平成27年度後期
工学部・情報工学科

情報理論

第1回小テスト（木曜2限クラス）
問題と解答例（40点満点）

2015. 10. 22

＜注意事項＞

- 教科書，資料等の持ち込み不可．電卓専用機使用可．
- 対数の数値は問題4の（参考）を参照のこと．
- 解答は有効数字3桁（4桁目を四捨五入）で示すこと．
- 問題用紙は回収しません．持ち帰ってください．

問題1（3点×3題＝9点）

胃痛（B）の原因として，ストレス（A_1），胃潰瘍（A_2），胃ガン（A_3）が考えられる．次のデータが分かっているとき，胃痛の原因をベイズ定理により確率的に求めよ．

＜事前確率：原因が生じる確率＞
$P(A_1) = 30\,\%$, $P(A_2) = 60\,\%$, $P(A_3) = 10\%$
＜原因から結果が生じる確率＞
$P(B|A_1) = 50\,\%, P(B|A_2) = 30\,\%, P(B|A_3) = 20\,\%$

＜求める確率：結果から推定される原因の確率＞
$P(A_1|B)$, $P(A_2|B)$, $P(A_3|B)$

＜解答例＞

＜事象＞
B 胃痛
A_1 ストレス
A_2 胃潰瘍
A_3 胃ガン
＜事前確率＞
$P(A_1) = 0.3$
$P(A_2) = 0.6$
$P(A_3) = 0.1$
＜原因→結果の確率＞
$P(B|A_1) = 0.5$
$P(B|A_2) = 0.3$
$P(B|A_3) = 0.2$

ベイズの定理による推定結果

（分母）$= 0.35$
$P(A_1)P(B|A_1) + P(A_1)P(B|A_1)$
$+P(A_1)P(B|A_1) + P(A_1)P(B|A_1)$

（分子）
$P(A_1|B) \to P(A_1)P(B|A_1) = 0.15$
$P(A_2|B) \to P(A_2)P(B|A_2) = 0.18$
$P(A_3|B) \to P(A_3)P(B|A_3) = 0.02$

＜結果→原因の確率＞
$P(A_1|B) = 0.429 = 42.9\,\%$
$P(A_2|B) = 0.514 = 51.4\,\%$
$P(A_3|B) = 0.0571 = 5.71\,\%$

問題2（10点）

二つのサイコロを振ったとき，その目の和が8であり，サイコロの目も分かっていた．後日，そのサイコロの目を忘れてしまった．このとき失われた情報量（ビット）を求めよ．

＜解答例＞
二つのサイコロの目の組合せは$6 \times 6 = 36$通りである．

①初めは，目の和が8であり，目の組み合わせも分かっているので，目の組合せは1通りである．
確率：$p_1 = 1/36$
自己情報量：$I_1 = -\log_2 p_1 = 5.17[bit]$

②次に，目の和が8であるが，目の組み合わせが不明であるので，次の5通りの組合せがある．
（2，6），（3，5），（4，4），（5，3），（6，2）
確率：$p_2 = 5/36$
自己情報量：$I_2 = -\log_2 p_2 = 2.85[bit]$

③失われた情報量： $I = I_1 - I_2 = 2.32\ [bit]$

問題3（10点）

あるイベントの入場者数は20,000人であった．その内訳は以下のようになっていた．結合エントロピーを求めよ．

（世代別）
30才未満：12,000人，30才以上：8,000人
（地域別）
関東圏：16,000人，関東圏以外：4,000人

＜解答例＞
事象a_1：30才未満 $p_{a1} = 12/20 = 3/5$
事象a_2：30才以上 $p_{a2} = 8/20 = 2/5$
事象b_1：関東圏 $p_{b1} = 16/20 = 4/5$
事象b_2：関東圏以外 $p_{b2} = 4/20 = 1/5$

結合事象の確率

$$(a_1, b_1) \rightarrow \quad p_{11} = \frac{3}{5} \times \frac{4}{5} = \frac{12}{25}$$
$$(a_1, b_2) \rightarrow \quad p_{12} = \frac{3}{5} \times \frac{1}{5} = \frac{3}{25}$$
$$(a_2, b_1) \rightarrow \quad p_{21} = \frac{2}{5} \times \frac{4}{5} = \frac{8}{25}$$
$$(a_2, b_2) \rightarrow \quad p_{22} = \frac{2}{5} \times \frac{1}{5} = \frac{2}{25}$$

これらの確率を結合エントロピーの式に代入する.

$$H(A) = -\sum_{i,j=1}^{2} p_{ij} \log_2 p_{ij}$$
$$= -\frac{12}{25}\log_2\frac{12}{25} - \frac{3}{25}\log_2\frac{3}{25} - \frac{8}{25}\log_2\frac{8}{25} - \frac{2}{25}\log_2\frac{2}{25}$$
$$= 1.69 \quad [bit]$$

問題4（4＋4＋3＝11点）

2人の学生の20科目の成績を以下に示す．2人の成績のエントロピー$H(A), H(B)$を求めよ．さらに，2人のエントロピーの値の違いについて考察せよ（エントロピーの意味と成績分布に基づいて違いを説明する）

成績	S	A	B	C
A君	2	10	7	1
B君	6	5	4	5

（参考）
$\log_2 3 = 1.58, \quad \log_2 5 = 2.32, \quad \log_2 7 = 2.81$

＜解答例＞

A君： $p_S = 2/20,\ p_A = 10/20,\ p_B = 7/20,\ p_C = 1/20$

$$H(A) = \sum_{i=(S,A,B,C)} -p_i \log_2 p_i = 1.58\ [bit]$$

B君： $p_S = 6/20,\ p_A = 5/20,\ p_B = 4/20,\ p_C = 5/20$

$$H(B) = \sum_{i=(S,A,B,C)} -p_i \log_2 p_i = 1.99\ [bit]$$

$H(A) < H(B)$の理由
エントロピーは曖昧さを表す尺度である．
A君の成績はA，Bに集中しており，予測し易い（曖昧さが小さい）．B君の成績はS～Cに万遍なく分布しており，予測が難しい（曖昧さが大きい）．