May, 2019

情報領域演習第二 第６回 K演習 (確率論) 宿題

宿題１－１：ベイズの定理

ある病気の感染率は 0.2% である。この病気の診断方法は、感染している患者を診断した時には 80% の確率で「感染あり」との結果を得る。また感染していない患者を診断した時には 10% の確率で「感染あり」という結果を得る。

1. 病気の感染の有無に関する周辺確率表を記せ。
2. 患者の感染の有無が判明しているときの、診断結果の「感染あり」「感染なし」の条件付き確率表を記せ。
3. 「診断結果が感染ありであった」という結果を知らされた。この患者が実際に感染している確率をベイズの定理を用いて求めよ。

略解：病気感染に関する周辺確率は次の表にまとめられる。

感染	あり	なし
周辺確率	0.002	0.998

(1)

診断の精度を条件付き確率で表すと、次の表になる。

		診断結果	
		あり	なし
真の状態	感染あり	0.80	0.20
	感染なし	0.10	0.90

(2)

診断結果が感染ありの場合の実際に感染している確率は

$$\frac{0.80 \times 0.002}{0.80 \times 0.002 + 0.10 \times 0.998} = \frac{8}{507} \tag{3}$$

である。

宿題1－2：ベイズの定理

ある病気の感染率は 0.1% である。この病気の診断方法 A は、感染している患者を診断した時の正答率が 70%、感染していない患者を診断したときの正答率も 70% である。なお正答率は、患者の真の状態を正しく当てる確率とする。

1. 病気の感染の有無に関する周辺確率表を記せ。
2. 患者の感染の有無が分かっているという条件の下で、方法 A の診断結果が「感染あり」、「感染なし」それぞれとなる条件付き確率表を記せ。
3. 「診断結果が感染ありであった」という結果を知らされた。この患者が実際に感染している確率をベイズの定理を用いて求めよ。
4. 別の診断方法 B は、感染している患者を診断した時の正答率が 80%、感染していない患者を診断したときの正答率は 60% である。診断方法 B で「診断結果が感染ありであった」という結果を知らされた患者が、実際に感染している確率をベイズの定理を用いて求めよ。

略解：病気感染に関する周辺確率は次の表にまとめられる。

感染	あり	なし
周辺確率	0.001	0.999

(4)

診断の精度を条件付き確率で表すと、次の表になる。

		診断結果	
		あり	なし
真の状態	感染あり	0.70	0.30
	感染なし	0.30	0.70

(5)

方法 A による診断結果が感染ありの場合の実際に感染している確率は

$$\frac{0.70 \times 0.001}{0.70 \times 0.001 + 0.30 \times 0.999} = \frac{7}{3004}$$

である。また、方法 B による診断結果が感染ありの場合の実際に感染している確率は

$$\frac{0.80 \times 0.001}{0.80 \times 0.001 + 0.40 \times 0.999} = \frac{2}{1001}$$

である。

宿題２：期待値の計算

1. 標本空間が $\{1, 2, 3\}$ の確率変数 X の確率関数が

$$p(1) = \frac{1}{2},\ p(2) = \frac{1}{6},\ p(3) = \frac{1}{3} \tag{6}$$

と与えられている。(a) まず累積分布関数 $F(x)$ を描け。 (b) 次に X の平均と分散を求めよ。

2. 練習問題２が一回に 1.5 の費用がかかる投資のリターン（回収額）を表しているとする。この投資の平均リターンを求めよ。またこの投資のリスク（標準偏差）を求めよ。

3. 標本空間が $\{x; 0 \le x \le 1\}$ の確率変数 X の確率密度関数が

$$f(x) = \left\{ \begin{array}{ll} 4x & 0 \le x < \frac{1}{2} \\ -4x + 4 & \frac{1}{2} \le x \le 1 \end{array} \right\} \tag{7}$$

と与えられている。(a) この確率分布の累積分布関数を描け。 (b) 確率変数 X の平均と分散を求めよ。

略解：累積分布関数のグラフは略。最初の問題の平均は

$$\mathrm{E}[X] = 1 \times \frac{1}{2} + 2 \times \frac{1}{6} + 3 \times \frac{1}{3} = \frac{3+2+6}{6} = \frac{11}{6} \tag{8}$$

分散は

$$\begin{aligned} \mathrm{Var}[X] &= \mathrm{E}\left[X^2\right] - \{\mathrm{E}[X]\}^2 \\ &= 1 \times \frac{1}{2} + 2^2 \times \frac{1}{6} + 3^2 \times \frac{1}{3} - \frac{11^2}{6^2} \\ &= \frac{1}{2} + \frac{4}{6} + \frac{9}{3} - \frac{121}{36} = \frac{18+24+108-121}{36} = \frac{29}{36} \end{aligned} \tag{9}$$

二つ目の問題の期待値は

$$\mathrm{E}[X - 1.5] = \mathrm{E}[X] - 1.5 = 0.5 - 1.5 = -1 \tag{10}$$

三つめの問題の期待値は

$$\begin{aligned} \mathrm{E}[X] &= \int_0^1 x f(x)\, dx \\ &= \int_0^{\frac{1}{2}} x(4x)\, dx + \int_{\frac{1}{2}}^1 x(-4x+4)\, dx \\ &= \frac{4}{3}\left[x^3\right]_0^{\frac{1}{2}} + \left[-\frac{4}{3}x^3 + 2x^2\right]_{\frac{1}{2}}^1 \\ &= \frac{1}{6} - \frac{7}{6} + \frac{3}{2} = \frac{1}{2} \end{aligned}$$

分散は

$$
\begin{aligned}
\mathrm{E}\left[X-\mu\right] &= \int_0^1 \left(x-\frac{1}{2}\right)^2 f\left(x\right) dx \\
&= \int_0^{\frac{1}{2}} \left(x-\frac{1}{2}\right)^2 \left(4x\right) dx + \int_{\frac{1}{2}}^1 \left(x-\frac{1}{2}\right)^2 \left(-4x+4\right) dx \\
&= 2\int_0^{\frac{1}{2}} \left(x-\frac{1}{2}\right)^2 \left(4x\right) dx \\
&= 2\int_0^{\frac{1}{2}} \left(4x^3-4x^2+x\right) dx \\
&= 2\left[x^4-\frac{4x^3}{3}+\frac{x^2}{2}\right]_0^{\frac{1}{2}} \\
&= 2\left\{\frac{1}{16}-\frac{1}{6}+\frac{1}{8}\right\} = \frac{3-8+6}{24} = \frac{1}{24}
\end{aligned}
$$

と求まる。

宿題３－１：確率不等式

1. 確率表

X	1	2	$\cdots$	N	総和
確率	p_1	p_2	$\cdots$	p_N	1

と期待値 μ と分散 σ^2 が与えられている。教科書の証明を参考に

$$
\begin{aligned}
\sigma^2 &= \mathrm{E}\left[\left(X-\mu\right)^2\right] \\
&= \sum_{k=1}^{N} \left(k-\mu\right)^2 p_k \\
&= \cdots
\end{aligned} \tag{11}
$$

から始め、不等式

$$
\sigma^2 \geq a^2 P\left(\left|X-\mu\right| \geq a\right) \tag{12}
$$

を示して、チェビシェフの不等式を証明せよ。

略解：これは離散分布の場合のチェビシェフの不等式で、連続分布の場合と同様に次のような不等式から示

せる。

$$
\begin{aligned}
\sigma^2 &= \mathrm{E}\left[(X-\mu)^2\right] \\
&= \sum_{k=1}^{N} (k-\mu)^2 p_k \\
&\geq \sum_{k:(\mu-k)\leq -a} (k-\mu)^2 p_k + \sum_{k:(k-\mu)\geq a} (k-\mu)^2 p_k \\
&\geq \sum_{k:(\mu-k)\leq -a} a^2 p_k + \sum_{k:(k-\mu)\geq a} a^2 p_k \\
&= a^2 \left\{ \sum_{k:(\mu-k)\leq -a} p_k + \sum_{k:(k-\mu)\geq a} p_k \right\} \\
&= a^2 P\left((X-\mu)^2 \geq a^2\right)
\end{aligned}
$$

宿題３－２：確率不等式

1. マルコフの不等式を用いて、平均年収が 400 万円の世代の中で、年収が 1000 万円以上の人の割合は最大でどれぐらいとなるかを求めよ。
2. マルコフの不等式を用いて、1 企業あたりの平均従業員数が 20 人の国では、1 万人以上の従業員を抱える企業の割合が最大でどれぐらいとなるかを求めよ。
3. チェビシェフの不等式を用いて、年収の平均が 400 万円、標準偏差が 200 万円の世代の中で、年収が 1000 万円以上の人の割合は最大でどれぐらいとなるかを求めよ。
4. チェビシェフの不等式を用いて、1 企業あたりの従業員数の平均が 400 人、標準偏差が 200 人の国で、1 万人以上の従業員を抱える企業の割合が最大でどれぐらいとなるかを求めよ。

略解：

$$P(X \geq 1000) \leq \frac{400}{1000} = \frac{2}{5} \tag{13}$$

$$P(X \geq 10000) \leq \frac{20}{10000} = \frac{1}{500} \tag{14}$$

$$P(X-400 \geq 1000-400) \leq P(|X-400| \geq 600) \leq \frac{200^2}{600^2} = \frac{1}{9} \tag{15}$$

$$P(X-400 \geq 10000-400) \leq P(|X-400| \geq 9600) \leq \frac{200^2}{9600^2} = \frac{1}{2304} \tag{16}$$