

情報理論レポート 1 (情報量)

提示：2010/10/6(水) 提出：2010/10/20(水)

注：A4 の用紙で作成し、ホチキス等できちんと左上をとめて提出すること。

1. 英文の情報量

26 文字からなるアルファベット $\mathcal{A} = \{a, b, \dots, z\}$ の各記号 $\alpha \in \mathcal{A}$ の出現確率 $P(\alpha)$ は、次の表で与えられる。(<http://www7.plala.or.jp/dvorakjp/hinshutu.htm> 参照。)

α	<i>e</i>	<i>a</i>	<i>t</i>	<i>i</i>	<i>o</i>	<i>s</i>	<i>n</i>	<i>r</i>
$P(\alpha)$	0.1183	0.0876	0.0849	0.0740	0.0731	0.0723	0.0702	0.0645
α	<i>h</i>	<i>l</i>	<i>d</i>	<i>c</i>	<i>u</i>	<i>m</i>	<i>p</i>	<i>f</i>
$P(\alpha)$	0.0441	0.0402	0.0402	0.0331	0.0306	0.0277	0.0210	0.0206
α	<i>g</i>	<i>y</i>	<i>w</i>	<i>b</i>	<i>v</i>	<i>k</i>	<i>j</i>	<i>x</i>
$P(\alpha)$	0.0189	0.0186	0.0176	0.0171	0.0104	0.0093	0.0021	0.0018
α	<i>q</i>	<i>z</i>						
$P(\alpha)$	0.0009	0.0009						

この出現確率 $P(\alpha)$ 基づいた情報源を

$$A = \left\{ \begin{array}{cccc} a & , & b & , & \dots & , & z \\ P(a) & , & P(b) & , & \dots & , & P(z) \end{array} \right\}$$

とし、

アルファベット中の各記号 $\alpha \in \mathcal{A}$ が均等に現れる情報源を

$$B = \left\{ \begin{array}{cccc} a & , & b & , & \dots & , & z \\ \frac{1}{26} & , & \frac{1}{26} & , & \dots & , & \frac{1}{26} \end{array} \right\}$$

とする。

また、単語 w が情報源 A から生成される確率を $P_A(w)$ 、自己情報量を $I_A(w)$ と表し、情報源 B から生成されたときの確率を $P_B(w)$ 、自己情報量を $I_B(w)$ と表す。

このとき次の問いに答えよ。

(1)

$w_1 = \text{information}$ 、 $w_2 = \text{theory}$ 、 $w_3 = \text{shannon}$ とし、次の確率および自己情報量を有効数字 3 桁で求めよ。

$$\begin{array}{l} P_A(w_1), I_A(w_1), P_B(w_1), I_B(w_1), \\ P_A(w_2), I_A(w_2), P_B(w_2), I_B(w_2), \\ P_A(w_3), I_A(w_3), P_B(w_3), I_B(w_3) \end{array}$$

(2) 情報源 A のエントロピー (平均情報量) $H(A)$ および情報源 B のエントロピー $H(B)$ をそれぞれ求めよ。