

東京大学大学院新領域創成科学研究科 基盤情報学専攻 入試
平成 17 年 専門科目 第 5 問

(1)

0 の後には、 $\frac{7}{8}$ の確率で 0 が続き、 $\frac{1}{8}$ の確率で 1 が続く。また、1 の後には、 $\frac{1}{2}$ の確率で 0 が続き、 $\frac{1}{2}$ の確率で 1 が続く。したがって、0 の出現する確率 $p(0)$ と、1 の出現する確率 $p(1)$ は、以下のように表わせる。

$$p(0) = p(0)p(0|0) + p(1)p(0|1)$$

$$p(1) = p(0)p(1|0) + p(1)p(1|1)$$

ここで、 $p(0|0)$ 、0 の次に 0 が続くという遷移確率であり、他も同様である。

遷移確率に値を代入すると、

$$p(0) = \frac{7}{8}p(0) + \frac{1}{2}p(1)$$

$$p(1) = \frac{1}{8}p(0) + \frac{1}{2}p(1)$$

であり、これは、 $p(0)$ と $p(1)$ についての連立方程式である。これを解くことで、求める確率が得られる。

$$p(0) = \frac{4}{5}$$

$$p(1) = \frac{1}{5}$$

(2)

0 の時のエントロピー H_0 は、

$$H_0 = -\frac{7}{8} \log_2 \frac{7}{8} - \frac{1}{8} \log_2 \frac{1}{8}$$

1 の時のエントロピー H_1 は、

$$H_1 = -\frac{1}{2} \log_2 \frac{1}{2} - \frac{1}{2} \log_2 \frac{1}{2}$$

である。

情報源のエントロピー H は、

$$H = p(0)H_0 + p(1)H_1 = 0.633$$

対数

自然対数は、

$$\ln x = \log_e x$$

常用対数は、

$$\log x = \log_{10} x$$

情報理論で用いる対数は、底を 2 とした対数であり、

$$\log_2 x$$

である。自然対数でも常用対数でもない対数というのは普段あまり使わないので、注意する。

たとえば、

$$\log_{10} 2 = 0.3010$$

などは良く知られているので使ってしまいそうになるが、情報理論では、

$$\log_2 2 = 1$$

である。

(3)

各シンボルを 0 と 1 で表現し直すと、

- (0, 1) \rightarrow 1
- (1, 1) \rightarrow 01
- (2, 1) \rightarrow 001
- (3, 1) \rightarrow 0001
- (3, 0) \rightarrow 0000

それぞれの出現確率は、遷移確率と (1) で求めた確率から求めることができる。

$$p(0, 1) = p(1) = \frac{1}{5}$$

$$p(1, 1) = p(0) \cdot p(1|0) = \frac{1}{10}$$

$$p(2, 1) = p(0) \cdot p(0|0) \cdot p(1|0) = \frac{7}{80}$$

$$p(3, 1) = p(0) \cdot p(0|0)^2 \cdot p(1|0) = \frac{49}{640}$$

$$p(3, 0) = p(0) \cdot p(0|0)^3 = \frac{343}{640}$$

(4)

ハフマンの符号化法を用いて符号化すると、

- $(0, 1) \rightarrow 00$
- $(1, 1) \rightarrow 010$
- $(2, 1) \rightarrow 0110$
- $(3, 1) \rightarrow 0111$
- $(3, 0) \rightarrow 1$

(5)

1 シンボルあたりの符号長、すなわち平均符号長は、

$$2p(0, 1) + 3p(1, 1) + 4p(2, 1) + 4p(3, 1) + 1p(3, 0) \doteq 1.892$$

である。

平均符号長とエントロピーの関係

エントロピーとの比較？分かりません。