1. 情報符号理論の枠組み

本講義では，Fano による Shanon 理論修正モデルに基づいて講義を進める．

1.1 理論の枠組み

現象はいろいろな説明の仕方がある．物理学者は与えられた状況を説明するために物理学の理論の枠組みを作り，そこで定められた概念と用語と法則を使い，与えられた現象を説明しようとする（図 1）．質点，力，位置，速度，あるいは，電気や光の概念を用いて，それらを現象の構成要素に当てはめて，各構成要素が物理法則によって挙動している，という説明や，将来はこのようなことが起きるだろう，といった予測を行う．

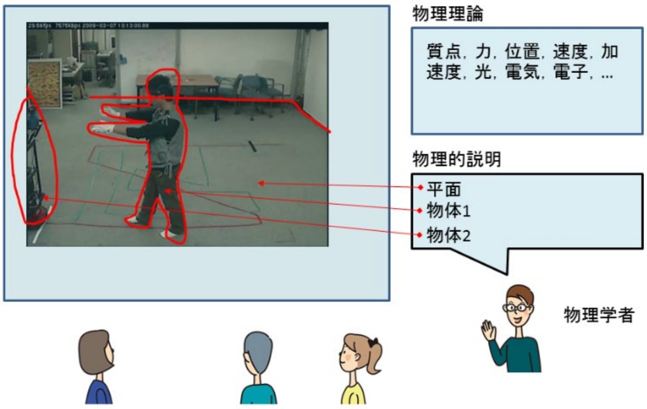

図１：現象の物理学的説明

他方，情報科学者は図 2 のように情報の理論の枠組みを使って現象を説明しようとする．物理概念の代わりに情報概念を用いて現象の説明や予測を試みる．

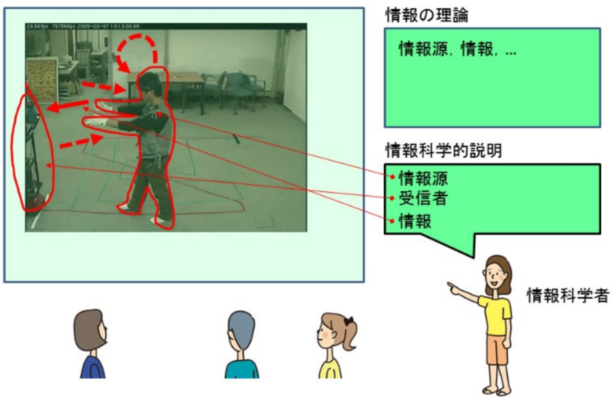

図 2: 現象の情報科学的説明

生命現象や社会的な現象は，物理レベルからの説明を試みると複雑になりすぎて理解しにくいが，知覚，記憶，情動といった情報に関わる概念を用いると説明が分かりやすくなることが多い．

1.2 情報とは何か？

情報符号理論の中心となるのは情報概念である．直観的には，情報とは「不確かさを減らすもの」と規定される．経済学，熱力学，量子力学などで生じるいくつかの現象は情報概念あるいはその対極であるエントロピー概念を用いた説明が広く行われてきた．

情報概念の効用がよくわかるのは，生命の行動を理解しようとしたときであろう[西垣 2004, 2008]．生命は，環境の中で生存するために，食物を獲得したり，外敵の存在を察知していち早く身を隠したり，時には連携して行動する．こうした行動を，質量，力，エネルギー，波動などの概念によって説明しようとしてもうまくいきそうもないことは明らかだろう．

情報概念を用いて生命の行動を理解しようとするならば，その第一歩は，

> 生命は，複雑なオブジェクトの錯綜し，不確かさに満ちた環境の中で，嗅覚，聴覚，視覚など，さまざまな手段を使って，自分の行動を妨げる不確かさを減らすことにより，（時には，情報を交換し合って連携し），食物を獲得し，外敵から身を守ることにより，生存している

と言い換えてみることである．さらに詳細な理解を得るためには，具体的に生命が行動のためにどのような手掛かりを利用しているか，それによって外界に関する不確かさが如何に減少しているか，また，手掛かりからどのように行動が生成されるか，といったあたりについて，掘り下げていくことになる．

現代社会において，我々の周りにある人工物は多かれ少なかれこうした生命と類似の能力――環境にある情報を感知し，行動する能力――を持っている．

（例 1）センサーライト

人間の体温に反応して光を灯す人工物．環境の中にある情報――この場合は人間がいるか否かの確からしさを減らすもの――を（ある確度で）捉えて，電球を点灯させる．

（例 2）ロボット

もっと複雑な人工物だ．人工物の中でも生命や人間に近い．環境のなかの情報をより繊細にとらえ，行動に結び付ける．

以上の考察をまとめてみると．

- 情報＝不確かさを減らすもの
- 生命は，環境の中のさまざまな情報に反応して，複雑な行動をする．
- 機械は，環境内の情報を検出し，組み込まれた機序に従って行動する．

広義の情報符号理論では，観測者と情報の関わり，情報の表現，情報の解釈に関わる広い現象について論じる．これは壮大な企てであり，まだ完結には程遠いと言っていいだろう．また，これについて論じることは本講義の視野を超えている．

これに対して，本講義で取り上げる狭義の情報符号理論では，次のような単純化をおこなう．

- 解釈の仕方を固定．与えられた確率モデルに従って解釈が行われるものとする．観測者の個性を捨象する．
- 記号系による情報表現を前提とした離散系の理論と，連続的な時系列の上に展開される連続系の理論.

こうした単純化の見返りに，情報の量的側面について数理理論を構築することができる．その典型的な応用は，情報圧縮，誤り訂正である．例えば，サンプリング理論は離散的なサンプリングでもとの連続時系列が完全に復元できることを理論的に示唆される．

1.3 Shannon モデルと Fano による修正モデル

情報の数学的理論の枠組みとして，Shannon モデル（図 3）が知られている．

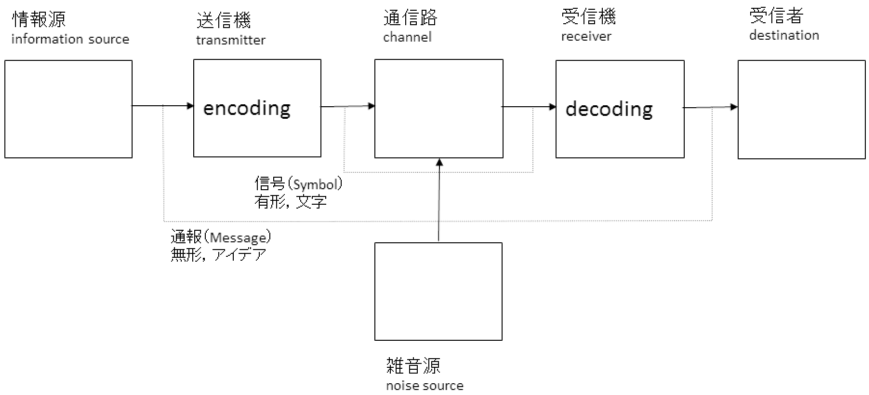

図 3: Shannon モデル

Shannon モデルでは，情報源で発生した情報は送信機において何らかの言語で表現されて信号として実体をもつ存在となり，これが通信路を経て，受信機に送られ，再び情報に変換され，受信者に伝えられる．通信路には雑音源からの雑音が混入し，通信路のインプットは必ずしもアウトプットと一致しない．そのような状況で通信路での誤りの混入は避けられないが，通信路での伝送の仕方を工夫すれば，1 メッセージあたりの誤り率は任意に下げることができる．Shannon 理論の主要な問いは，情報源で発生したメッセージをどこまで効率的に（平均的に短く）表現できるか，通信路での誤りをどこまで下げることができるか，誤りを抑えたうえで通信路の伝送速度をどれだけ速めることができるか，といったもので

ある．

通信路の誤りに対応するために用いられる通信路誤り訂正や検出手法では，情報源からのメッセージを符号化した信号をさらに符号化したものを通信路を介して受信者に送信する．これを反映したのが Fano によって修正された Shannon モデル（図 4）である．

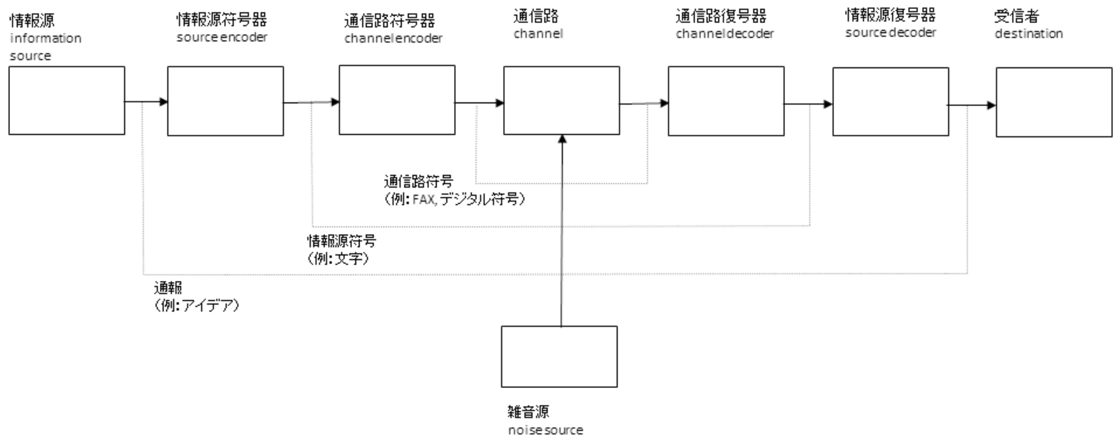

図 4: Fano によって修正された Shannon モデル

1.4 情報量の定量的尺度

情報の量的側面がどのように定式化されるか，次のような例を用いて説明しよう．

図 5 の写真で，今，自分の友人（「山田さん」）がこの路地で先に行ってしまい，左に行ったか右に行ったかわからなくなってしまったとしよう．

図 5: 例：街角で友人（「山田さん」）を見失ったとき，「山田さんはこの路地の奥に行ったよ」という通報がもたらす価値は？

単純のために可能性は左に行ったか，右に行ったかのいずれかであるとし，さらに左に行った確率をp，右に行った確率を$1-p$とする．すると，このとき，「山田さんはこの路地の奥（左）に行ったよ」ということを教えてもらうと，その言葉の持つ情報量は，$\log_2 \frac{1}{p}$ビット

とする. $p=\frac{1}{2}$のときだと, その言葉は 1 ビットの情報量であるが, pがゼロに近づくにつれ, 情報量（サプライズ）は増え，逆に 1 に近づくときは，情報量は減る．

自分の友人の行先を聞くことによって得られる平均的な情報量は,

$$p\log_2\frac{1}{p}+(1-p)\log_2\frac{1}{1-p}\text{ビット}$$

となる．この値は，$p=\frac{1}{2}$のとき最も大きい．つまり，左に行った確率も右に行った確率も等しいとき最大になる．

我々は劇場メタファーを設定しているから，そこでは我々の「台本」すなわちモデルに従ってシーンが移り変わっていく．モデルの品質が高ければ，そこで演じられる「劇」は実際に起きることにそれだけ近いことになる．

1.5 相互情報量の概念

今度は，知りたいことについての直接的な答えが得られない場合について考えてみよう．前の例と同様に，自分の友人が街の中の分かれ道で左か右のどちらに行ったかわからなくなってしまったとする．また，左の道から来た人に「若い男性を見かけませんでしたか？」という質問をすることはできるが，その質問に対する答えは必ずしも正確ではないとする．この場合，全事象は，友人が左右のいずれに行ったか，左から来た人が見かけたと言うか否かの 4 つのケースから構成される．さらに，

- 友人が実際に左に行った場合，左の道から来た人に「若い男性を見かけませんでしたか？」と尋ねると，(そんなに注意していたわけではないなどの理由で)「見ていない」と答える割合は全体の 20%，「見かけた」と答える割合は全体の 30%である．
- 友人が右に行ったとする．左の道から来た人が「若い男性を見ていない」と答える割合は全体の 40%，（ほかの若い男性を見たなどの理由で）「見かけた」と答える割合は全体の 30%である．

とする．

	左	右
見逃	20%	40%
目撃	30%	10%

図 6: 間接的な通報のもたらす価値は？

上に述べた状況について検討するための前提は，結合確率を用いて次のように記述できる．

$$P(\text{左},\text{「見てない」}) = 0.2 \quad P(\text{左},\text{「見た」}) = 0.3$$
$$P(\text{右},\text{「見てない」}) = 0.4 \quad P(\text{右},\text{「見た」}) = 0.1$$

ここで， $P(\text{左},\text{「見てない」}) = 0.2$は友人が左に行き，かつ，左から来た人が「見てない」と報告する確率が 0.2 であることを示している．他も同様．

上記の確率モデルからいろいろな確率が導ける．そもそも，友人が左に行った確率，右に行った確率はそれぞれ，

$$P(\text{左}) = P(\text{左},\text{「見てない」}) + P(\text{左},\text{「見た」}) = 0.2 + 0.3 = 0.5$$
$$P(\text{右}) = P(\text{右},\text{「見てない」}) + P(\text{右},\text{「見た」}) = 0.4 + 0.1 = 0.5$$

である．一方，左から来た人が「見てない」と言うか，「見た」と言うかについては，

$$P(\text{「見てない」}) = P(\text{左},\text{「見てない」}) + P(\text{右},\text{「見てない」}) = 0.2 + 0.4 = 0.6$$
$$P(\text{「見た」}) = P(\text{左},\text{「見た」}) + P(\text{右},\text{「見た」}) = 0.3 + 0.1 = 0.4$$

となる．さらに，「見てない」という答えを得たという条件下で，友人が右に行ったか，左に行ったかについては，

$$P(\text{左} \mid \text{「見てない」}) = \frac{P(\text{左},\text{「見てない」})}{P(\text{左},\text{「見てない」}) + P(\text{右},\text{「見てない」})} = \frac{0.2}{0.2+0.4} = \frac{1}{3}$$
$$P(\text{右} \mid \text{「見てない」}) = \frac{P(\text{右},\text{「見てない」})}{P(\text{左},\text{「見てない」}) + P(\text{右},\text{「見てない」})} = \frac{0.4}{0.2+0.4} = \frac{2}{3}$$

であることが，結合確率モデルから導ける．同様に，

$$P(\text{左} \mid \text{「見た」}) = \frac{P(\text{左},\text{「見た」})}{P(\text{左},\text{「見た」}) + P(\text{右},\text{「見た」})} = \frac{0.3}{0.3+0.1} = \frac{3}{4}$$
$$P(\text{右} \mid \text{「見た」}) = \frac{P(\text{右},\text{「見た」})}{P(\text{左},\text{「見た」}) + P(\text{右},\text{「見た」})} = \frac{0.1}{0.3+0.1} = \frac{1}{4}$$

である．一方，左右のいずれに行ったかがわかった場合については，

$$P(\text{見てない} \mid \text{左}) = \frac{P(\text{左},\text{「見てない」})}{P(\text{左},\text{「見てない」}) + P(\text{左},\text{「見た」})} = \frac{0.2}{0.2+0.3} = \frac{2}{5}$$
$$P(\text{「見た」} \mid \text{左}) = \frac{P(\text{左},\text{「見た」})}{P(\text{左},\text{「見てない」}) + P(\text{左},\text{「見た」})} = \frac{0.3}{0.2+0.3} = \frac{3}{5}$$
$$P(\text{見てない} \mid \text{右}) = \frac{P(\text{右},\text{「見てない」})}{P(\text{右},\text{「見てない」}) + P(\text{右},\text{「見た」})} = \frac{0.4}{0.4+0.1} = \frac{4}{5}$$
$$P(\text{「見た」} \mid \text{右}) = \frac{P(\text{右},\text{「見た」})}{P(\text{右},\text{「見てない」}) + P(\text{右},\text{「見た」})} = \frac{0.1}{0.4+0.1} = \frac{1}{5}$$

となる．

この状況で，「若い男性を見かけましたか？」という問いに対する必ずしも正確ではない回答を得ることで，どれくらいのゲインがあると思えばいいだろうか？

この問いに答えるために，次の 2 つの量を計算してみよう．第一は，左から来た人から何も教えてもらっていない状況で，友人が左右どちらに行ったかを教えてもらったときに得る平均情報量．第二は，左から来た人から「見た」か「見てない」かを教えてもらったあとで，友人が左右どちらに行ったかを教えてもらったときに得る情報量．

第一の場合は，「左に行った」というメッセージがもたらす情報量は，

$$\log_2 \frac{1}{P(\text{左})} = 1\text{ビット}$$

である．同様に，「右に行った」というメッセージのもたらす情報量も 1 ビットであり，左に行ったか右に行ったかという問いに対するメッセージからは，平均すると 1 ビットの情報量が期待される．

第二の目撃したか否かについて教えてもらった場合については次のようになる．「見てない」という報告を受けたとすれば，左に行った確率は$\frac{1}{3}$，右に行った確率は$\frac{2}{3}$であるので，左に行ったか右に行ったかという問いに対するメッセージからは，平均すると

$$\frac{1}{3}\log_2 \frac{1}{\frac{1}{3}} + \frac{2}{3}\log_2 \frac{1}{\frac{2}{3}} \approx 0.918\text{ビット}$$

の情報量が期待される．一方，「見た」という報告を受けたとすれば，友人が左に行った確率は $\frac{3}{4}$，右に行った確率は$\frac{1}{4}$であるので，左に行ったか右に行ったかという問いに対するメッセージからは，平均すると

$$\frac{3}{4}\log_2 \frac{1}{\frac{3}{4}} + \frac{1}{4}\log_2 \frac{1}{\frac{1}{4}} \approx 0.811\quad\text{ビット}$$

の情報が得られる．そもそも「見てない」という報告が生じる確率は$\frac{3}{5}$，「見た」という報告が生じる確率は$\frac{2}{5}$であるので，左に行ったか右に行ったかという問いに対するメッセージから期待される情報量をその割合で平均すると，

$$\frac{3}{5} \times 0.918 + \frac{2}{5} \times 0.811 \approx 0.8755\quad\text{ビット}$$

となる．

左から来た人から何も教えてもらわなかったときに友人が左右どちらに行ったかというメッセージから期待される情報量は 1 ビットであったから，「見た」か「見てない」かを教えてもらうことで，友人が左右どちらに行ったかというメッセージによって平均的に得られる情報量は，

$$1 - 0.8755 = 0.1245\quad\text{ビット}$$

ほど減ったことになる．これがまさに，「若い男性を見かけましたか？」という問いに対する正確さが保証されていない回答が，平均的にもたらす情報量である．

換言すると，友人が左右のどちらに行ったかというメッセージを発生する情報源は，当初

は 1 ビットの不確かさをもっていたが,「若い男性を見かけましたか？」という問いに対する回答が得られるとたとえそれが正確でなくても，平均的に約 0.8755 ビットの不確かさに減少する．そして，減少した 0.1245 ビットは，まさしく「若い男性を見かけましたか？」という問いに対する回答の情報量である．

問い　友人が左右のいずれに行ったかを伝えてくれるメッセージが，若い男性を目撃したか否かについての不確かさを平均的にどれだけ減らすか，算出せよ．

問い　P(左,「見てない」), P(左,「見た」), P(右,「見てない」), P(右,「見た」)の値を変えたとき，若い男性を見かけたか否かというメッセージがもつ情報量はどう変化するか？

1.6 なぜ log か？

生起確率pの事象が起きたことを知らせる「文」の情報量を$I(p)$と一般的に表記しよう．これまでの議論では，$I(p) = C \log_e p$としてきた．なぜそのようにしたのか？

ここでは，$I(p)$を定めるために log という関数を用いたことが恣意的なものではなく，自然なものであることを示そう．そのために，$I(p)$の定義を決めるときに用いられた（はずの）原則について再考してみよう．

1. 「文」の情報量は「文」の価値を表すものと考える．「文」の価値が高いほど，その「文」の情報量も大きいとする．
2. 珍しい事象の生起を告げる「文」ほどその価値が高い．
3. 事象が珍しいほど，それを告げるためには長い「文」が必要になる．
4. 事象を伝えるために必要な「文」の長さの最小値をその「文」の価値とする．
5. 事象 A と事象 B が独立であるとき，事象 A と事象 B の両方が起きたことを告げる「文」は，概ね「A が起き，B も起きた」となるので，その長さの最小値は，事象 A が起きたことを告げる「文」（「A が起きた」）の長さの最小値と，事象 B が起きたことを告げる「文」（「B が起きた」）の長さの最小値の和である．
6. $I(p)$の関数はpの微分可能関数にしておきたい．

第 5，6 項から$I(p)$が$C \log_e p$という関数であることが決まる．つまり，事象 A の生起確率をp，事象 B の生起確率をqとすると，事象 A と事象 B が独立であるとき，事象 A と事象 B の両方が起きる確率は$p \times q$であるので，第 5 項は，

$$I(p \times q) = I(p) + I(q)$$

であることを要請している．このような条件を満足する微分可能な関数は，$I(p) = C \log_e p$しかない（Cは定数）．

【演習】上記を確かめてみよう．

補足

生起確率pのメッセージのもつ情報量を単に$\frac{1}{p}$にせず，それに log をつけて$\log_2 \frac{1}{p}$と定義する理由は，直観的には次のようにも考えられる．一般に，pが与えられたとき，$\log_2 p$はpを 2 進数表示したときの桁数を表す．生起確率pのメッセージの生起確率が2^{-10}であったとき，そのメッセージを 2 進数を用いて表示するのであれば，どのような 2 進数表示を割り当てたらいいだろうか？

そのようなメッセージが他にもあり，それぞれに異なる 2 進数表示を割り当てなければならないとすると，生起確率が2^{-10}のメッセージには 10 桁の 2 進数表示を割り当てるのが素直であるように思える．というのは，他のメッセージもすべて同じく生起確率が2^{-10}とすれば，そのようなメッセージは全部で2^{10}個あることになり，それらを区別するためには，10 桁の 2 進数表示が必要であるからだ．他のメッセージのなかに生起確率が2^{-3}のメッセージがあるとすれば，それを表示するために，そのために 10 桁の 2 進数表示2^7個を確保して，その2^7個の 10 桁の 2 進数表示をひとまとめにして，$\log_2 \frac{2^{10}}{2^7} = 10 - 7 = 3$桁の 2 進数表示にしてしまう．このように，生起確率pのメッセージに長さ$\log_2 \frac{1}{p}$の 2 進数表示を割り当てるようにすると，すべてのメッセージに対する 2 進数表示の平均の桁数（＝長さ）は，そうでない割り当てをした時よりも必ず小さくなることが示される．つまり，各メッセージに対して異なる 2 進数表示を割り当てるという条件を課す限り，それよりも平均桁数を減らすことはできないという意味で，下限を与える．シャノンは確率モデルの下で発生するメッセージに対して有限個の記号を用いて異なる表示を割り当てたとき，その表示の平均長の下限を与える量として，情報量として規定したのである．このことはクラフトの不等式（第 2 回講義）を経て，情報源符号化定理（第 4 回講義）の中で示される．