平成28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1限クラス)
(問題と解答例／50点満点)

2016. 10. 20

＜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
- 計算式を示すこと．
- 対数の数値は問題5(参考)を参照のこと．
- 解答は分数(既約)または小数(有効数字3桁以内or小数点以下3桁以内)で示すこと．

＜問題用紙を試験終了後に回収します＞

問題1(10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，胃痛の原因をベイスの定理により確率的に求めよ．

＜事前確率:原因が生じる確率＞

$P(A_1) = 60\ \%,\quad P(A_2) = 30\ \%,\quad P(A_3) = \ 10\%$

＜原因から結果が生じる確率＞

$P(B|A_1) = 30\ \%, P(B|A_2) = 50\ \%, P(B|A_3) = 20\ \%$

＜求める確率:結果から推定される原因の確率＞

$P(A_1|B),\quad P(A_2|B),\quad P(A_3|B)$

＜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＞

$P(A_1) = 60\ \%$
$P(A_2) = 30\ \%$
$P(A_3) = 10\%$

＜原因→結果の確率＞

$P(B|A_1) = 30\ \%$
$P(B|A_2) = 50\ \%$
$P(B|A_3) = 20\ \%$

ベイズの定理による推定結果

(分母)$= 0.35$
(分子)
$P(A_1|B) \rightarrow P(A_1)P(B|A_1) = 0.18$
$P(A_2|B) \rightarrow P(A_2)P(B|A_2) = 0.15$
$P(A_3|B) \rightarrow P(A_3)P(B|A_3) = 0.02$

＜結果→原因の確率＞

$P(A_1|B) = 0.514 = 51.4\ \%$
$P(A_2|B) = 0.429 = 42.9\ \%$
$P(A_3|B) = 0.057 = \ \ 5.7\ \%$

問題2(10点)

ある壺の中に赤玉が4個，青玉が3個，白玉が1個入っている．この壺から1個の玉を取り出すときのエントロピー(平均情報量)を求めよ．

＜解答例＞

赤玉を取り出す確率　$p_1 = 4/8$
青玉を取り出す確率　$p_2 = 3/8$
白玉を取り出す確率　$p_3 = 1/8$

エントロピー

$$H = \sum_{i=1}^{3} -p_i \log_2 p_i$$

$$= -\frac{1}{2}\log_2\left(\frac{1}{2}\right) - \frac{3}{8}\log_2\left(\frac{3}{8}\right) - \frac{1}{8}\log_2\left(\frac{1}{8}\right) = 1.41[bit]$$

問題3(10点)

二つのサイコロを振ったとき，その目の和が6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＞

① 目の和が6であり，目の組み合わせも分かっている（目の組み合わせ＝1通り）事象
確率：$p_1 = 1/36$
自己情報量：$I_1 = -\log_2 p_1 = 5.17[bit]$

② 目の和が6であり，目の組み合わせが不明である事象
目の和が6の組み合わせ＝(1,5), (2,4), (3,3), (4,2),(5,1)
確率：$p_1 = 5/36$
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4（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：14,000人，30才以上：6,000人

（地域別）
関東圏：6,000人，関東圏以外：14,000人

＜解答例＞

事象a_1： 30才未満 $p_{a1} = 14/20$
事象a_2： 30才以上 $p_{a2} = 6/20$
事象b_1： 関東圏 $p_{b1} = 6/20$
事象b_2： 関東圏以外 $p_{b2} = 14/20$

結合事象の確率

$$(a_1, b_1) \rightarrow p_{11} = \frac{14}{20} \times \frac{6}{20} = \frac{7}{10} \times \frac{3}{10}$$
$$(a_1, b_2) \rightarrow p_{12} = \frac{14}{20} \times \frac{14}{20} = \frac{7}{10} \times \frac{7}{10}$$
$$(a_2, b_1) \rightarrow p_{21} = \frac{6}{20} \times \frac{6}{20} = \frac{3}{10} \times \frac{3}{10}$$
$$(a_2, b_2) \rightarrow p_{22} = \frac{6}{20} \times \frac{14}{20} = \frac{3}{10} \times \frac{7}{10}$$

これらの確率を結合エントロピーの式に代入する．

$$H(A) = -\sum_{i,j=1}^{2} p_{ij} \log_2 p_{ij} = 1.76\ [bit]$$

問題5（10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	3	14	2	1
B君	6	5	5	4

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

＜解答例＞

A君： $p_S = 3/20,\ p_A = 14/20,\ p_B = 2/20,\ p_C = 1/20$
$$H(A) = \sum_{i=(S,A,B,C)} -p_i \log_2 p_i = 1.32\ [bit]$$
B君： $p_S = 6/20,\ p_A = 5/20,\ p_B = 5/20,\ p_C = 4/20$
$$H(B) = \sum_{i=(S,A,B,C)} -p_i \log_2 p_i = 1.99\ [bit]$$

$H(A) < H(B)$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）．