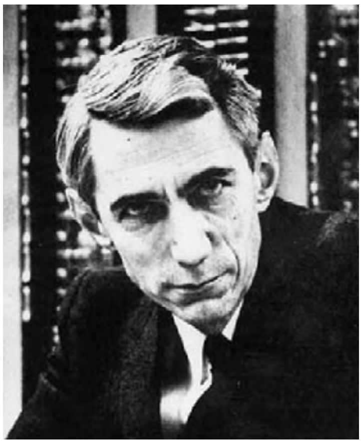
情報理論とは

情報の伝達を如何に効率よく，信頼性高く行うかに関する理論．
20世紀の半ばにシャノン(C.E. Shannon)により構築された．

Claude E Shannon's father was also named Claude Elwood Shannon and his mother was Mabel Catherine Wolf. Shannon was a graduate of the University of Michigan, being awarded a degree in mathematics and electrical engineering in 1936. Although he had not been outstanding in mathematics, he then went to the Massachusetts Institute of Technology where he obtained a Master's Degree in electrical engineering and his Ph.D. in mathematics in 1940.



Shannon wrote a Master's thesis *A Symbolic Analysis of Relay and Switching Circuits* . His doctoral thesis was on population genetics.
Shannon joined AT&T Bell Telephones in New Jersey in 1941 as a research mathematician and remained at the Bell Laboratories until 1972.
D Slepian, a colleague at the Bell Laboratories wrote:-
Many of us brought our lunches to work and played mathematical blackboard games but Claude rarely came. He worked with his door closed, mostly. But if you went in, he would be very patient and help you along. He could grasp a problem in zero time. He really was quite a genius. He's the only person I know whom I'd apply that word to.
Shannon published *A Mathematical Theory of Communication* in the *Bell System Technical Journal* (1948). This paper founded the subject of information theory and he proposed a linear schematic model of a communications system. This was a new idea. Communication was then thought of as requiring electromagnetic waves to be sent down a wire. The idea that one could transmit pictures, words, sounds etc. by sending a stream of 1s and 0s down a wire, something which today seems so obvious as we take this information from a server in St Andrews, Scotland, and view it anywhere in the world, was fundamentally new.（ウィキペディアより）

情報伝達のモデル

情報が伝えられてゆく過程は大まかに次の3つから構成される．

（1）伝送される情報

（2）情報を送る人（送信者）と受け取る人（受信者）

（3）情報が伝送される通り道（伝送媒体・手段）

例えばAさんがBさんに携帯電話で電話をかけるときは，

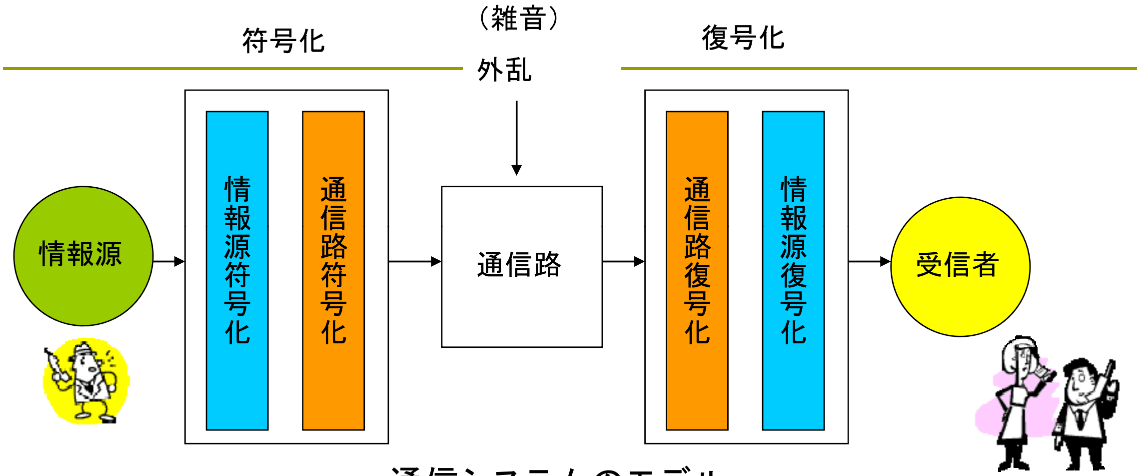
（1）の情報は，Aさんの発する言葉

（2）の送信者はAさんであり，受信者はBさん

（3）の伝送媒体・手段は電波や光ケーブル

である．

前のページの情報伝達のモデルを少し詳しく見ると・・・以下のようなシステムになっていることがわかる.



通信システムのモデル

情報源・・・送信者の発する情報

符号化・・・英文の通信を考える. 英文のまま通信することはできないから英文を符号化する必要がある.

A ⟶	0x41	0100 0001
B ⟶	0x42	0100 0010

情報源符号化

情報源の統計的性質を利用して効率の向上を図る符号化

例えば各英文字を頻度の高い文字には短い系列を割り当て, 頻度の低い文字には長い系列を割り当てる.

具体的方法　ハフマン符号化, ランレングス符号化, 算術符号化（シャノン・ファノ符号）

情報源復号化　0,1系列から受信者が利用できる元の情報, 例えばアルファベット記号に変換する作業

通信路符号化

現実の通信路においては種々の原因に起因して雑音による誤りが生じる. これをできるだけ排除して誤りの無い情報の伝達を実現するためには, これに対する対策を講じる必要がある. この誤り対策のための符号化を通信路符号化という.

例1

0 1 0 ⟶ 0 0 0　1 1 1　0 0 0（各記号を繰り返し3回送る）

（3回中1回誤っても正しく復号できる）

詳しくは符号理論（3年後期）で論じられる.

具体的方法として　ハミング符号, BCH符号等がある.

情報源通信路符号化の比較

符号化	最終目標	実現状況	符号の長さ	定理
情報源符号化	効率化	エネルギー・時間の節約	最短符号の実現	情報源符号化定理（シャノンの第1基本定理）
通信路符号化	信頼性の向上	誤りの検出・訂正	冗長性の付加	通信路符号化定理（シャノンの第2基本定理）

情報源の種類

- 離散情報源 （ディジタル情報源）・・・・授業ではこちらを仮定する.
- 連続的情報源 （アナログ情報源）

記号の定義

情報源アルファベット（”山田”，”片岡”，”渡辺”）といった情報を記号的に $S=\{s_1,s_2,\cdots,s_M\}$ と表したもの．時刻iでの情報源の出力を X_i で表す．

$X_0,X_1,\cdots,X_{n-1}$ の確率分布を

$$P_{X_0X_1\cdots X_{n-1}}(x_0,x_1,\cdots,x_{n-1})=[X_0=x_0,X_1=x_1,\cdots X_{n-1}=x_{n-1}\text{となる確率}]$$

とする．$x_0,x_1,\cdots,x_{n-1}$ は S の任意の元．

【重要】何故確率を導入するのか！？

頻度の大きい情報源には短い符号化を与え，頻度の小さい情報源には長い符号化を与えれば直感的に効率的となることから確率導入の意義がわかるだろう.

結合確率分布 $P_{X_0 \cdots X_n}(x_0, \cdots, x_n)$ がわかれば全てがわかる！

問1 S={0,1}とする． $P_{X_0}(0)$ を求めよ．

x_0	x_1	x_2	$P_{X_0X_1X_2}(x_0, x_1, x_2)$
0	0	0	0.648
0	0	1	0.072
0	1	0	0.032
0	1	1	0.048
1	0	0	0.072
1	0	1	0.008
1	1	0	0.048
1	1	1	0.072

条件付確率

これからマルコフ情報源という概念が登場するが，そこでは条件付確率が重要となる．ところで，条件付確率について高校で学習していないようなのでここで，ついでに勉強してしまおう．

定義 $X_0 = x_0$ の下での X_1 の条件付確率

$$P_{X_1|X_0}(x_1 | x_0) = \frac{P_{X_0X_1}(x_0, x_1)}{P_{X_0}(x_0)}$$

問2 S={0,1}とする．$P_{X_1|X_0}(1|0), P_{X_1|X_0}(0|0)$ を図で説明せよ．

問3 $P_{X_0|X_1X_2}(0|1,1)$を求めよ．

問4 {0,1}の値をとる3つの確率変数X,Y,Zについて $P_{XYZ}(x,y,z)$ が図のように与えられている．空欄を埋めよ．

x	y	z	$P_{XYZ}(x,y,z)$
0	0	0	0.504
0	0	1	0.056
0	1	0	0.07
0	1	1	0.07
1	0	0	0.096
1	0	1	0.024
1	1	0	0.054
1	1	1	0.126

（1）

x	y	z	$P_{Z\|X,Y}(z\|x,y)$
0	0	0	
0	0	1	
0	1	0	0.5
0	1	1	0.5
1	0	0	0.8
1	0	1	0.2
1	1	0	0.3
1	1	1	0.7

（2）

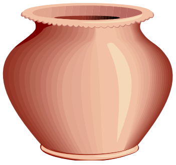
x	y	$P_{Y\|X}(y\|x)$
0	0	
0	1	
1	0	0.4
1	1	0.6

（3）

x	$P_X(x)$
0	
1	

演習

<u>問5</u>　つぼに2つの黒球と2つの白球が入っている．つぼから一度に1つだけボールを取り出し，戻さないこととする．（1）最初に取り出した球が黒球である確率は？（2）2個目に取り出した球が黒であったと告げられたとする．最初に取り出した球が黒球である確率は？



余談ですがつぼにボールが入っているという状況を渡辺は見たことがありません^^）

まさに問題のための問題？

っていうのが結構あるんだよね・・・

<u>問6</u>　問5のつぼの問題に戻る．最初に取り出した球が黒球であったとする．2回目に取り出した球が黒球である確率を求めよ．2とおりの方法を考えてみよう．

事象Bが事象Aの<u>事前に</u>起こっているときは事後に起こっているときと比べ条件付確率は簡単になることが多い