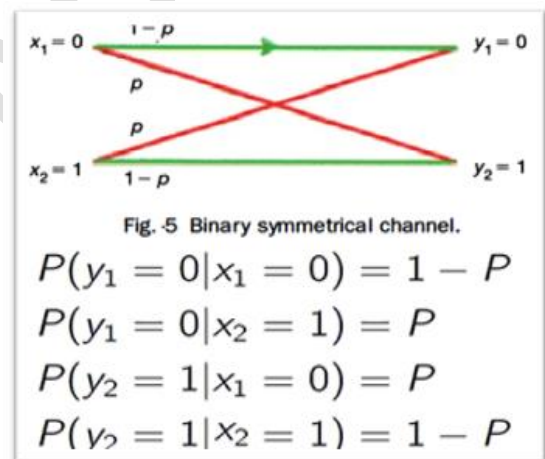
The binary symmetric channel (BSC) is defined by the channel diagram shown in Fig.5, and its channel matrix is given by prob. (5).

$$[P(Y|X)] = \begin{bmatrix} 1-p & p \\ p & 1-p \end{bmatrix} \quad (5)$$

The channel has two inputs ($x_1 = 0, x_2 = 1$) and two outputs ($y_1 = 0, y_2 = 1$).

The *channel is symmetric because*

the probability of receiving a 1 if a 0 is sent is the same as the probability of receiving a 0 if a 1 is sent. This common transition probability is denoted by p .



For the BSC of Fig. 5, the mutual information is (Prob. 5.)

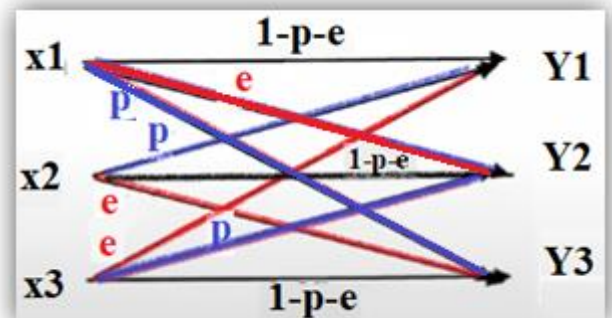
$$I(X; Y) = H(Y) + p \log_2 p + (1 - p) \log_2(1 - p)$$

and the channel capacity per symbol is

$$C_s = 1 + p \log_2 p + (1 - p) \log_2(1 - p)$$

2. Ternary Symmetric Channel (TSC) :

In this model, the number of symbols in X & Y are equal to 3. Therefore, transition probability diagram of TSC is given by



$$P(Y|X) = \begin{bmatrix} 1-p-e & e & p \\ p & 1-p-e & e \\ e & p & 1-p-e \end{bmatrix}$$

Communication Channel & Channel Capacity

Exercises

- **EX.1** : Find the channel capacity for BSC shown then draw the channel.

$$P(Y \setminus X) = \begin{bmatrix} 0.7 & 0.3 \\ 0.3 & 0.7 \end{bmatrix}$$

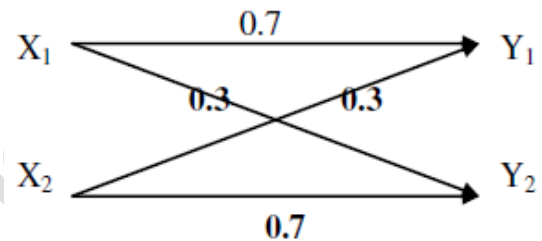
Sol.

$$C_s = 1 + \underbrace{p \log_2 p + (1 - p) \log_2 (1 - p)}_K$$

ملاحظة : قيمة K تحسب لقيم الصف الأول من مصفوفة الاحتمالية

$$K = 0.7 \log_2 0.7 + 0.3 \log_2 0.3 = -0.8823$$

$$C = 1 + (-0.8823) = 0.118 \text{ bit / symbol}$$



- **EX.2** : Binary Symmetric Channel (BSC) have $p(y1|x1)=0.7$, Draw channel model.

Sol.:

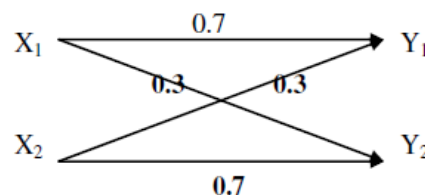
$$\sum_{j=1}^n P(y_j | x_i) = 1 \text{ for all } i \quad (1)$$

$$p(y2|x1) = 1 - p(y1|x1) = 1 - 0.7 = 0.3$$

Symmetric Channel:

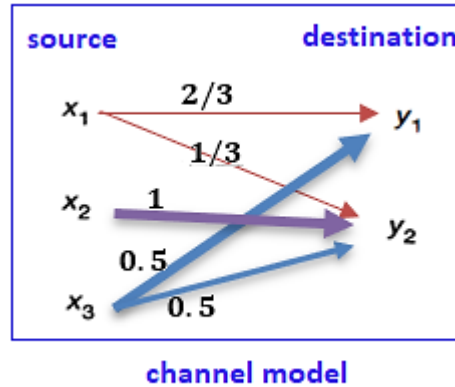
- 1- Equal number of symbols in X & Y, i.e. the conditional probability $P(Y|X)$ is a square matrix (2*2).
- 2- Each row in the conditional probability matrix $P(Y|X)$ comes from other row in the matrix after changing (shift) the positions of its elements.

$$P(Y \setminus X) = \begin{bmatrix} 0.7 & 0.3 \\ 0.3 & 0.7 \end{bmatrix}$$

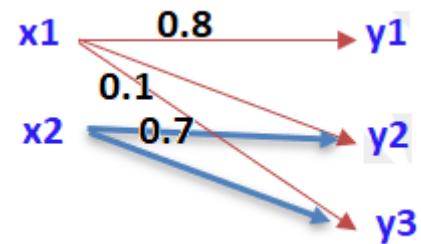


- **EX.3** : Draw channel model if you have the following $p(y|x) =$
- | | |
|---------------|---------------|
| $\frac{2}{3}$ | $\frac{1}{3}$ |
| 0 | 1 |
| 0.5 | 0.5 |

Sol.



Ex. 4: Suppose The Channel Model In The Following Figure, $P(X1)=0.6$, Find $H(X)$, $H(Y)$, $H(X,Y)$, Noise And Loss Entropy.



Sol. :

From the figure we find $p(y|x)$.

$$p(y|x) = \begin{matrix} & y1 & y2 & y3 \\ \begin{matrix} x1 \\ x2 \end{matrix} & \begin{matrix} 0.8 \\ 0 \end{matrix} & \begin{matrix} 0.1 \\ 0.7 \end{matrix} & \begin{matrix} 0.1 \\ 0.3 \end{matrix} \end{matrix}$$

Because $\sum p(x)=1$ and $p(x1) = 0.6$ then $p(x2) = 1-0.6=0.4$

$$P(x) = \begin{matrix} x1 & x2 \end{matrix} \begin{matrix} 0.6 & 0.4 \end{matrix}$$

$$p(y|x) = \frac{p(x,y)}{p(x)}, \quad p(x,y) = p(y|x) * p(x)$$

$$p(x,y) = \begin{matrix} & y1 & y2 & y3 \\ \begin{matrix} x1 \\ x2 \end{matrix} & \begin{matrix} 0.8 \\ 0 \end{matrix} & \begin{matrix} 0.1 \\ 0.7 \end{matrix} & \begin{matrix} 0.1 \\ 0.3 \end{matrix} \end{matrix} * \begin{matrix} x1 & x2 \end{matrix} \begin{matrix} 0.6 & 0.4 \end{matrix}$$

$$p(x,y) = \begin{matrix} & y1 & y2 & y3 \\ \begin{matrix} x1 \\ x2 \end{matrix} & \begin{matrix} 0.48 \\ 0 \end{matrix} & \begin{matrix} 0.06 \\ 0.28 \end{matrix} & \begin{matrix} 0.06 \\ 0.12 \end{matrix} \end{matrix}$$

$$P(y_i) = \sum_{i=1}^n p(x_i, y_j)$$

$$P(y) = [0.48 \quad 0.34 \quad 0.18]$$

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

Marginal Entropy

$$H(x) = -[0.6 \log_2 0.6 + 0.4 \log_2 0.4] = 0.97 \text{ bit/symbol}$$

$$H(y) = -[0.48 \log_2 0.48 + 0.34 \log_2 0.34 + 0.18 \log_2 0.18] = 1.48 \text{ bit/symbol}$$

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

$$H(x, y) = -[0.48 \log_2 0.48 + 0.06 \log_2 0.06 + 0.06 \log_2 0.06 + 0.28 \log_2 0.28 + 0.12 \log_2 0.12] = 1.876 \text{ bit/symbol}$$

Noise Entropy

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

or

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

$$H(y|x) = 1.876 - 0.97 = 0.906 \text{ bit/symbol}$$

Loss Entropy

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

$$H(x|y) = 1.876 - 1.48 = 0.39 \text{ bit/symbol}$$

Ex. 5 : A source produces dots "." & dashes "-" with probability $P(\text{dot}) = 0.65$, if time duration of a dot is 200 ms and that for a dash is 800 ms. Find the average source entropy $R(X)$.

Sol:

$$P(\text{dot}) = 0.65 \quad P(\text{dash}) = 1 - P(\text{dot}) = 1 - 0.65 = 0.35$$

$$\tau_{\text{dot}} = 200 \text{ ms}, \quad \tau_{\text{dash}} = 800 \text{ ms}$$

$$R(X) = \frac{H(X)}{\tau'}$$

$$\tau' = \sum_{i=1}^2 \tau_i \cdot P(X_i)$$

$$H(X) = \sum p(x) \cdot \log p(x) = - [0.65 \log_2(0.65) + 0.35 \log_2(0.35)] = 0.934 \text{ bit/symbol}$$

$$\tau' = [200 \cdot 0.65 + 800 \cdot 0.35] = 410 \text{ ms}$$

$$R(X) = \frac{0.934}{410 \cdot 10^{-3}} = 2.278 \cdot 10^3 \text{ bit/sec}$$

Ex. 6: The joint prob. is given by $P(X_i Y_j) = \begin{bmatrix} 0.5 & 0.25 \\ 0 & 0.125 \\ 0.0625 & 0.0625 \end{bmatrix}$

- Find :**
1. Marginal entropies
 2. System Entropies
 3. Noise and losses entropies
 4. Transinformation
 5. Draw the channel model

Sol:

$$1. \quad P(X) = \sum_{j=1}^3 P(X_i, Y_j) = [0.75 \quad 0.125 \quad 0.125]$$

$$P(Y) = \sum_{i=1}^3 P(X_i, Y_j) = [0.5625 \quad 0.4375]$$

$$H(X) = - \sum_{i=1}^3 P(X_i) \cdot \log_2 P(X_i) = - [0.75 \log_2(0.75) + 2 \cdot 0.125 \log_2(0.125)]$$

$$= 1.06127 \text{ bits/symbol}$$

$$H(Y) = - \sum_{j=1}^2 P(Y_j) \cdot \log_2 P(Y_j) = - [0.5625 \log_2(0.5625) + 0.4375 \log_2(0.4375)]$$

$$= 0.9887 \text{ bits/symbol}$$

$$2. H(X,Y) = - \sum_{i=1}^2 \sum_{j=1}^3 P(X_i, Y_j) \cdot \log_2 P(X_i, Y_j) = - [0.5 \log_2(0.5) + 0.25 \log_2(0.25) + 0.125 \log_2(0.125) + 2 \cdot 0.0625 \log_2(0.0625)] = 1.875 \text{ bits/symbols}$$

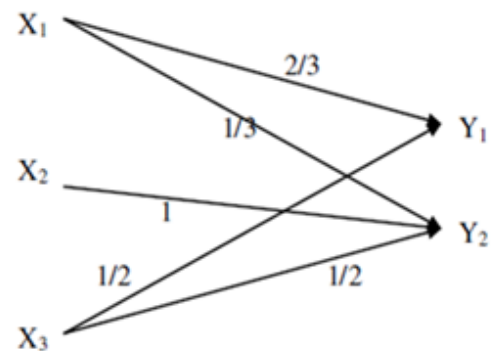
$$3. H(Y|X) = H(X,Y) - H(X) = 1.875 - 1.06127 = 0.81373 \text{ bit/symbol.}$$

$$H(X|Y) = H(X,Y) - H(Y) = 1.875 - 0.9887 = 0.8863 \text{ bit/symbol}$$

$$4. I(X,Y) = H(X) - H(X|Y) = 0.17497 \text{ bits/symbol}$$

$$5. \text{ To draw a channel, we find } P(Y_j|X_i) = \frac{P(X_i, Y_j)}{P(X_i)}$$

$$P(Y_j|X_i) = \begin{pmatrix} \frac{0.5}{0.75} & \frac{0.25}{0.75} \\ \frac{0}{0.125} & \frac{0.125}{0.125} \\ \frac{0.0625}{0.125} & \frac{0.0625}{0.125} \end{pmatrix} = \begin{pmatrix} \frac{2}{3} & \frac{1}{3} \\ 0 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$$



Source Coding

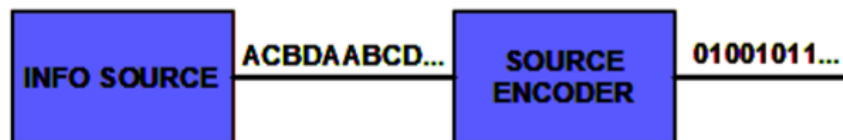
(Entropy Encoding)

An important problem in communications is the efficient representation of data generated by a discrete source. The process by which this representation is accomplished is called source encoding.

The device that performs the representation is called a **source encoder**. Our primary interest is in the development of an efficient source encoder that satisfies two functional requirements:

1. The code words produced by the encoder are in **binary form**.
2. The source code is uniquely decodable, so that the original source sequence can be reconstructed perfectly from the encoded binary sequence.

Source coding does not change or alter the source entropy, i.e. the average number of information bits per source symbol. In this sense source entropy is a fundamental property of the source.



Source coding techniques:

- 1- Fixed length Code
- 2- Variable length Code

1- Fixed length Code

A fixed length code is a kind of channel coding that was used in earlier communication systems. The idea is simple: let's suppose you want to encode an alphabet $\{A_1, A_2, A_3, \dots, A_k\}$. A fixed length code C is a set of code words $\{C_1, \dots, C_k\}$, all of the same number of bits, each of which encodes a specific element of the alphabet. if the alphabet of messages is $\{A, B, C, D, E\}$ then 3 bits suffice to represent these five elements, and you can choose the

$A = 000 \ B = 001 \ C = 010 \ D = 011 \ E = 100$

Example: (The abcd-file).

Consider file with symbols 'a', 'b', 'c' and 'd':

bdacddaadaabbcaabbaaabaacbdab.....

and suppose that

$$P(a) = 1/2$$

$$P(b) = 1/4$$

$$P(c) = 1/8$$

$$P(d) = 1/8$$

This we want to code (binary)

As all code words $\in \{00,01,10,11\}$ consist of two bits:

(average) # bits per symbol = 2

a \leftrightarrow 00

b \leftrightarrow 01

c \leftrightarrow 10

d \leftrightarrow 11

code word

For example:

Coding: "ababc" \longrightarrow 00**0**100**0**1**1**0

Decoding:

This coding is decodable : As all code words $\in \{00,01,10,11\}$ consist of two bits:

(average) # bits per symbol = 2

0001000110 \longrightarrow ababc

2- Variable length Code

- A. Shannon Fano Coding (Claude Shannon and Robert Fano)
- B. Huffman Coding
- C. Arithmetic Coding

A- Shannon Fano Coding

- Shannon Fano Algorithm is an entropy encoding technique for lossless data compression of multimedia. Named after Claude Shannon and Robert Fano, it assigns a code to each symbol based on their probabilities of occurrence. It is a variable length encoding scheme, that is, the codes assigned to the symbols will be of varying length.
- It is A top - down approach.

HOW DOES IT WORK?

The steps of the algorithm are as follows:

1. Create a list of probabilities or frequency counts for the given set of symbols so that the relative frequency of occurrence of each symbol is known.
2. Sort the list of symbols in decreasing order of probability, the most probable ones to the left and least probable to the right.
3. Split the list into two parts, each with approximately same number of counts, i.e. split in two so as **to minimize difference** in **total probability** or **counts** .
4. Assign the value 0 to the left part and 1 to the right part.
5. Repeat the steps 3 and 4 for each part, until all the symbols are split into individual subgroups.

<https://youtu.be/yNUdPSsgUHY> : رابط المحاضرة 8 :

Example1: For the given set of symbols

“E B E C DED A EE C C EEE A D B E B D A EEE C B D C C A D B C E B E A E “

- 1- Construct the Code word for each symbol by using Shannon-Fano technique.
- 2- find size of transited message.

1- Solution:

Step 1 : Count symbols in stream:

Symbol	A	B	C	D	E	Total
Count	5	6	7	6	15	39

Step 2 :

Sort the list of symbols in decreasing order.

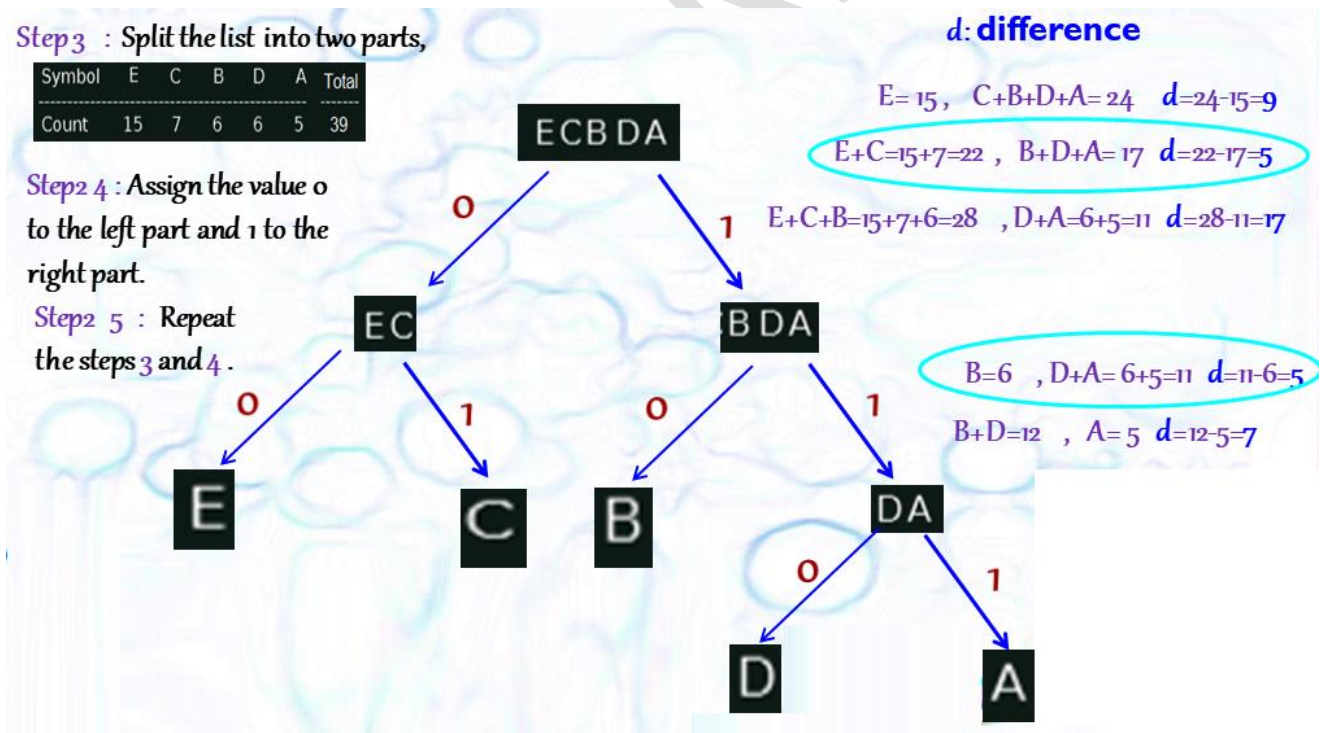
Symbol	E	C	B	D	A	Total
Count	15	7	6	6	5	39

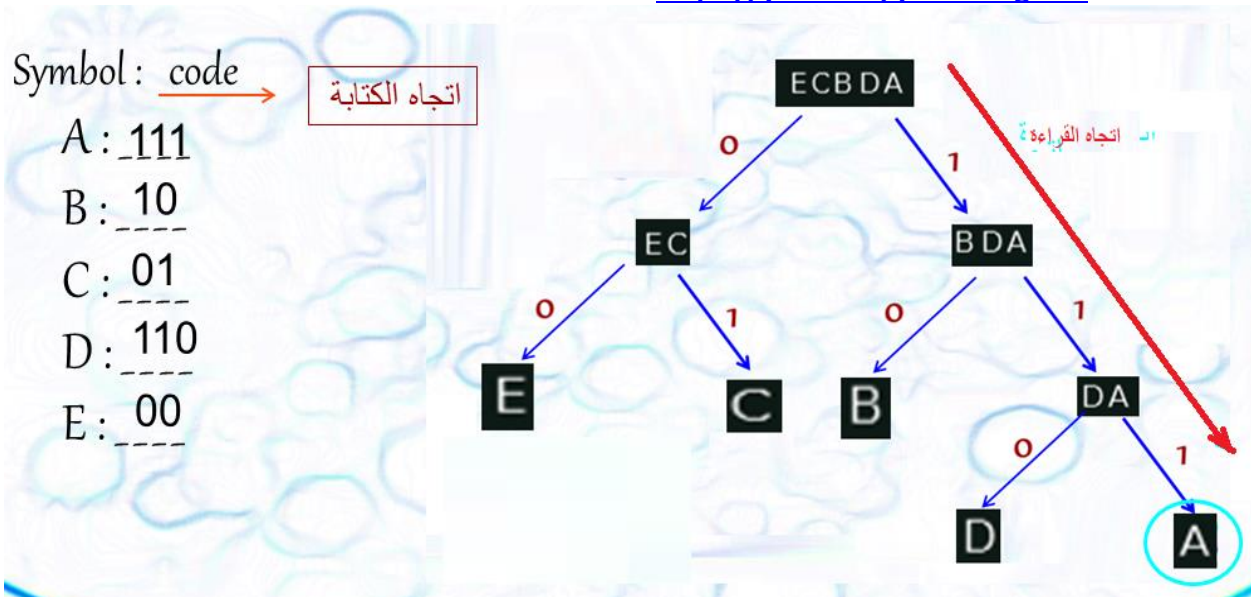
Step 3 : Split the list into two parts,

Symbol	E	C	B	D	A	Total
Count	15	7	6	6	5	39

Step 4 : Assign the value 0 to the left part and 1 to the right part.

Step 5 : Repeat the steps 3 and 4.





“EBECDEDAEECCEEEADBEBDAAEEECBDCCADBCEBEAE”

2- find size of transited message

2- Solution:

Symbol	Count	Code	#of bits
E	15	00	30
C	7	01	14
B	6	10	12
D	6	110	18
A	5	111	15
TOTAL (# of bits): 89			

size of transited message (coded data stream) = 89 bits

size of transited message (Raw token 8 bits per (39 chars) = 312 bits

Example1:

Given task is to construct Shannon codes for the given set of symbols using the Shannon-Fano lossless compression technique.

SYMBOL	A	B	C	D	E
PROBABILITY OR FREQUENCY	0.22	0.28	0.15	0.30	0.05

THE SYMBOLS AND THEIR PROBABILITY / FREQUENCY
ARE TAKEN AS INPUTS.
(In case of Frequency, the values can be any number)

Solution:

Let $P(x)$ be the probability of occurrence of symbol x :

Step 1& Step2: Sort the list of symbols in decreasing order of probability

SYMBOL	D	B	A	C	E
PROBABILITY OR FREQUENCY	0.30	0.28	0.22	0.15	0.05

INPUTS ARE SORTED ACCORDING
TO THEIR PROBABILITY / FREQUENCY
(Here they are sorted according to their probability)

Step 3:

1. Upon arranging the symbols in decreasing order of probability:

$$P(D) + P(B) = 0.30 + 0.2 = 0.58$$

and,

$$P(A) + P(C) + P(E) = 0.22 + 0.15 + 0.05 = 0.42$$

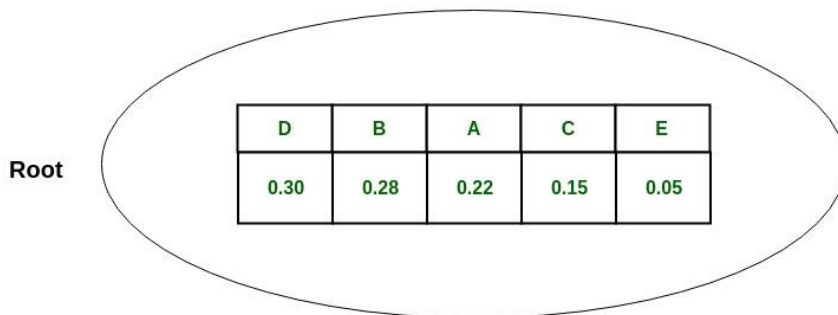
And since the almost equally split the table, the most is divid edit the block quote table is block quoten to

{D, B} and {A, C, E}

and assign them the values 0 and 1 respectively.

Step2:

Tree:



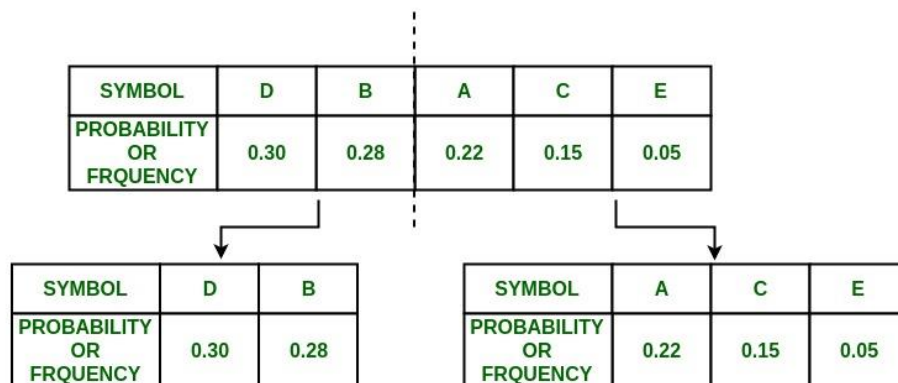
TREE AFTER STEP 2

2. Now, in {D, B} group,

P(D) = 0.30 and P(B) = 0.28

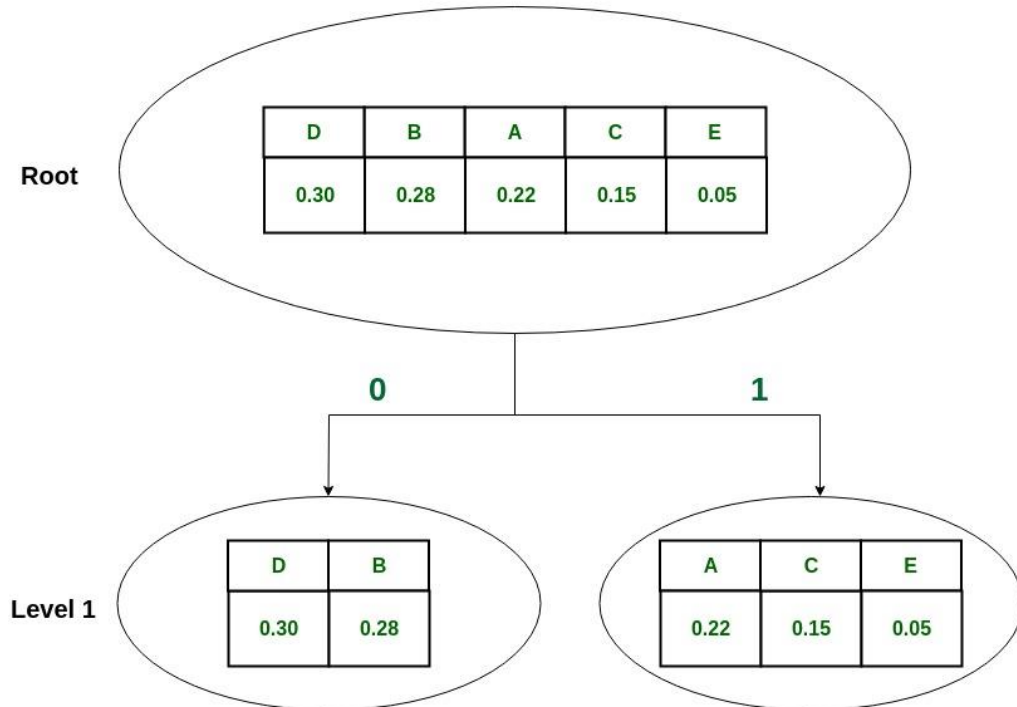
which means that **P(D) > P(B)**, so divide {D, B} into {D} and {B} and assign 0 to D and 1 to B.

Step3:



THE SYMBOLS ARE DIVIDED INTO TWO
 SUCH THAT THE TOTAL PROBABILITY / FREQUENCY
 OF LEFT SIDE ALMOST SAME AS THAT OF RIGHT SIDE

Tree:



TREE AFTER STEP 3

3. In {A, C, E} group,

$$P(A) = 0.22 \text{ and } P(C) + P(E) = 0.20$$

So the group is divided into

{A} and {C, E}

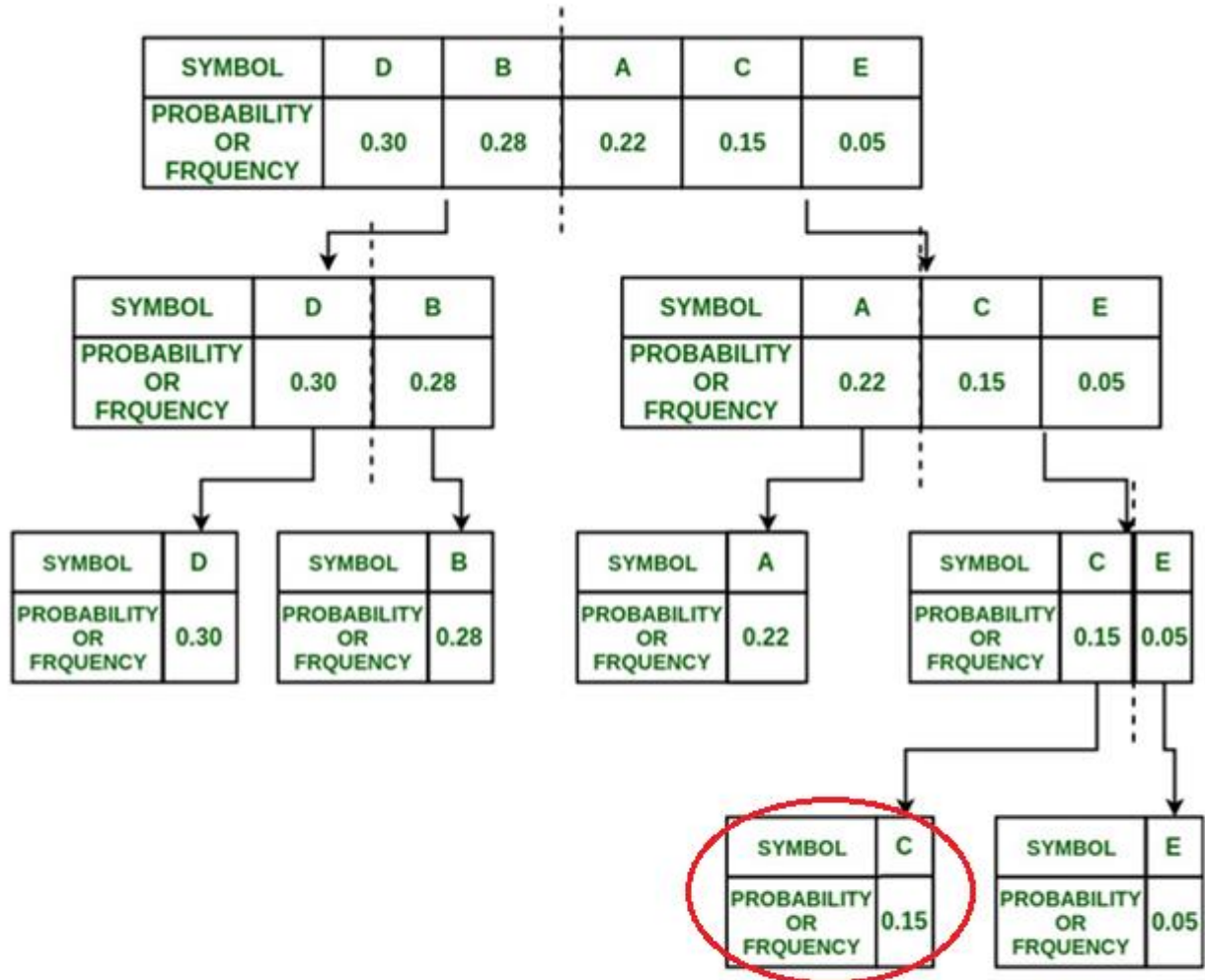
and they are assigned values 0 and 1 respectively.

4. In {C, E} group,

$$P(C) = 0.15 \text{ and } P(E) = 0.05$$

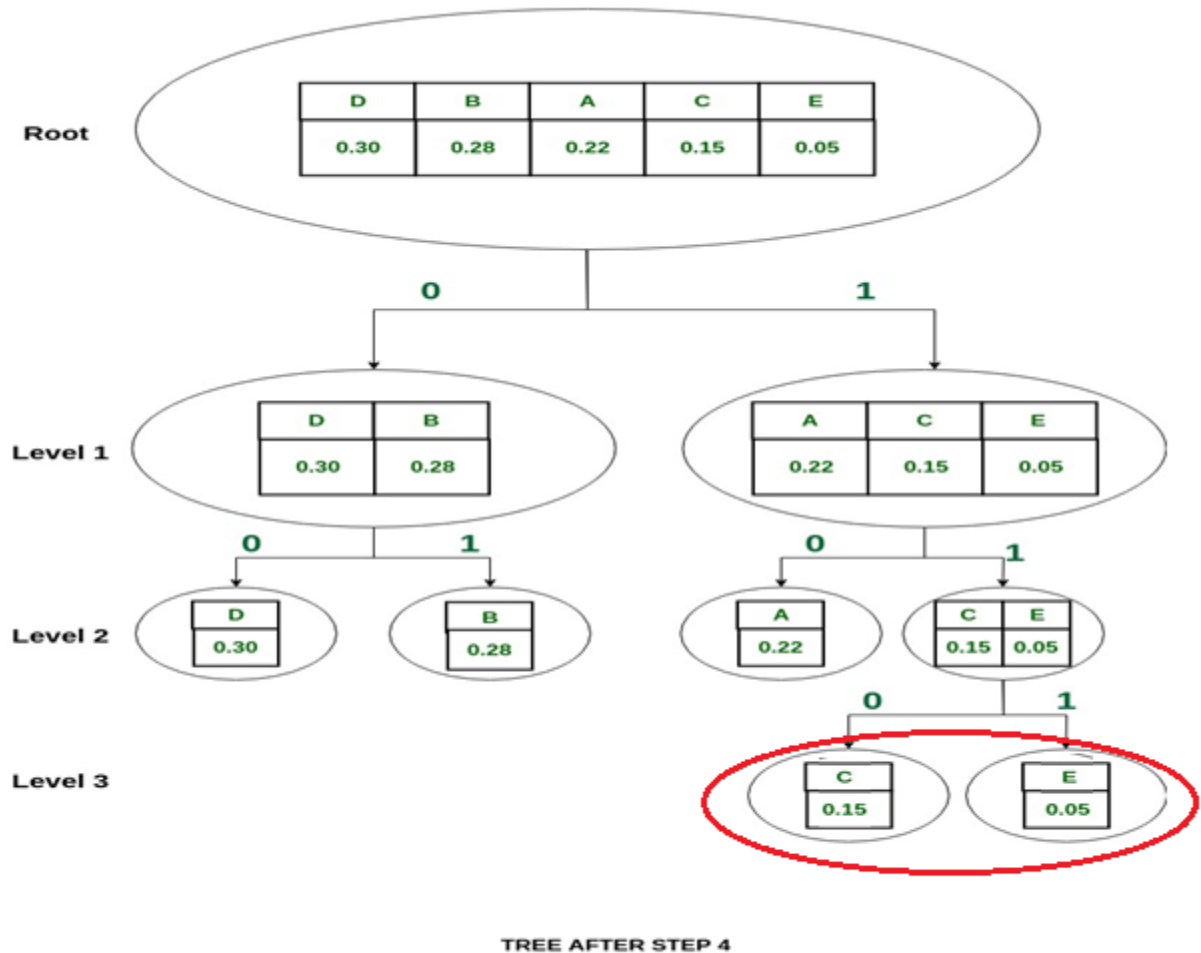
So divide them into {C} and {E} and assign 0 to {C} and 1 to {E}

Step4:



THE SYMBOLS ARE CONTINUED TO BE DIVIDED INTO TWO
 TILL EACH SYMBOL BECOME SEPARATED

Tree:



Note: The splitting is now stopped as each symbol is separated now.

The Shannon codes for the set of symbols are:

SYMBOL	D	B	A	C	E
PROBABILITY OR FREQUENCY	0.30	0.28	0.22	0.15	0.05
SHANNON-FANO CODE	00	01	10	110	111

Source Coding

(Entropy Encoding)

Variable length Code

- A. Shannon Fano Coding (Claude Shannon and Robert Fano)
- B. Huffman Coding
- C. Arithmetic Coding

B- Huffman coding:

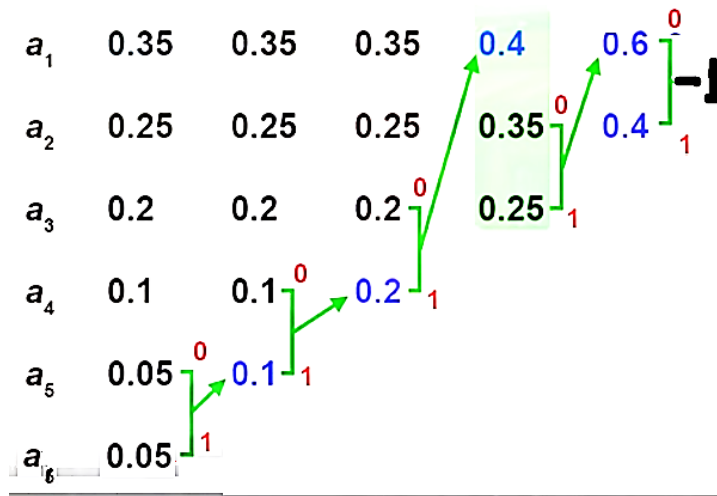
- Huffman coding is an effective method for lossless data compression, based on the principle of assigning shorter codes to more frequent symbols. It uses a bottom-up approach to build a binary tree representing the coding symbols.
- Huffman coding is characterized by its high efficiency, relative simplicity, and ability to provide lossless compression. However, it requires prior knowledge of the frequency of each symbol and the coding tree must be stored with the compressed data for decoding
- This method is widely used in various applications such as image, audio, video, and file compression, and is an essential component of many modern compression standards, examples: JPEG, MPEG, MP3..

Huffman Algorithm:

1. Initialization: put all nodes in a list, keep it sorted at all times
2. Find 2 symbols with the smallest probability/ frequency and then merge them to create a new "node" and treat it as a new symbol.
3. Repeat steps 1 and 2 until there is only 1 symbol left. At this point, we created a binary tree.
4. Assign the value 0 to the top part and 1 to the bottom part.

Example1 : If $P(a_3)=0.2$, $P(a_4)=0.1$, $P(a_5)=0.05$, $P(a_6)=0.05$, $P(a_1)=0.35$, $P(a_2)=0.25$. Using Huffman code to find Code word for each symbol.

Solution:



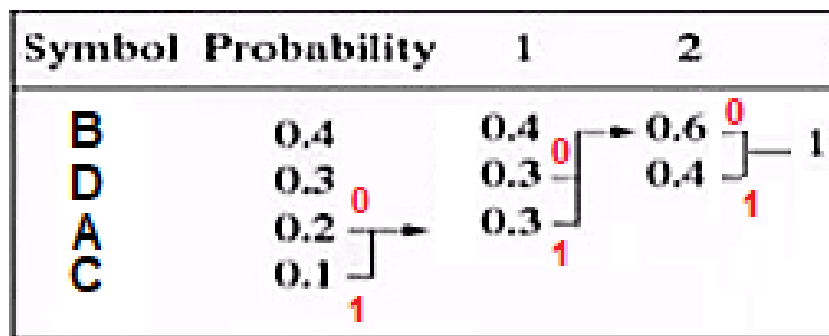
Huffman coding

a_1	0.35	00
a_2	0.25	01
a_3	0.2	10
a_4	0.1	110
a_5	0.05	1110
a_6	0.05	1111

<https://youtu.be/yNUdPSsgUHY> : رابط المحاضرة 9

Example2 : If you have message of 4 variables with length 100 symbols, Where the count of each variable is: A=20, B=40, C=10, D=30. Fill the following table by applying Huffman coding **algorithm**. [10 M]

Solution:



variable	Code word	number of variables in this message	Probability of variable
A	010	20	0.2
B	1	40	0.4
C	011	10	0.1
D	00	30	0.3

Example 3:

Let us assume we have the following text: "COMPUTER SCIENCE"

First, we calculate the frequency of each character:

- **C: 3**
- O: 1
- M: 1
- P: 1
- U: 1
- T: 1
- **E: 3**
- R: 1
- S: 1
- I: 1
- N: 1
- ' ':1 (for the space)

Total number of characters: 16

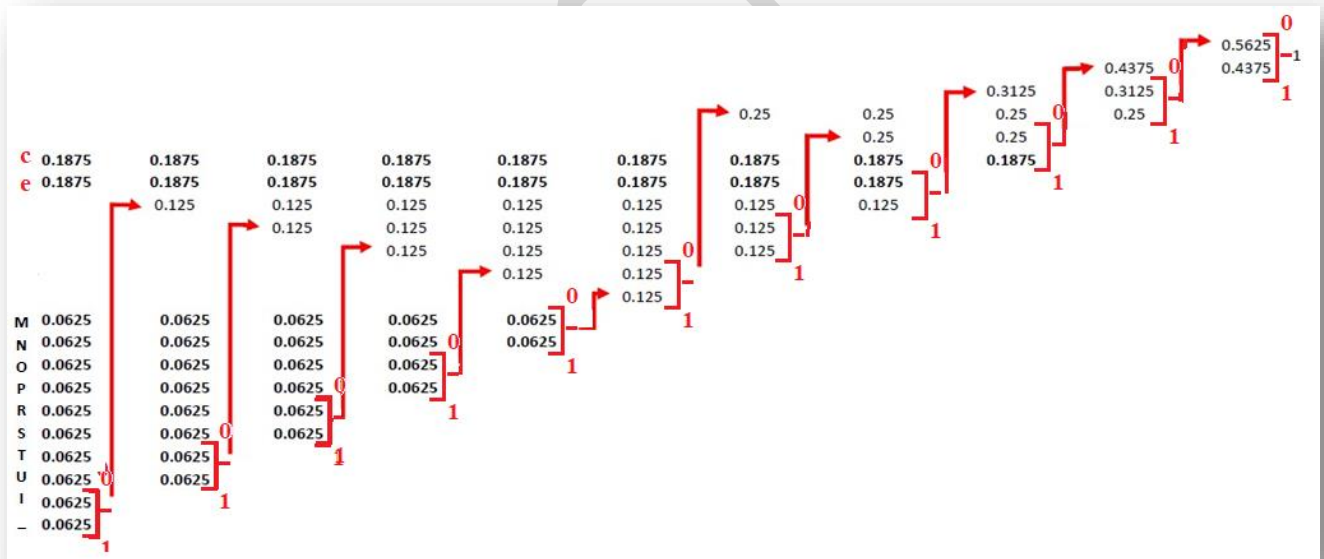
Probabilities:

- **$P(C) = 3/16 = 0.1875$**
- $P(O) = 1/16 = 0.0625$
- $P(M) = 1/16 = 0.0625$
- $P(P) = 1/16 = 0.0625$

- $P(U) = 1/16 = 0.0625$
- $P(T) = 1/16 = 0.0625$
- $P(E) = 3/16 = 0.1875$
- $P(R) = 1/16 = 0.0625$
- $P(S) = 1/16 = 0.0625$
- $P(I) = 1/16 = 0.0625$
- $P(N) = 1/16 = 0.0625$
- $P(' ') = 1/16 = 0.0625$ (for the space)

Applying Huffman Algorithm: We begin by merging the symbols with the lowest probabilities and continue until we obtain a complete tree.

ملاحظة : ترتب ابجديا في حالة تساوي الاحتمالات لتوحيد فك الشفرة



Final Huffman Code Outputs:

C : 11
e :000
m : 0110
n:0111
o:0100
p:0101
r :1010
s:1011
t :1000
u :1001
l :0010
' ': 0011 (المسافة)

"computer science"=

"11 0100 0110 1010 1001 1000 000 1010 0011 1011 11 0010 000 0111 11 000"

Compression Efficiency:

- Original message length (assuming 8 bits per ASCII character): $16 \times 8 = 128$ bits
- Compressed message length: $3 \times 2 + 3 \times 3 + 10 \times 4 = 55$ bits
- Compression ratio: $(128-55) \div 128 = 73 \div 128 = 0.57 \times 100 = 57\%$

Benefits of Huffman Coding:

1. **Compression Efficiency:** It provides a high compression ratio, especially with data having non-uniform distribution.
2. **Implementation Simplicity:** The algorithm is relatively easy to implement and understand.
3. **Lossless Compression:** It preserves all original information.

4. **Unique Decodability:** The encoded message can be decoded without ambiguity.
5. **Flexibility:** It can be applied to various types of data.

Entropy Calculation and Coding Efficiency:

Entropy $H(X)$ represents the theoretical minimum average code length required for source coding, and is calculated as follows:

$$H(X) = -\sum P(x_i) \times \log_2 P(x_i)$$

Where $P(x_i)$ is the probability of symbol x_i .

Example of Entropy Calculation:

For the previous example "COMPUTER SCIENCE":

$$H(X) = -(6 \times 0.1875 \times \log_2(0.1875) + 10 \times 0.0625 \times \log_2(0.0625)) = \text{--bits/symbol}$$

Arithmetic coding

- Arithmetic coding is a form of **entropy encoding** used in lossless data compression.
- Arithmetic coding **differs** from other forms of **entropy encoding**, such as **Huffman coding**, in that rather than separating the input into component symbols and replacing each with a code, arithmetic coding encodes the entire **message** into a **single number (message given a unique tag value)**, an arbitrary-precision fraction q where $0.0 \leq q < 1.0$. It represents the current information as a range, defined by two numbers.
- **Tag value** is calculated as **average** lower and upper limit.

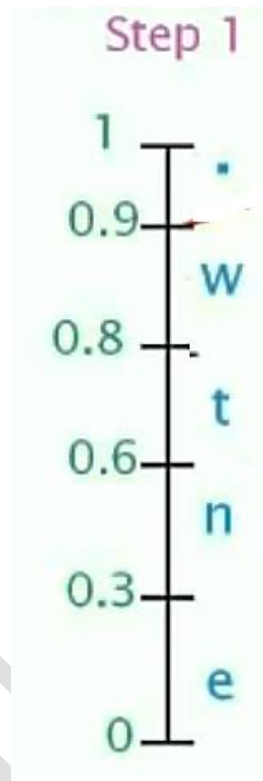
Arithmetic Coding Algorithm

The main steps of arithmetic coding are as follows:-

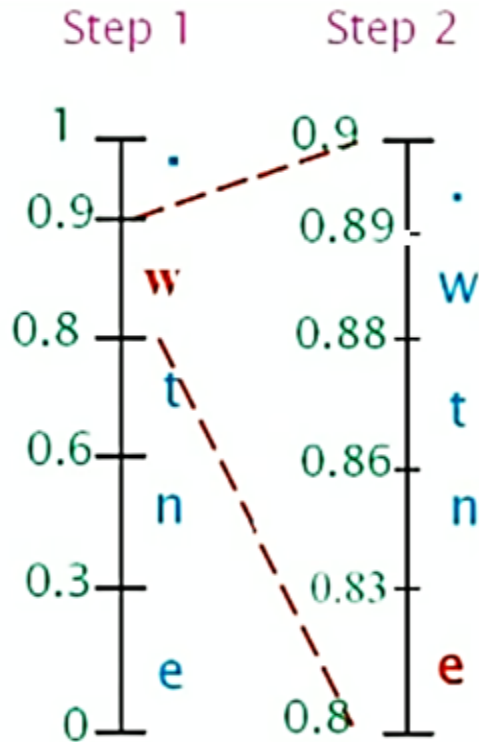
1. Defining two variables Low and High which define an interval [Low, High), i.e. Start by defining the “current interval” as [0,1).
2. Repeat the following two steps for each symbol in the input stream:
 - 2.1 Divide the current interval into subintervals whose sizes are proportional to the symbols probabilities.
 - 2.2 Select the subinterval corresponding to the symbol that actually occurs next in the file, and make it the new current interval.
3. When the entire input stream has been processed in this way, the output should be any number that uniquely identifies the current interval.

Example 1:

Consider the transmission of a message “went.” comprising a string of characters with probability $e \rightarrow 0.3$, $n \rightarrow 0.3$, $t \rightarrow 0.2$, $w \rightarrow 0.1$, $., ' \rightarrow 0.1$.



ملاحظة : يعتمد التسلسل الأبجدي (او قيمة الاحتمالية بحيث الأعلى تكون من الأسفل) على خط الاحتمالية من 0 إلى 1 ومن الأسفل إلى الأعلى.



Step2

$$d = \text{upper bound} - \text{lower bound} = 0.9 - 0.8 = 0.1$$

Range of 'symbol' = lower limit : upper limit

d: difference

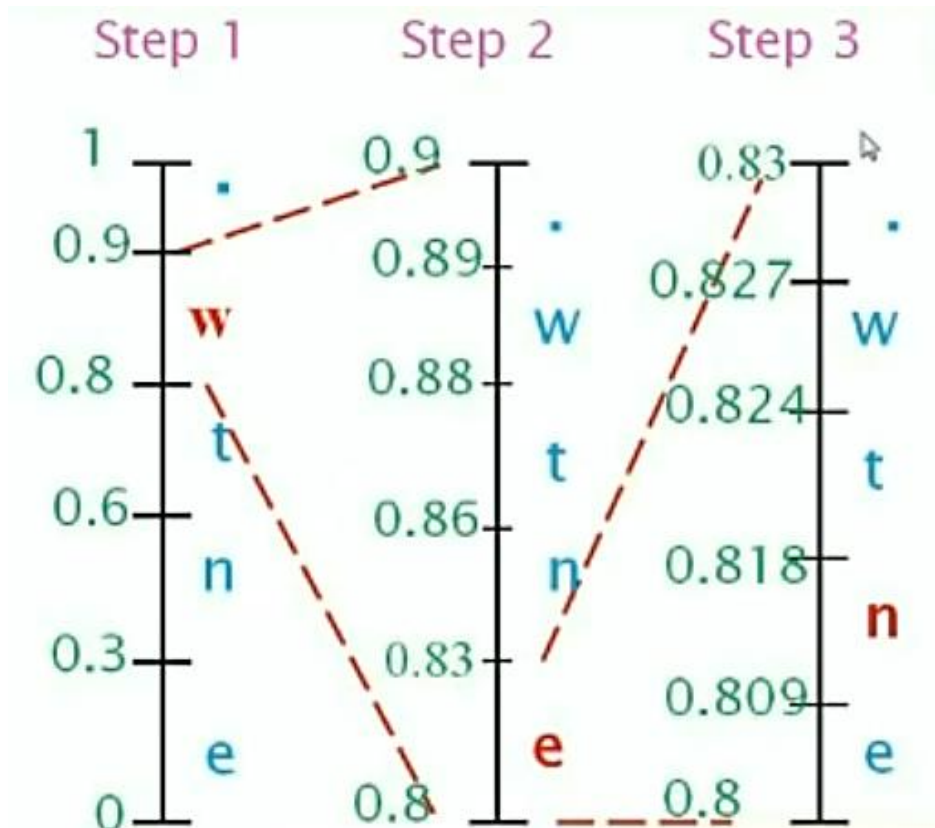
$$\text{lower limit} + d (\text{probability of 'symbol'})$$

$$\text{Range of 'e'} = 0.8 : [0.8 + 0.1 (0.3)] = 0.8 : 0.83$$

$$\text{Range of 'n'} = 0.83 : [0.83 + 0.1 (0.3)] = 0.83 : 0.86$$

$$\text{Range of 't'} = 0.86 : [0.86 + 0.1 (0.2)] = 0.86 : 0.88$$

$$\text{Range of 'w'} = 0.88 : [0.88 + 0.1 (0.1)] = 0.88 : 0.89$$



Step3

$$d = \text{upper bound} - \text{lower bound} = 0.83 - 0.8 = 0.03$$

Range of 'symbol' = lower limit : upper limit

lower limit + d (probability of 'symbol')

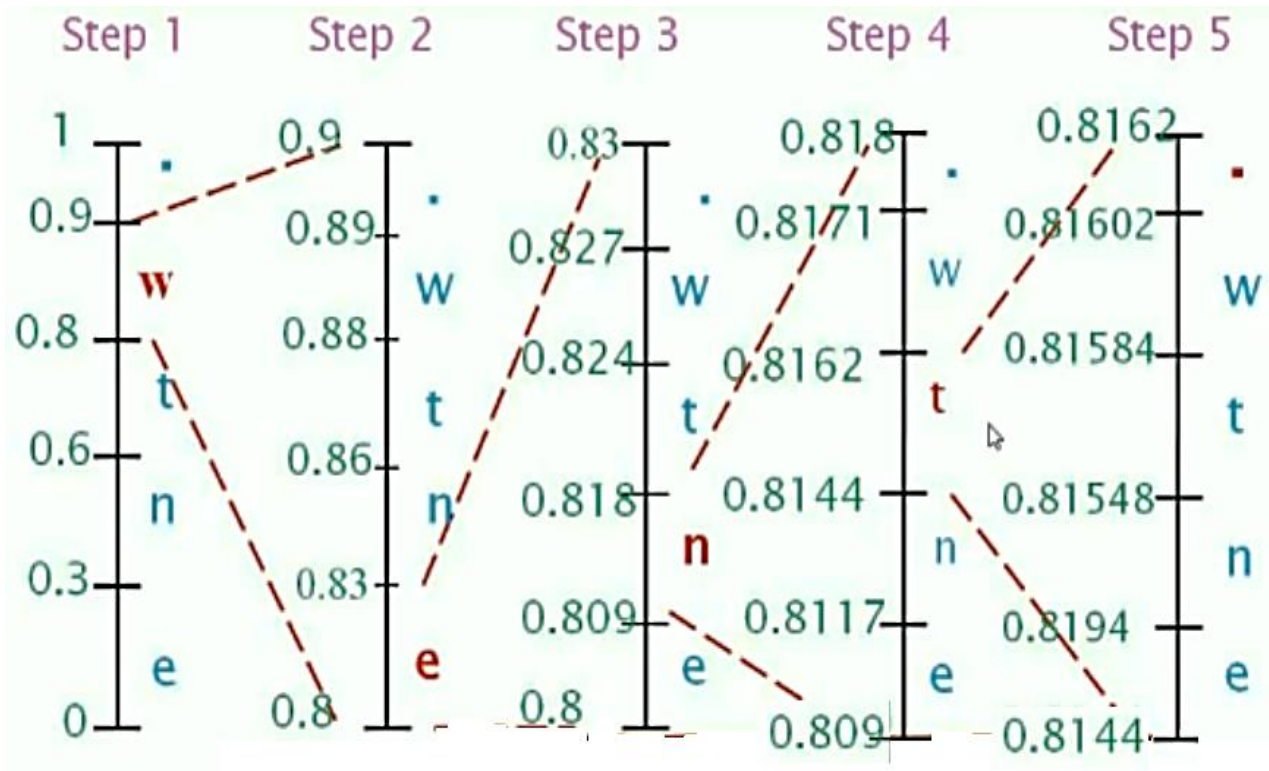
Range of 'e' = 0.8 : [0.8 + 0.03 (0.3)] = 0.8 : 0.809

Range of 'n' = 0.809 : [0.809 + 0.03 (0.3)] = 0.809 : 0.818

Range of 't' = 0.818 : [0.818 + 0.03 (0.2)] = 0.818 : 0.824

Range of 'w' = 0.824 : [0.824 + 0.03 (0.1)] = 0.824 : 0.827

Range of '.' = 0.827 : [0.827 + 0.03 (0.1)] = 0.827 : 0.83



نتوقف عندما نصل اخر رمز

Arithmetic Code word And Tag

Hence the arithmetic code word is
 $0.81602 < \text{codeword} < 0.8162$

Generation of Tag:

Tag = (Upper limit of code word + Lower limit
 of code word)/2

$$\begin{aligned} \text{Tag} &= (0.8162 + 0.81602) / 2 \\ &= 0.81611 \end{aligned}$$

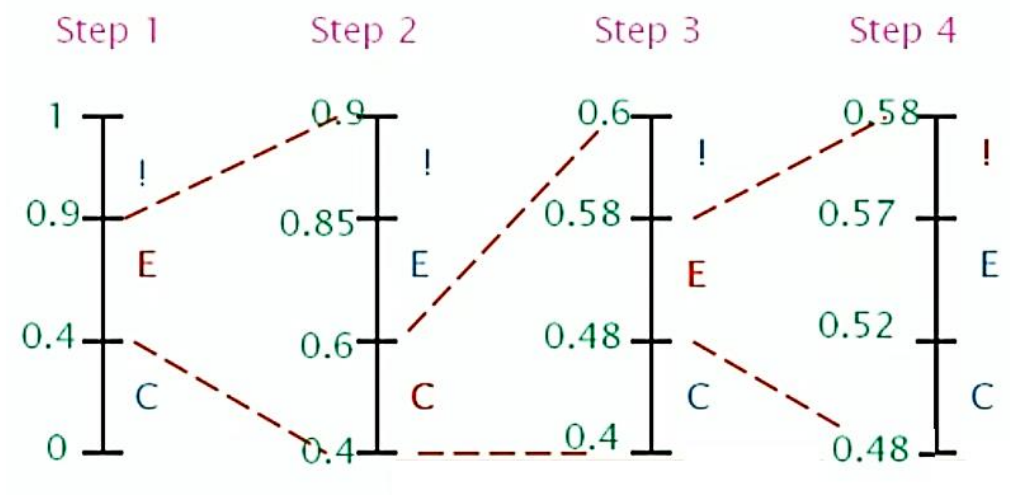
Arithmetic Decoding Procedure

► Decode the message 0.575 given the coding model.

Symbol	!	C	E
Probability	0.1	0.4	0.5

في خوارزمية فك الترميز decoding يتم تطبيق نفس خوارزمية الترميز في الخطوة الأولى ثم يتم ملاحظة قيمة الرسالة المستلمة ويتم تحديد الفترة التي تقع فيها القيمة ليتم تحديد الحرف الذي سيتم العمل عليه في الخطوة التالية وتكرر الخطوة الأخيرة إلى أن نصل إلى القيمة المستلمة.

Solution :



Decoded Message

Since 0.575 lies between 0.57 & 0.58 which is the range of termination character, stop the process

The decoded message is obtained by tracing the expanded characters i.e. "ECE!"

H.W. Write the Decoding steps for Arithmetic coding.

Channel Coding

Components of a communication system : (Information Source, **Transmitter**, **Channel**, **Receiver**, and Destination).

A communication system sometimes encounter several obstacles عقبات to the surrounding environment with the components, the **communication channel** is most of these components **is liable** عرضة to be affected للتأثر by what may alter or corrupt بما قد يغير أو غير المتعمد or unintentional غير المتعمد the transmitted data (eg: error or **noise deliberate** متعمد). (...).

Therefore, when this system requires **high reliability**, we are forced to develop a mechanism to ensure that it does not occur any error in the receiver, digital systems have been able to do so flexibly.

Purpose of Channel Coding :

for **Reliable Communication** either

1. to protect information from **channel noise**, **distortion** and **jamming**. Or
2. to protect information from the 3rd parity (enemy) which is the subject of encryption.

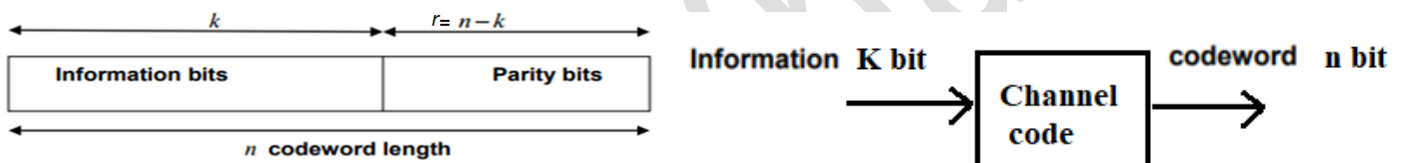
Error Detecting And Correcting Codes

Concept:

The basic idea behind error detecting or correcting codes is to add extra bits (or digits) to the information such that the receiver can use it to detect or correct errors with limited capabilities.

These extra redundant bits are called parity or check or correction bits.

So, if for each k digits, r parity bits are added then, the transmitted $k+r=n$ bits will have r redundant bits and the code is called (n,k) code with code efficiency or rate of (k/n) . In general, the ability of detection or correction depends on the techniques used and the n, k parameters.



code write as (n,k)

channel code efficiency $= k/n * 100\%$

$n=k+r$

k : bit in data

r : rudandance bit

n :code worde bit

Example:

code $(8,7)$ that is mean $n=8$ bit, $k=7$ bit and $r=8-7=1$ bit

efficiency= $K/N= 7/8=87.5\%$

code $(5,4)$ then $n=5$, $k=4$, $r=1$ bit

efficiency= $K/N= 4/5*100= 80\%$

Channel coding techniques:

1- Error Detecting Codes

1-1 Parity check (odd_even detector).

1-2 Repetition codes.

2- Error corrected Codes

☐ Forward error correction (FEC)

- Hamming code.

☐ Automatic repeat-request (ARQ).

☐ Hybrid ARQ combines ARQ and FEC.

1- Error Detecting Codes:

There are regular ways to detect the error where the sender sends original information attached to with a fixed number of check bits derived from bits of information using an algorithm

at the receiver in the future the same algorithm is applied to compare the resulting cells with if there is a mismatch, there is an error ... and we have irregular ways in which converting the entire message.

1-1 parity check bit (odd_even detector) method:

A parity bit is a bit that is added to a group of source bits to ensure that the number of set bits (i.e., bits with value 1) in the outcome is even or odd.

idea: add a redundant bit to each block of data that forces it to have even parity; e.g.,

- if the block of data has even parity, add a 0 to the end
- if the block of data has odd parity, add a 1 to the end

			$r = 1$	
0000	even parity	→	00000	even parity
0001	odd parity	→	00011	even parity
0010	odd parity	→	00101	even parity
0011	even parity	→	00110	even parity
0100	odd parity	→	01001	even parity
			↑ parity bit	

Example: write code word truth table for data of 3 bits.

Solution:

data (k bits)	code words (n=k+1 bits)
000	0000
001	0011
010	0101
011	0110
100	1001
101	1010
110	1100
111	1111

Note: This method fails if there is an error in an even number of bits

Example: Using parity check code (8,7), find code word if data are: data information1=[0111010]
 data information2=[1100111],
 data information3=[0000000]

solution:

code word ci=[li:parity bit]

c1=[01110100]

c2=[11001111]

c3=[00000000]

1-2 Repetition codes

A *repetition code* is a coding scheme that repeats the bits across a channel to achieve error-free communication. Given a stream of data to be transmitted, the data are divided into blocks of bits. Each block is transmitted some predetermined number of times.

For example, to send the bit pattern "1011", the four-bit block can be repeated three times, thus producing "1011 1011 1011". However, if this twelve-bit pattern was received as "1010 1011 1011" – where the first block is unlike the other two – it can be determined that an error has occurred.

A repetition code is very inefficient, and can be susceptible to problems if the error occurs in exactly the same place for each group (e.g., "1010 1010 1010" in the previous example would be detected as correct). The advantage of repetition codes is that they are extremely simple, and are in fact used in some transmissions of numbers stations.

Basic definitions:

1- Systematic and nonsystematic code:

Systematic code: If the check bits (the redundancy bits added with message bits to create code word) and the message bits can be **separated or identified from the codewords** then it is called systematic code.

Example: Input data $D=[a_1 a_2 a_3 \dots a_k]$, K bits
 $R=[c_1 c_2 c_3 \dots c_r]$ r bits

- Output systematic (n,k) **codeword** is

$$C=[a_1 a_2 a_3 \dots a_k c_1 c_2 c_3 \dots c_r] \text{ n bits.}$$

Non-systematic code: If the check bits (the redundancy bits added with message bits to create codeword) and the message bits **can not be separated or identified** from the

codewords then it is called non-systematic code. They are mixed in the block of the codeword.

Example: Input data $D=[a_1 a_2 a_3 \dots a_k]$, K bits
 $R=[c_1 c_2 c_3 \dots c_r]$ r bits

- Output nonsystematic (7,4) **codeword** is $C=[c_2 a_1 c_3 a_2 c_1 a_4 a_3]$

2- Hamming Distance (**HD**): is the difference between two strings of equal length , i.e. is the minimum number of positions of errors at which the corresponding symbols are different.

Example1: The Hamming distance between:

- **1011101** and **1001001** is 2.
- **2173896** and **2233796** is 3.

The Hamming distance between two codewords C_i and C_j is denoted by d_{ij} which is the number of bits that differ.

Example2: Find the Hamming distance between the two codewords:

$[C1]=[1011100]$ and

$[C2]=[1011001]$.

Solution:

Here, the no. of bits that differ is 2, hence $d_{12}=2$.

Example: Find the minimum Hamming distance for the 3 codewords:

$[C1]=[1011100]$,

[C2]=[1011001]
 [C3]=[1011000].

Solution:

Here $d_{12}=2$,
 $d_{13}=1$ and
 $d_{23}=1$.

Hence $\min(d_{ij})=1=(HD)$.

Note that the calculation of HD becomes more difficult if no of codewords increases.

3- Hamming weight: This is the number of 1's in the non-zero bits in codeword C_i . It is denoted by ω_i .

Example, if
 and for
 so on.

Examples

String	Hamming weight
11101	4
11101000	4
00000000	0
789012340567	10

[C1]=[1011000], then $\omega_1=3$,
 [C2]=[0001010], then $\omega_2=2$, and

Hamming Code

Channel Coding Techniques:

2-Error corrected Codes

Three approaches can be used to cope with data transmission errors.

- Forward error correction (FEC): Error detection and correction.
where controlled redundant information is added to the symbol stream. The redundant information can be utilized for error detection and error correction, such as Hamming code method.
- Automatic repeat-request (ARQ): **Error** detection
where the symbol frame is retransmitted if symbol errors are detected in the reception.
- Hybrid ARQ combines ARQ and FEC.

Hamming Code

Method Hamming Code is one method of **error detection and error correction** the most simple. This method uses logic operation XOR (Exclusive OR) in the process of error detection, and the process of error correction, while input and output of data from the method of Hamming Code in the form of binary numbers.

Hamming Code method invented by **Richard W. Hamming** in the 1940s.

Hamming Code Method is one of the error detection methods that can detect some errors, but it is only **capable of correcting one error**.

This error detection method is suitable for use in situations where errors are occurring randomly.

Hamming Code Method inserts multiple pieces of check bits into data. The number of check bits inserted depending on the length of the data.

Calculating the Hamming Code

The key to the Hamming Code is the use of extra parity bits to allow the identification of a single error. Create the code word as follows:

Example

Let us understand hamming code error correction through an example. Assume,
Data = 10011010

1- Calculate the number of parity bits to be added to data (r):

$$2^r \geq k + r + 1 \dots\dots (i) \quad \text{where, } r = \text{number of parity bits,} \\ k = \text{number of message bits} = 8$$

Using **hit and trial** method, we will find the value of 'r'.

Let's assume we first take $r = 4$.

Now, using equation (i)

$$16 \geq 8 + 4 + 1 \text{ (Therefore, the inequality holds)}$$

2- Calculating the parity bits position

The parity bits are added at power of 2's position. For example, position will be 1st (2^0), 2nd (2^1), 4th (2^2), 8th (2^3), ...and so on.

(Positions 1, 2, 4, 8, 16, 32, 64, etc.)

$$\begin{aligned} r=4 \\ r1 &= \text{Positon1}(P1) = 2^0 = 1 \\ r2 &= \quad = \quad (P2) = 2^1 = 2 \\ r3 &= \quad = \quad (P4) = 2^2 = 4 \\ r4 &= \quad = \quad (P8) = 2^3 = 8 \end{aligned}$$

3- Mark all bit positions as parity bits and data bit onto codeword

Position Check Bit (redundant bits)	
$r_i = 2^{i-1}$	
Check Bit	Position
r1	1
r2	2
r3	4
r4	8
r5	16
r6	32
r7	64
r8	128
r9	256

Now, **total bits** that will be sent to the **receiver** will be the

$$\text{Codeword}(n) = \text{message(data) bits}(k) + \text{parity bits}(r) = 8+4 = 12 \text{ bit}$$

1	2	3	4	5	6	7	8	9	10	11	12
2^0	2^1		2^2				2^3				
P1	P2		P4				P8				
r1	r2		r3				r4				

All other bit positions are for the data to be encoded.

(positions 3, 5, 6, 7, 9, 10, 11, 12)

1	2	3	4	5	6	7	8	9	10	11	12
2^0	2^1		2^2				2^3				
P1	P2	D1	P4	D2	D3	D4	P8	D5	D6	D7	D8
		1		0	0	1		1	0	1	0

4- Each parity bit calculates the parity for some of the bits in the code word. The position of the parity bit determines the sequence of bits that it alternately checks and skips.

Position 1: check 1 bit, skip 1 bit, check 1 bit, skip 1 bit, etc. (1,3,5,7,9,11,13,15,...)

Position 2: check 2 bits, skip 2 bits, check 2 bits, skip 2 bits, etc. (2,3,6,7,10,11,14,15,...)

Position 4: check 4 bits, skip 4 bits, check 4 bits, skip 4 bits, etc.

(4,5,6,7,12,13,14,15,20,21,22,23,...)

Position 8: check 8 bits, skip 8 bits, check 8 bits, skip 8 bits, etc. (8-15,24-31,40-47,...)

Position 16: check 16 bits, skip 16 bits, check 16 bits, skip 16 bits, etc. (16-31,48-63,80-95,...)

Position 32: check 32 bits, skip 32 bits, check 32 bits, skip 32 bits, etc. (32-63,96-127,160-191,...)

etc.

Set a parity bit to 1 if the total number of ones in the positions it checks is odd. Set a parity bit to 0 if the total number of ones in the positions it checks is even.

Calculate the parity for each parity bit (a ? represents the bit position being set):

- Position 1 checks bits 1,3,5,7,9,11:

? _ 1 _ 0 0 1 _ 1 0 1 0. Even parity so set position 1 to a 0: **0 _ 1 _ 0 0 1 _ 1 0 1 0**

- Position 2 checks bits 2,3,6,7,10,11:

0 ? 1 _ 0 0 1 _ 1 0 1 0. Odd parity so set position 2 to a 1: **0 1 1 _ 0 0 1 _ 1 0 1 0**

- Position 4 checks bits 4,5,6,7,12:

0 1 1 ? 0 0 1 _ 1 0 1 0. Odd parity so set position 4 to a 1: **0 1 1 1 0 0 1 _ 1 0 1 0**

- Position 8 checks bits 8,9,10,11,12:

0 1 1 1 0 0 1 ? 1 0 1 0. Even parity so set position 8 to a 0: **0 1 1 1 0 0 1 0 1 0 1 0**

1	2	3	4	5	6	7	8	9	10	11	12
2 ⁰	2 ¹		2 ²				2 ³				
P1	P2	D1	P4	D2	D3	D4	P8	D5	D6	D7	D8
0	1	1	1	0	0	1	0	1	0	1	0

Code word: 011100101010.

Decoding a message in Hamming Code

Once the receiver gets an incoming message, it performs recalculations to detect errors and correct them. The steps for recalculation are –

- **Step 1** – Calculation of the number of redundant bits.
- **Step 2** – Positioning the redundant bits.
- **Step 3** – Parity checking.
- **Step 4** – Error detection and correction

Step 1 – Calculation of the number of redundant bits

Using the same formula as in encoding, the number of redundant bits are ascertained.

$$2^r \geq k + r + 1$$

where k is the number of data bits and r is the number of redundant bits.

Step 2 – Positioning the redundant bits

The r redundant bits (Check bits) placed at bit positions that is calculated based on the formula for calculating the position following the check bits:

$$C_i = 2^{i-1}$$

So that got the check bit position table as follows:

Position Check Bit

Check Bit	Position
C_1	1
C_2	2
C_3	4
C_4	8
C_5	16
C_6	32
C_7	64
C_8	128
C_9	256

Step 3 – Parity checking

Parity bits are calculated based upon the data bits and the redundant bits using the same rule as during generation of c_1, c_2, c_3, c_4 etc. Thus

c_1 = parity (1, 3, 5, 7, 9, 11 and so on)

c_2 = parity (2, 3, 6, 7, 10, 11 and so on)

c_3 = parity (4-7, 12-15, 20-23 and so on)

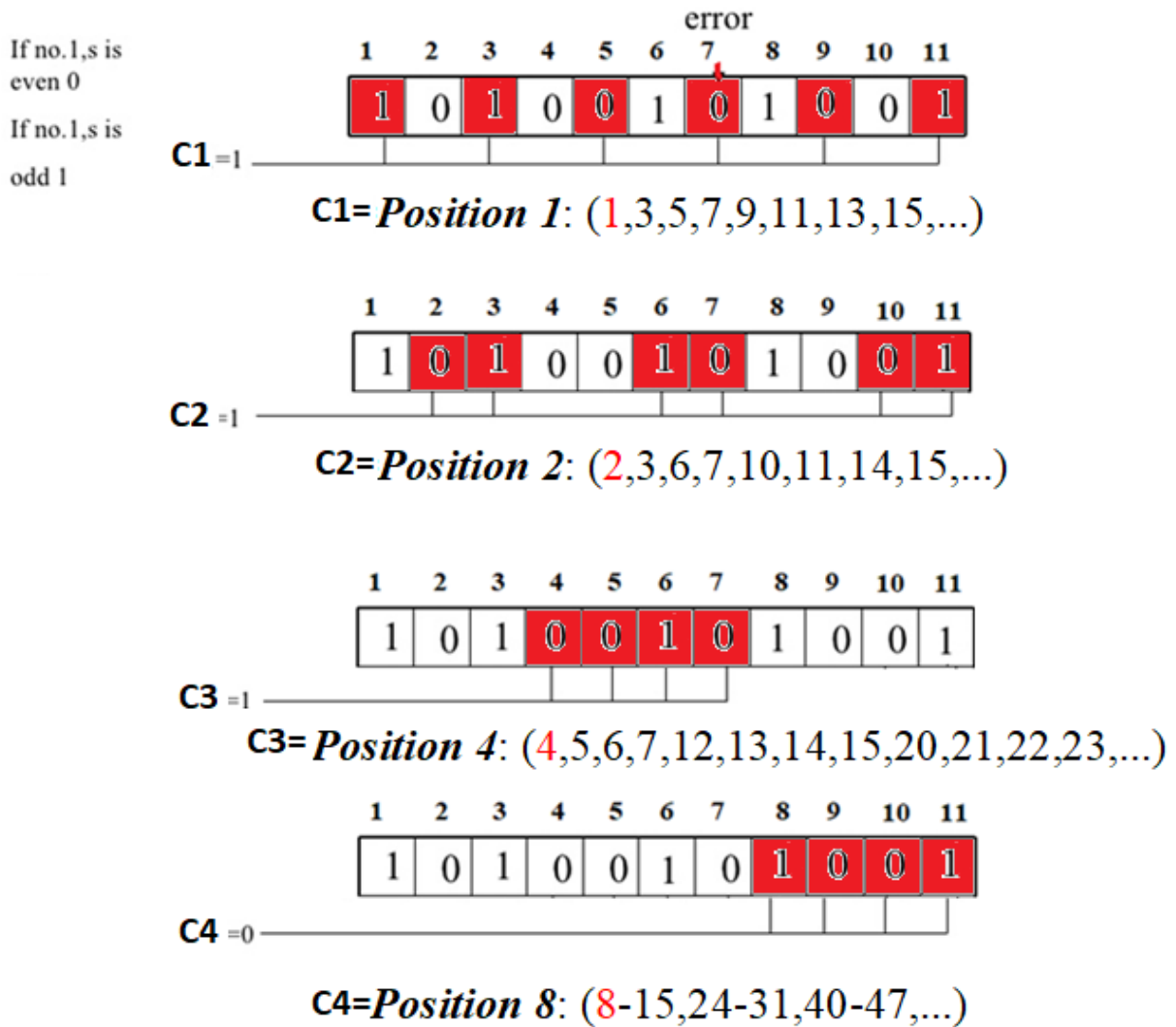
c_4 = parity (8-15, 24-31 and so on)

Step 4 – Error detection and correction

The decimal equivalent of the parity bits binary values is calculated. If it is 0, there is no error. Otherwise, the decimal value gives the bit position which has error. For example, if $c_4c_3c_2c_1 = 1001$, it implies that the data bit at position 9, decimal equivalent of 1001, has error. The bit is flipped to get the correct message.

Error detecting using hamming code

Example: suppose the input data is transmitted 10100111001.
 Data received by the receiver in the form of 10100101001, the
 error detection and error correction of the method of Hamming
 Code is



$$\begin{matrix} 8 & 4 & 2 & 1 \\ C4 & C3 & C2 & C1 \\ 0 & 1 & 1 & 1 \end{matrix} \Rightarrow (0111)_2 = (7)_{10} \text{ OR } 1+2+4=7$$

It mean the 7 bit is corrupted

هذا يعني : تلف البت السابع