通信工学概論

Introduction to Communication Engineering

第7回講義資料

Lecture notes 7

情報量

Shannon Entropy

豊橋技術科学大学

Toyohashi University of Technology

電気・電子情報工学系

Department of Electrical and Electronic Information Engineering

准教授　竹内啓悟

Associate Professor　Keigo Takeuchi

情報とは何か？(What is information?)

情報理論では、情報が何を表現しているかは議論しない。
情報を表現するのに必要な文字列の長さを議論する。

In information theory, we do not discuss what information represents. We discuss the length of a string required for representing the information.

Q1：豊橋技術科学大学を卒業しましたか？(Did you graduated from TUT?)

A1：0

はい(Yes)=0、いいえ(No)=1

Q2：あなたは今日の朝食でご飯を食べましたか？(Did you eat rice this morning?)

A1：1

A1の方が重要な質問に対する答えであるが、質問はどちらも二択なので、答えの情報量はどちらも同じとみなす。

A1 is an answer to more important question than A2. However, the answers have the same amount of information as each other since both questions are two choices.

複数の確率変数(Multiple random variables)

3個の離散確率変数(X, Y, Z)を同時に取り扱いたい。

Treat three discrete random variables (X, Y, Z) jointly.

同時確率(Joint probability)

三つの離散確率変数(X, Y, Z)が(x, y, z)を取る確率$p_{x,y,z} \geq 0$を定める。

Determine the probability $p_{x,y,z}$ with which three discrete random variables (X, Y, Z) take (x, y, z).

$$\mathbb{P}(X = x, Y = y, Z = z) = p_{x,y,z}, \qquad \sum_{(x,y,z)} p_{x,y,z} = 1$$

ただし、総和は(x, y, z)が取りうるすべての組み合わせに関して取られる。

The summation is over all possible combinations of (x, y, z).

例7.1 (Example 7.1)

三つの確率変数は$(0,0,0)$、$(0,0,1)$、$(0,1,0)$、$(0,1,1)$、$(1,0,0)$、$(1,0,1)$、$(1,1,0)$、$(1,1,1)$を確率$1/8$で取る。

The three random variables take $(0,0,0)$, $(0,0,1)$, $(0,1,0)$, $(0,1,1)$, $(1,0,0)$, $(1,0,1)$, $(1,1,0)$, $(1,1,1)$ with probability $1/8$.

周辺確率(Marginal probability)

$$\mathbb{P}(X,Y)=\sum_{z}\mathbb{P}(X,Y,Z=z),\qquad \mathbb{P}(X)=\sum_{y}\mathbb{P}(X=x,Y=y)$$

総和はzまたはyが取りうるすべての値に関して取られる。

The summation is over all possible values of z or y.

他の確率変数の組み合わせに関する周辺確率も同様に定義される。

We define marginal probability with respect to the other combinations of the random variables similarly.

周辺確率の計算(Computation of the marginal probability)

例7.1の場合に周辺確率を計算する。(Compute marginal probability in Example 7.1.)

$$\mathbb{P}(X=0,Y=0)=\mathbb{P}(X=0,Y=0,Z=0)+\mathbb{P}(X=0,Y=0,Z=1)=\frac{1}{4}$$

同様に、

Similarly,

$$\mathbb{P}(X=0,Y=1)=\mathbb{P}(X=1,Y=0)=\mathbb{P}(X=1,Y=1)=\frac{1}{4}$$

さらに、

Furthermore,

$$\mathbb{P}(X=0)=\mathbb{P}(X=0,Y=0)+\mathbb{P}(X=0,Y=1)=\frac{1}{2}$$

条件付き確率(Conditional probability)

$$\mathbb{P}(Y,Z|X)=\frac{\mathbb{P}(X,Y,Z)}{\mathbb{P}(X)},\qquad \mathbb{P}(Z|X,Y)=\frac{\mathbb{P}(X,Y,Z)}{\mathbb{P}(X,Y)}$$

他の確率変数の組み合わせに関する条件付き確率も同様に定義される。

We define conditional probability with respect to the other combinations of the random variables similarly.

条件付き確率の計算(Computation of the conditional probability)

例7.1の場合に条件付き確率を計算する。

Compute conditional probability in Example 7.1.

$$\mathbb{P}(Y=0,Z=0|X=0)=\frac{\mathbb{P}(X=0,Y=0,Z=0)}{\mathbb{P}(X=0)}=\frac{1}{4}$$

$$\mathbb{P}(Z=0|X=0,Y=0)=\frac{\mathbb{P}(X=0,Y=0,Z=0)}{\mathbb{P}(X=0,Y=0)}=\frac{1}{2}$$

この場合、

In this case,

$$\mathbb{P}(Z|X,Y)=\mathbb{P}(Z)$$

独立性(Independence)

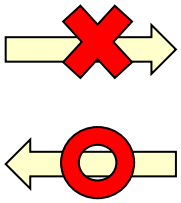
同時確率が各確率変数の周辺確率の積に分解されるとき、
確率変数(X, Y, Z)は独立と呼ばれる。

Random variables (X, Y, Z) are said to be independent if the joint probability is decomposed into the product of the individual marginal probability of the random variables.

$$\mathbb{P}(X, Y, Z) = \mathbb{P}(X)\mathbb{P}(Y)\mathbb{P}(Z)$$

注意 $\mathbb{P}(X, Y) = \mathbb{P}(X)\mathbb{P}(Y), \mathbb{P}(X, Z) = \mathbb{P}(X)\mathbb{P}(Z), \mathbb{P}(Y, Z) = \mathbb{P}(Y)\mathbb{P}(Z)$

Remark



$$\mathbb{P}(X, Y, Z) = \mathbb{P}(X)\mathbb{P}(Y)\mathbb{P}(Z)$$

独立性の意義(Significance of independence)

(X, Y, Z)が独立のとき、条件付き確率は条件に依存しない。

If (X, Y, Z) is independent, the conditional probability does not depend on any conditioning.

$$\mathbb{P}(Y, Z|X) = \frac{\mathbb{P}(X, Y, Z)}{\mathbb{P}(X)} = \frac{\mathbb{P}(X)\mathbb{P}(Y)\mathbb{P}(Z)}{\mathbb{P}(X)} = \mathbb{P}(Y)\mathbb{P}(Z) = \mathbb{P}(Y, Z)$$

二番目と四番目の等号は、独立性の定義から従う。

The second and fourth equalities follow from the definition of the independence.

期待値(Expectation)

決定論的な関数$f(x,y,z)$に対して、(For a deterministic function $f(x,y,z)$,)

$$\mathbb{E}[f(X,Y,Z)] = \sum_{(x,y,z)} f(x,y,z)\mathbb{P}(X=x,Y=y,Z=z)$$

総和は(x,y,z)が取りうるすべての組み合わせに関して取られる。

The summation is over all possible combinations of (x,y,z).

条件付き期待値(Conditional expectation)

決定論的な関数$g(y,z)$と$h(z)$に対して、(For deterministic functions $g(y,z)$ and $h(z)$,)

$$\mathbb{E}[g(Y,Z)|X=x] = \sum_{(y,z)} g(y,z)\mathbb{P}(Y=y,Z=z|X=x)$$

$$\mathbb{E}[h(Z)|X=x,Y=y] = \sum_{z} h(z)\mathbb{P}(Z=z|X=x,Y=y)$$

総和は(y,z)またはzが取りうるすべての組み合わせに関して取られる。

The summation is over all possible combinations of (y,z) or z.

連続確率変数の場合(Case of continuous random variables)

3個の連続確率変数(X, Y, Z)を同時に取り扱いたい。

Treat three continuous random variables (X, Y, Z) jointly.

離散の場合の確率と総和をそれぞれ確率密度関数と積分に置き換える。

Replace the probability and summation in the discrete case with a probability density function and an integral, respectively.

同時確率(Joint probability)

(X, Y, Z)が3次元空間内の集合Aに入る確率は以下で定義される。

We define the probability with which (X, Y, Z) is included into a set A in the 3-dimensional space.

$$\mathbb{P}\big((X, Y, Z) \in A\big) = \int_{(x,y,z)\in A} p_{X,Y,Z}(x, y, z)dxdydz$$

ただし、同時確率密度関数$p_{X,Y,Z}(x, y, z)$は以下を満たす。

Here, the joint probability density function $p_{X,y,Z}(x, y, z)$ satisfies the following:

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p_{X,Y,Z}(x, y, z)dxdydz = 1$$

連続確率変数の場合(Case of continuous random variables)

周辺確率密度関数(Marginal probability density functions)

$$p_{X,Y}(x,y) = \int_{-\infty}^{\infty} p_{X,Y,Z}(x,y,z)dz\,, \qquad p_X(x) = \int_{-\infty}^{\infty} p_{X,Y}(x,y)dy$$

条件付き確率密度関数(Conditional probability density functions)

$$p_{Y,Z|X}(y,z|x) = \frac{p_{X,Y,Z}(x,y,z)}{p_X(x)}, \qquad p_{Z|X,Y}(z|x,y) = \frac{p_{X,Y,Z}(x,y,z)}{p_{X,Y}(x,y)}$$

独立性(Independence)

同時確率密度関数が各確率変数の周辺確率密度関数の積に分解されるとき、確率変数(X,Y,Z)は独立と呼ばれる。

Random variables (X,Y,Z) are said to be independent if the joint probability density function is decomposed into the product of the individual marginal probability density functions of the random variables.

$$p_{X,Y,Z}(x,y,z) = p_X(x)p_Y(y)p_Z(z)$$

連続確率変数の場合(Case of continuous random variables)

期待値(Expectation)

決定論的な関数$f(x,y,z)$に対して、(For a deterministic function $f(x,y,z)$,)

$$\mathbb{E}[f(X,Y,Z)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y,z)p_{x,y,z}(x,y,z)dxdydz$$

条件付き期待値(Conditional expectation)

決定論的な関数$g(y,z)$と$h(z)$に対して、(For deterministic functions $g(y,z)$ and $h(z)$,)

$$\mathbb{E}[g(Y,Z)|X=x] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(y,z)p_{Y,Z|X}(y,z|x)dydz$$

$$\mathbb{E}[h(Z)|X=x,Y=y] = \int_{-\infty}^{\infty} h(z)p_{Z|X,Y}(z|x,y)dz$$

期待値の性質(Properties of expectation)

連続確率変数を仮定するが、離散の場合にも同じ性質が成り立つ。

Continuous random variables are assumed. However, the same properties hold for the discrete case.

一貫性(Consistency)

同時確率密度関数$p_{X,Y,Z}$に関するXの期待値は、周辺確率密度関数p_Xに関するXの期待値と一致する。

The expectation of X with respect to a joint probability density function $p_{X,Y,Z}$ is equal to that of X with respect to the corresponding marginal probability density function p_X.

$$\because \quad \mathbb{E}[X] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x p_{X,Y,Z}(x,y,z)\,dxdydz$$

$$= \int_{-\infty}^{\infty} x \left\{ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p_{X,Y,Z}(x,y,z)\,dydz \right\} dx = \int_{-\infty}^{\infty} x p_X(x)\,dx$$

最後の等号は、周辺確率密度関数の定義から従う。

The last equality follows from the definition of the marginal probability density function. ■

期待値の性質(Properties of expectation)

線形性(Linearity)

実数aとbに対して、(For real numbers a and b,)

$$\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$$

XとYに関する何の条件も不要なことに注意

Note that X or Y does not require any conditions.

$\because$ 積分の線形性から、(From the linearity in integral,)

$$\mathbb{E}[aX + bY] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(ax + by)p_{X,Y}(x,y)dxdy$$

$$= a\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}xp_{X,Y}(x,y)dxdy + b\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}yp_{x,y}(x,y)dxdy = a\mathbb{E}[X] + b\mathbb{E}[Y]$$

最後の等号は、期待値の一貫性から従う。

The last equality follows from the consistency in expectation. ■

期待値の性質(Properties of expectation)

確率変数の積(Product of random variables)

fとgを決定論的な関数とする。(Let f and g denote deterministic functions.)

(X,Y)が独立ならば、(If (X,Y) is independent,)

$$\mathbb{E}[f(X)g(Y)] = \mathbb{E}[f(X)]\mathbb{E}[g(Y)]$$

積の場合は、無条件で期待値の分解はできない。

In the case of the product, the expectation cannot be decomposed with no conditions.

∵ 独立性の定義$p_{X,Y}(x,y) = p_X(x)p_Y(y)$から、

From the definition of the independence $p_{X,Y}(x,y) = p_X(x)p_Y(y)$,

$$\mathbb{E}[f(X)g(Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x)g(y)p_{X,Y}(x,y)dxdy$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x)g(y)p_X(x)p_Y(y)dxdy$$

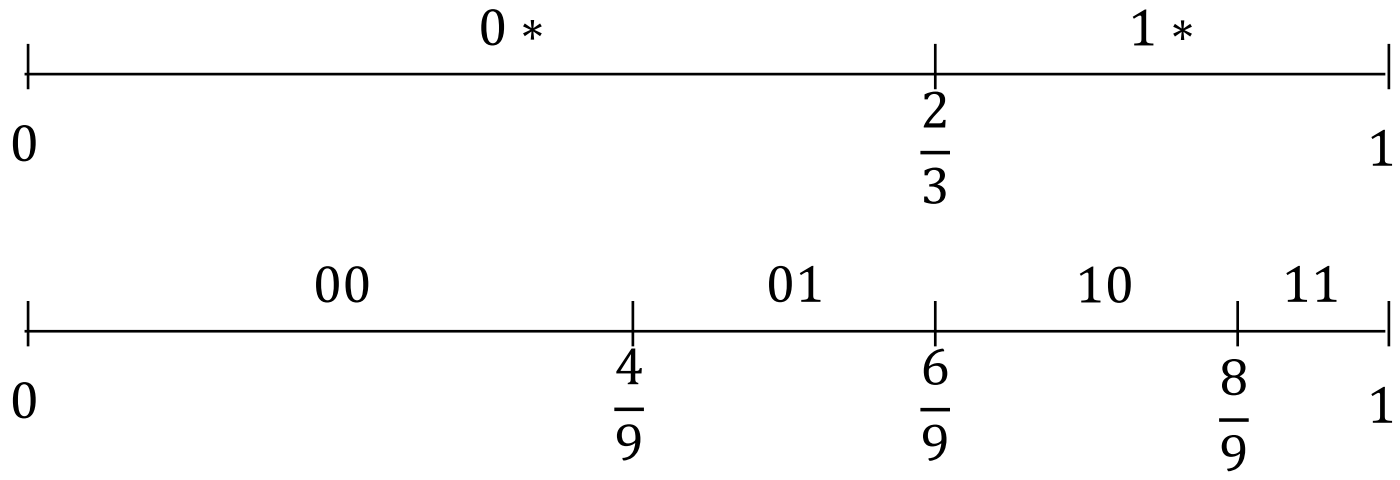
$$= \int_{-\infty}^{\infty} f(x)p_X(x)dx \int_{-\infty}^{\infty} g(y)p_Y(y)dy = \mathbb{E}[f(X)]\mathbb{E}[g(Y)] \quad \blacksquare$$

算術符号化(Arithmetic coding)

0または1を取る長さ2の二進系列Sを考える。ただし、各位置で0が発生する確率を$2/3$とし、異なる位置の数字は独立に発生するものとする。

Consider a binary sequence S of length 2 that takes 0 or 1. Assume that 0 occurs with probability $2/3$ in each position and that 0 or 1 occurs independently in different positions.

区間分割(Interval division)



系列の発生確率が対応する区間の長さに等しくなるように分割する。

The occurrence probability of each sequence is equal to the length of the corresponding interval.

算術符号化(Arithmetic coding)

符号化(Encoding)

各入力系列$\boldsymbol{s}$に対して、$m_{\boldsymbol{s}} = \lceil -\log_2 \mathbb{P}(\boldsymbol{S} = \boldsymbol{s}) \rceil + 1$とする。$\boldsymbol{s}$の符号語として、対応する区間の中点の2進数表現を$m_{\boldsymbol{s}}$ビットで打ち切ったものを選ぶ。

For each input sequence $\boldsymbol{s}$, let $m_{\boldsymbol{s}} = \lceil -\log_2 \mathbb{P}(\boldsymbol{S} = \boldsymbol{s}) \rceil + 1$. As a codeword of $\boldsymbol{s}$, we select the sequence obtained by truncating the binary representation of the midpoint in the corresponding interval up to $m_{\boldsymbol{s}}$ bits.

$\lceil a \rceil$は実数a以上の最小の整数($\lceil a \rceil$ is the minimum integer greater than or equal to a.)

復号 Decoding

各符号語が対応する区間に入っていれば、復号できる。(詳細省略)

If each codeword is included into the corresponding interval, decoding is possible.

平均符号長$\mathbb{E}[L]$(Average code length $\mathbb{E}[L]$)

$\boldsymbol{S} = S_1 S_2$、$\boldsymbol{s} = s_1 s_2$とすると、$\mathbb{P}(\boldsymbol{S} = \boldsymbol{s}) = \mathbb{P}(S_1 = s_1)\mathbb{P}(S_2 = s_2)$なので、

For $\boldsymbol{S} = S_1 S_2$ and $\boldsymbol{s} = s_1 s_2$, we have $\mathbb{P}(\boldsymbol{S} = \boldsymbol{s}) = \mathbb{P}(S_1 = s_1)\mathbb{P}(S_2 = s_2)$. Thus,

$$\mathbb{E}[L] = \sum_{\boldsymbol{s}} m_{\boldsymbol{s}} \mathbb{P}(\boldsymbol{S} = \boldsymbol{s}) \leq -\sum_{i=1,2} \sum_{s_i = 0,1} \mathbb{P}(S_i = s_i) \log_2 \mathbb{P}(S_i = s_i) + 2$$

不等式の導出で$\lceil a \rceil \leq a + 1$を使った。(In the derivation of the inequality, $\lceil a \rceil \leq a + 1$ has been used.)

算術符号化(Arithmetic coding)

符号語が対応する区間に含まれていることを確認せよ。

Confirm that each codeword is included into the corresponding interval.

入力、符号語、区間を入力の発生確率が大きい順にそれぞれ$\boldsymbol{s}^{(1)},\dots,\boldsymbol{s}^{(4)}$、$c_1,\dots,c_4$、$[L_1,R_1),\dots,[L_4,R_4)$とする。$c_i > L_i$を証明したい。
$m_{\boldsymbol{s}^{(i)}}$を2進数表現$c_i = 0.b_1^{(i)}\cdots b_{m_{\boldsymbol{s}^{(i)}}}^{(i)}$のビット長とすると、$c_i$の定義から、

Let $\boldsymbol{s}^{(1)},\dots,\boldsymbol{s}^{(4)}$, $c_1,\dots,c_4$, and $[L_1,R_1),\dots,[L_4,R_4)$ denote the inputs, codewords, and intervals arranged in descending order with respect to the input occurrence probability, respectively. Prove $c_i > L_i$. Let $m_{\boldsymbol{s}^{(i)}}$ be the bit length of the binary representation $c_i = 0.b_1^{(i)}\cdots b_{m_{\boldsymbol{s}^{(i)}}}^{(i)}$. From the definition of c_i,

$$\frac{L_i+R_i}{2} - c_i < \sum_{j=m_{\boldsymbol{s}^{(i)}}+1}^{\infty} 2^{-j} = 2^{-m_{\boldsymbol{s}^{(i)}}} = 2^{-\left(\lceil -\log_2 \mathbb{P}\left(\boldsymbol{S}=\boldsymbol{s}^{(i)}\right)\rceil+1\right)} \le \frac{\mathbb{P}\left(\boldsymbol{S}=\boldsymbol{s}^{(i)}\right)}{2}$$

$$\because -\lceil a\rceil \le -a \text{ for } a>0$$

区間の長さ$R_i - L_i = \mathbb{P}\left(\boldsymbol{S}=\boldsymbol{s}^{(i)}\right)$を使うと、

Using the interval length $R_i - L_i = \mathbb{P}\left(\boldsymbol{S}=\boldsymbol{s}^{(i)}\right)$ yields

$$c_i > \frac{L_i+R_i}{2} - \frac{\mathbb{P}\left(\boldsymbol{S}=\boldsymbol{s}^{(i)}\right)}{2} = \frac{L_i+R_i}{2} - \frac{R_i-L_i}{2} = L_i \quad \blacksquare$$

情報量(The amount of information)

エントロピー(Entropy)

離散確率変数Xに対して、Xのエントロピー$H(X)$を以下で定義する。

For a discrete random variable X, the entropy $H(X)$ of X is defined as

$$H(X) = -\sum_{x} \mathbb{P}(X = x) \log_2 \mathbb{P}(X = x), \qquad 0 \log_2 0 = 0$$

総和はxの取りうる値全体に及ぶ。(The summation is over all possible values of x.)

順定理(Direct theorem)

長さNの独立同一分布する入力ビット系列$\boldsymbol{S} = S_1 \cdots S_N$の場合に
算術符号化を適用すると、平均符号長$\mathbb{E}[L_N]$は以下を満たす。

Applying the arithmetic coding to an independent and identically distributed input bit sequence $\boldsymbol{S} = S_1 \cdots S_N$ of length N, we find that the average code length $\mathbb{E}[L_N]$ satisfies

$$\frac{1}{N}\mathbb{E}[L_N] \leq -\sum_{s=0,1} \mathbb{P}(S_1 = s) \log_2 \mathbb{P}(S_1 = s) + \frac{2}{N} \to H(S_1) \text{ as } N \to \infty$$

情報量(The amount of information)

逆定理(Converse theorem)

いかなる符号化をしても、平均符号長$\mathbb{E}[L_N]$は以下の不等式を満たす。

For any coding method, the average code length $\mathbb{E}[L_N]$ satisfies the following inequality:

$$\lim_{N\to\infty}\frac{1}{N}\mathbb{E}[L_N] \geq H(S_1)$$

証明は情報理論の講義を参照(Learn Information Theory for the proof.)

順定理と逆定理を合わせて、情報源符号化定理と呼ぶ。

The combination of the direct and converse theorems is called source coding theorem.

エントロピーの意味(Meaning of the entropy)

エントロピーは、独立同一分布する無限に長いビット系列を圧縮したときの圧縮率の理論限界に等しい。

The entropy is equal to the theoretical limit of the compression rate in compressing an infinitely long bit sequence that has independent and identically distributed elements.

エントロピーはデータの本質的な情報量（不確かさ）を表す。

The entropy means the amount of essential information (uncertainty) on data.

エントロピーの計算(Computation of entropy)

例7.1 (Example 7.1)

離散確率変数Xは、確率pで0を取り、確率$1-p$で1を取るとする。Xのエントロピーを計算せよ。さらに、エントロピーが最大になるpを答えよ。

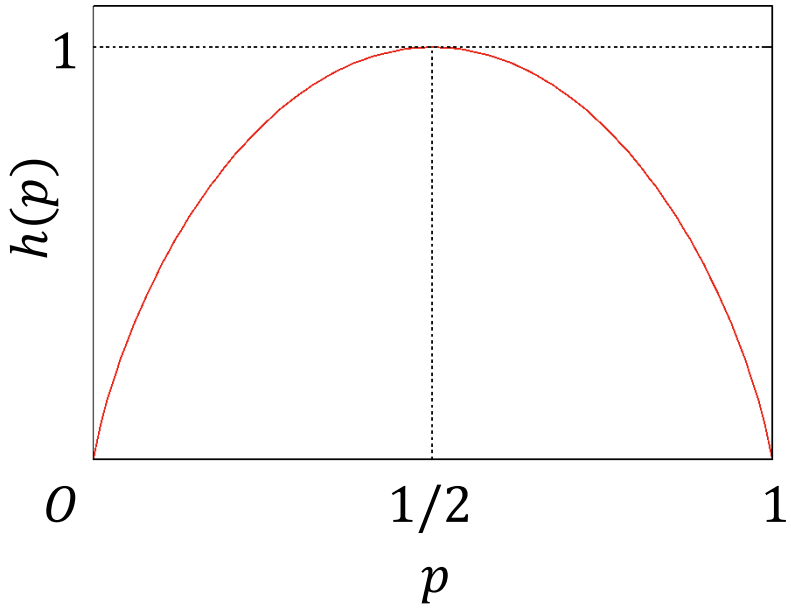
A discrete random variable X takes 0 and 1 with probability p and $1-p$, respectively. Compute the entropy of X. Furthermore, answer p maximizing the entropy.

$$H(X) = -\mathbb{P}(X=0)\log_2 \mathbb{P}(X=0) - \mathbb{P}(X=1)\log_2 \mathbb{P}(X=1)$$

$$= -p\log_2 p - (1-p)\log_2(1-p) \equiv h(p)$$

関数$h(p)$は$p = 1/2$のときに最大値1を取る。

The function $h(p)$ takes the maximum 1 at $p = 1/2$.



演習(Exercise)

離散確率変数Xは、-1、0、1をそれぞれ確率$1/6$、$1/3$、$1/2$で取る。
Xのエントロピーを計算せよ。

A discrete random variable X takes -1, 0, and 1 with probability $1/6$, $1/3$, and $1/2$, respectively. Compute the entropy of X.