2020.8.18

情報統計力学 試験問題

必修問題

I. 以下を示せ。

1. 事象 A, B について，$P(A) > 0, P(B) > 0$ とする。B が起こる条件のもとで A が起こる確率を，事象 B が与えられたときに A が起こる条件付き確率といい，$P(A|B)$ と書く。これは次のようになる。

$$P(A|B) = \frac{P(A \cap B)}{P(B)}.$$

   以下のベイズの公式を示せ。

$$P(A|B)P(B) = P(B|A)P(A) \tag{1}$$

2. 次の等式を示せ。ただし，$P(B) > 0, P(B^c) > 0$ とする．ここで、B^c は，B の補集合、$B^c = \Omega - B$，Ω は起こりうる事象の集合、標本空間。

$$P(A) = P(A|B)P(B) + P(A|B^c)P(B^c) \tag{2}$$

II. いくつかのエントロピーの間の関係。
X を離散型確率変数として、有限 (n) 個の値をとるとする。$p(x)$ を確率関数とすると、X の情報エントロピー H_X は、

$$H_X = -\sum_{i=1}^{n} p(x_i) \log_2 p(x_i) \tag{3}$$

で与えられる。X, Y を確率変数とし、確率関数を $p(x, y)$ とする。確率変数 (X, Y) の結合エントロピーは、

$$H_{X,Y} = -\sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i, y_j) \log_2 p(x_i, y_j) \tag{4}$$

と定義される。また、$p(x) > 0$ として、x が与えられたときの y の条件付き確率分布を $p(y|x)$ とする。

$$p(y|x) = p(x, y)/p(x) \tag{5}$$

となる。 条件付きエントロピーは、以下で定義される。

$$H_{Y|X} = -\sum_{i=1}^{n} p(x_i) \sum_{j=1}^{m} p(y_j|x_i) \log_2 p(y_j|x_i). \tag{6}$$

相互情報量 $I_{X,Y}$ は次式で定義される。

$$I_{X,Y} = \sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i, y_j) \log_2 \frac{p(x_i, y_j)}{p(x_i)p(y_j)}. \tag{7}$$

次式を示せ。

$$
\begin{aligned}
&(a) \qquad H_{X,Y} = H_X + H_{Y|X}, && (8)\\
&(b) \qquad X, Y \text{ が独立なら、} H_{Y|X} = H_Y,\ H_{X,Y} = H_X + H_Y. && (9)\\
&(c) \qquad I_{X,Y} = H_X + H_Y - H_{X,Y},\ I_{X,Y} = H_Y - H_{Y|X} = H_X - H_{X|Y}. && (10)
\end{aligned}
$$

選択問題
次の２問のうちいずれか一つを回答せよ。

III-1. ベイズの公式の応用。乳がん検診に関する確率の問題を考える。以下のような確率を定義する。

P(陽性) : 検診で乳がんと判定される確率。P(乳がん) : 乳がんである確率。

P(陽性 | 乳がん) : 乳がんであるときに、検診で乳がんであると判定される確率。

P(乳がん | 陽性) : 検診で乳がんであると判定されたとき、実際に乳がんである確率。

いま、P(陽性 | 乳がん) $= 0.9$ とする。つまり、乳がんであるときに、検診で乳がんであると判定される確率が 9 割であるとする。また、実際の乳がんである人の割合は 0.8 パーセント、つまり、P(乳がん) $= 0.008$ とする。さらに，乳がんでないのに陽性となる確率が、P(陽性 | 乳がんでない) $= 0.07$ とする。
(a) ベイズの公式 (1) を用いて、検診で陽性と判定され、実際に乳がんである確率が約９パーセント、つまり、P(乳がん | 陽性) $\simeq 0.09$ となる事を示せ。
(b) １回目で陽性と判定されて、２回目の検診を受て、再び陽性と判定されたとき、実際に乳がんである確率が約 58 パーセント、つまり、P(乳がん | 陽性) $\simeq 0.58$ となる事を示せ。

III-2. 統計力学と情報理論におけるエントロピーの関係。
k_{B} をボルツマン定数とする。系が，体積 V の断熱容器に入れられており, 外部とのエネルギーのやりとりが行われていないとする。系のエネルギーは，E であり，その状態数を W とする。系の状態を，$i, (i = 1, \cdots, W)$ で表すとき，状態 i にある確率は，ミクロカノニカル分布

$$P_i = \frac{1}{W}$$

で与えられる。このとき,

$$S \quad = \quad k_{\mathrm{B}} T \ln W$$

である。一方，情報エントロピーは,

$$H_E = -\sum_{i=1}^{W} P_i \log_2 P_i$$

で与えられる。

$$H_E = \frac{1}{k_{\mathrm{B}} \ln 2} S$$

となることを示せ。