※ 事象と事象系

- 何通りかの出来事(事象という)が起きる可能性があって、すべての事象についてそれがおこる確率がわかっているとき、それら全体のことを事象系という。

  例 1：事象系：一桁分のデジタル信号を受け取る
  - 事象：「0 を受け取った」「1 を受け取った」

  例 2：事象系：コインを投げる
  - 事象：「表が出た」「裏が出た」

  例 3：事象系：サイコロを投げる
  - 事象：「1 が出た」「2 が出た」...「6 が出た」

➢ 事象系は、事象と確率をまとめて行列の形で次の様に表わされる。

$$X_1 = \begin{bmatrix} x_0 & x_1 \\ 1/2 & 1/2 \end{bmatrix} \quad X_2 = \begin{bmatrix} x_{表} & x_{裏} \\ 1/2 & 1/2 \end{bmatrix} \quad X_3 = \begin{bmatrix} x_1 & x_2 & x_3 & x_4 & x_5 & x_6 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \end{bmatrix}$$

X_1のx_0とx_1はそれぞれ 0, 1 の信号を受け取るという事象を表わす。X_2とX_3に含まれる事象も上の例であげた結果に対応する。どのような事象系も、一般的に次の様に書ける。

$$X = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \\ p_1 & p_2 & \cdots & p_n \end{bmatrix}$$

ここで、$p_1, p_2, \cdots, p_n$は$x_1, x_2, \cdots, x_n$がおこる確率をあらわす。

➢ 事象系に含まれる事象の確率の和は必ず 1 にならなければならない。
  - ➢ × 和が 1 より小さい → 起こりうる事象を網羅していない
  - ➢ × 和が 1 より大きい → 一つの出来事を複数の事象としてカウントしている

練習問題 1.

「ジョーカー2 枚を含む 54 枚のトランプのセットからランダムに 1 枚のカードを引く」ときの結果にかかわる事象系を、上の例で示したような行列の形で記述せよ。ただし、事象は以下のとおりとする。

x_A：エースを引いた

x_N：エース以外の数字のカードを引いた

x_P：絵札を引いた

x_J：ジョーカーを引いた

情報理論 第 2 回　エントロピー

※ エントロピー

- エントロピー：事象系のすべての事象についての情報量を求め、事象系全体で平均値をとったもの
  - 事象系の事象の確率の組み合わせによって決まるので、事象Xの関数という意味で$H(X)$と書く。
  - エントロピーが意味するもの
    - 練習問題 1 の例のような事象系だと、起こった事象によって得られる情報量はことなる。エントロピーとは、何回も試行を繰り返したときに 1 回あたり平均どれくらい情報量が増えるかを表わす量(平均情報量ともいう)。

- すべての事象の確率が同じ場合の平均 ⇒ エントロピー
  - 例 1, 例 2

    0, 1 を受け取る確率 (表と裏が出る確率)はどちらも 1/2

    →0 を受け取った時 (表が出た時) に増える情報量は$-\log\frac{1}{2}=1$

    →1 を受け取った時 (裏が出た時) に増える情報量も$-\log\frac{1}{2}=1$

    →平均の情報量も 1

    →エントロピーは 1
  - 例 3

    どの目が出る確率も 1/6

    →どの目が出た場合も増える情報量は$-\log\frac{1}{6}=2.58$

    →平均の情報量も 2.58

    →エントロピーは 2.58

  つまり、この場合はどれか一つの情報量を求めればそれがエントロピーになる。

- 確率が異なる事象がある場合の平均
  - 当たりが出る確率が 1%のくじを引いて、当たれば 1000 円、はずれなら 10 円もらえる。一回あたり平均して何円もらえるか？
  - 100 回引いた場合のトータルでもらえる金額は

    $$\left.\begin{array}{l}\text{当たり 1 回} \rightarrow 1000\times1=1000\text{ 円}\\ \text{はずれ 99 回} \rightarrow 10\times99=990\text{ 円}\end{array}\right\}\ 1990\text{ 円}$$

    ⇒1 回あたりにすると 1990/100=19.9 円
  - 確率を使って平均値を求める
    - (平均金額) = (当たる確率)×(当たった時の金額) + (外れる確率)×(外れた時の金額)

      = 0.01×1000+0.99×10

      = 10+9.9

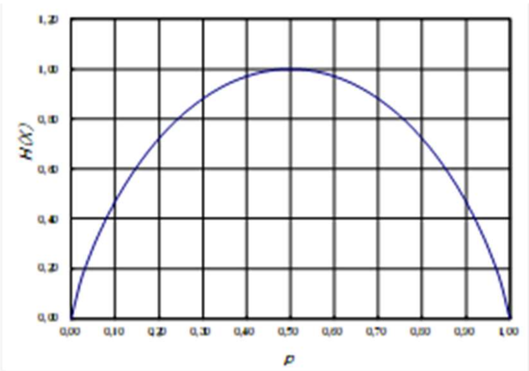
      =19.9
  - もうちょっと一般的に書くと
    - (平均金額) = (確率 1)×(金額 1) + (確率 2)×(金額 2)
  - 福引だと、1 等、2 等、3 等…のようにもっと多くの事象があるが、基本的な形は同じ
    - (平均金額) = (1 等の確率)×(1 等の金額) + (2 等の確率)×(2 等の金額)+…

- ✧ =(確率 1)×(金額 1) + (確率 2)×(金額 2) + …
- ➢ それぞれの場合の確率をp_i、金額をm_i、事象の総数をnとすると、平均金額は次の様に書ける。
  - ✧ $\bar{m} = p_1 m_1 + p_2 m_2 + \cdots + p_n m_n = \sum_{i=1}^{n} p_i m_i$
- ➢ 金額mを情報量Iに読み替えると、エントロピー＝平均情報量は次の様になることがわかる。
  - ✧ $H = \bar{I} = p_1 I_1 + p_2 I_2 + \cdots + p_n I_n = \sum_{i=1}^{n} p_i I_i = \sum_{i=1}^{n} p_i(-\log p_i) = -\sum_{i=1}^{n} p_i \log p_i$

- ● 事象系の事象が2通りだけの場合
  - ➢ 2つの事象の確率を足すと1になる　⇒　一方をもう一方で表わせる

$$X_1 = \begin{bmatrix} x_1 & x_2 \\ p & 1-p \end{bmatrix}$$

  - ➢ エントロピーは次のようになる
    - ✧ $H(X) = -p\log p - (1-p)\log(1-p)$
  - ➢ pを 0～1 の範囲で変えると、エントロピーは図のように変化する



- ● エントロピーは「不確定さ」をあらわす
  - ➢ 事象系「表が出る確率がpのコインを投げる」
    - ✧ pが 0.5：結果が出るまで表か裏かまったくわからない→不確定さが大きい
    - ✧ pが 0.1：結果が出る前でも、「裏」といっておけばけっこう当たる→不確定さが小さい
    - ✧ pが 0：結果が出る前でも、「裏」といっておけば絶対に当たる→不確定さが 0

確率がどちらかに偏っていると「予想が当たりやすい」⇒　エントロピーが小さい

練習問題 2.
表が出る確率が 1/3 のコインを投げ、「表が出た」「裏が出た」という事象からなる事象系のエントロピーを求めよ。ただし、$\log 3 = 1.58$とし、計算結果は四捨五入して小数第二位までにすること。

練習問題 3.
表が出る確率が 1/4 のコインを投げ、「表が出た」「裏が出た」という事象からなる事象系のエントロピーを求めよ。ただし、$\log 3 = 1.58$とし、計算結果は四捨五入して小数第二位までにすること。

練習問題 4.
表が出る確率が 1/5 のコインを投げ、「表が出た」「裏が出た」という事象からなる事象系のエントロピーを求めよ。ただし、$\log 5 = 2.32$とし、計算結果は四捨五入して小数第二位までにすること。

※　練習問題 2～4 の結果を比較すると、どのようなことがいえるか？