確率・統計

1　確率空間

定義 1.1. 以下の条件を満たす、3 つの組 $(\Omega, \mathcal{A}, P)$ を**確率空間**という。

(1) Ω は空でない集合である。Ω を**標本空間**という。Ω は起こりうる全ての結果の集まりであると解釈される。

(2) $\mathcal{A}$ は Ω の部分集合からなる空でない集合であり、以下の条件を満たす。
  (i) $\emptyset \in \mathcal{A}$
  (ii) $X \in \mathcal{A}$ ならば、$X^c \in \mathcal{A}$（ただし、$X^c = A \setminus X$ とする）
  (iii) $\mathcal{A}$ の元 $A_1, A_2, \ldots$ に対して、$\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$

  $\mathcal{A}$ を**事象空間**という。$\mathcal{A}$ は起こりうる結果の中で確率を計算することが出来る結果の集まりであると解釈される。

(3) P は $\mathcal{A}$ から $[0,1]$ への関数であり、以下の条件を満たす。
  (i) $P(\Omega) = 1$
  (ii) $\mathcal{A}$ の元 $A_1, A_2, \ldots$ に対して、全ての $i \neq j$ に対して、$A_i \cap A_j = \emptyset$ ならば、$P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$

  P を**確率**という。P は事象が起こる確率を表していると解釈される。

例 1.1. 2 つのサイコロ X_1, X_2 を投げた結果の確率について考える。X_1 の目が x_1 であり、X_2 の目が x_2 であるという結果を (x_1, x_2) で表すことにする。このとき、起こりうる全ての結果の集まりは $\{(x_1, x_2) \mid 1 \leq x_1, x_2 \leq 6\}$ である。そこで、$\Omega = \{(x_1, x_2) \mid 1 \leq x_1, x_2 \leq 6\}$ とおく。Ω の任意の部分集合 Φ に対して、Φ が起こる確率を $P(\Phi)$ で表すことにする。このとき、$\mathcal{A} = \Omega$ とおけば、$(\Omega, \mathcal{A}, P)$ は確率空間である。例えば、$P(\{(1,1)\}) = 1/36$ である。

定義 1.2. $(\Omega, \mathcal{A}, P)$ を確率空間とする。このとき、$X, Y \in \mathcal{A}$ に対して、$P(Y) \neq 0$ であるとき、

$$P(X|Y) := \frac{P(X \cap Y)}{P(Y)}$$

と定義する。$P(X|Y)$ を**条件付き確率**という。

例 1.2. 子供が二人いる家族に会ったところ、男子がいることが分かった。このとき、男子がもう一人いる確率を求める。ただし、男子と女子が生まれる確率はそれぞれ 1/2 とする。男子を b で表し、女子を g で表すことにする。子供が二人いる家族に対して、第一子の性別が x_1 であり、第二子の性別が x_2 であることを (x_1, x_2) で表すことにする。このとき、標本空間は $\Omega = \{(b,b), (b,g), (g,b), (g,g)\}$ である。$\mathcal{A} = \Omega$ と

し、$P : \mathcal{A} \to [0,1]$ を $P(\{(x_1, x_2)\}) = 1/4$ で定義する。$(\Omega, \mathcal{A}, P)$ は確率空間である。男子がいる事象は $Y = \{(b,b),(b,g),(g,b)\}$ であり、男子が二人いる事象は $X = \{(b,b)\}$ なので、男子がもう一人いる確率は $P(X|Y) = P(X \cap Y)/P(Y) = (1/4)/(3/4) = 1/3$ である。

定理 1.1. （ベイズの定理） $(\Omega, \mathcal{A}, P)$ を確率空間とする。このとき、$X, Y \in \mathcal{A}$ に対して、$P(X), P(Y) \neq 0$ であるとき、

$$P(X|Y) = \frac{P(Y|X)P(X)}{P(Y)}$$

が成り立つ。

例 1.3. 罹患率 0.01 %の病気の罹患状況について考える。この病気の検査方法では、実際に罹患している人間が陽性と判断される確率が 98 %であり、罹患してない人が陰性と判断される確率は 80 %となる。陽性だと診断される確率は 20 %である。このとき、陽性と診断された場合に本当に罹患している確率を求める。X を「病気である」という事象とし、Y を「病気であると判断される」という事象とする。仮定により、

$$P(X) = 0.0001, \quad P(Y|X) = 0.98, \quad P(Y^c|X^c) = 0.8, \quad P(Y) = 0.2$$

である。よって、ベイズの定理により、

$$P(X|Y) = \frac{P(Y|X)P(X)}{P(Y)} = \frac{0.98 \times 0.0001}{0.2} = 4.9 \times 10^{-4}$$

である。よって、陽性と診断された場合に本当に罹患している確率は 0.049 %である。

2　確率変数と確率分布

定義 2.1. $(\Omega, \mathcal{A}, P)$ を確率空間とする。このとき、Ω から実数全体の集合 $\mathbb{R}$ への関数 $X : \Omega \to \mathbb{R}$ を**確率変数**という。任意の $x \in \mathbb{R}$ に対して、

$$F_X(x) := P(\{\omega \in \Omega \mid X(\omega) \leq x\})$$

と定義する。F_X を X の**分布関数**という。

定義 2.2. $(\Omega, \mathcal{A}, P)$ を確率空間とし、$X : \Omega \to \mathbb{R}$ を確率変数とする。$X(\Omega)$ が有限集合であるとき、X を**離散型確率変数**といい、任意の $x \in X(\Omega)$ に対して、

$$P_X(x) := P(\{\omega \in \Omega \mid X(\omega) = x\}) \tag{1}$$

と定義する。P_X を X の**確率関数**という。(1) の右辺は $P(x)$ と略記される。

例 2.1. 例 1.1 の確率空間 $(\Omega, \mathcal{A}, P)$ を考える。$X : \Omega \to \mathbb{R}$ を $X((x_1, x_2)) = x_1 + x_2$ で定義する。X は確率変数である。よって、

$$P(3) = P(\{\omega \in \Omega \mid X(\omega) = 3\}) = P(\{(1,2),(2,1)\}) = 2/36 = 1/18$$

である。

定義 2.3. $(\Omega, \mathcal{A}, P)$ を確率空間とし、$X : \Omega \to \mathbb{R}$ を確率変数とする。ある関数 f_X が存在して、

$$F_X(x) = \int_{-\infty}^{x} f_X(x)dx$$

が成り立つとき、X を**連続型確率変数**といい、f_X を X の**確率密度関数**という。

例 2.2. 0 から 1 の目盛りが振られたルーレットを二つ回す。それぞれの針が指す目盛り α, β についての確率は以下の確率空間 $(\Omega, \mathcal{A}, P)$ で記述される。

(1) 標本空間 : $\Omega = \{(\alpha, \beta) \mid 0 \leq \alpha, \beta < 1\}$
(2) 事象空間 : $\mathcal{A} = \{X \subset \Omega \mid X$ は面積が測れる $\}$
(3) 確率 : $P(X) = \frac{X \text{ の面積}}{\Omega \text{ の面積}}$

確率変数 $X : \Omega \to \mathbb{R}$ を $X((\alpha, \beta)) = \max\{\alpha, \beta\}$ で定義する。X の確率密度関数を求める。任意の $x \in \mathbb{R}$ に対して、

$$F_X(x) = \{(\alpha, \beta) \in \Omega \mid X((\alpha, \beta)) \leq x\} = \{(\alpha, \beta) \in \Omega \mid \alpha, \beta \leq x\}$$

であるから、

$$F_X(x) = \begin{cases} 0 & x < 0 \\ x^2 & 0 \leq x < 1 \\ 1 & x \geq 1 \end{cases}$$

である。F_X は微分可能であり、$f_X(x) = dF_X(x)/dx$ とおけば、

$$f_X(x) = \begin{cases} 0 & x < 0 \\ 2x & 0 \leq x < 1 \\ 1 & x \geq 1 \end{cases}$$

である。よって、

$$F_X(x) = \int_{-\infty}^{x} f_X(x)dx$$

であるから、X は連続型確率変数であり、f_X は X の確率密度関数である。

以下では、代表的な確率分布について紹介する。

定義 2.4. (二項分布) n を自然数とし、$0 < p < 1$ とする。二項分布は以下の確率空間と確率変数からなる。

(1) 標本空間 : $\Omega = \{(\omega_1, \ldots, \omega_n) \mid \omega_1, \ldots, \omega_n \in \{0, 1\}\}$
(2) 事象空間 : $\mathcal{A} = \Omega$ の部分集合全体の集合
(3) 確率 : $P(\{(\omega_1, \ldots, \omega_n)\}) = p^{\sum_{i=1}^{n} \omega_i}(1-p)^{n-\sum_{i=1}^{n} \omega_i}$
(4) 確率関数 : $Y_n : \Omega \to \mathbb{R}$ を $Y_n((\omega_1, \ldots, \omega_n)) = \sum_{i=1}^{n} \omega_i$ で定義する。
(5) Y_n の確率関数 P_{Y_n} は $P_{Y_n}(k) = {}_nC_k p^k (1-p)^{n-k}$ で与えられる。

P_{Y_n} を**二項分布**といい、Y_n は二項分布に従うという。$n = 1$ であるとき、P_{Y_n} を**ベルヌーイ分布**という。P_{Y_n} を $B(n, p)$ で表す。

定義 2.5. （正規分布）$\mu \in \mathbb{R}, \sigma > 0$ とし、$(\Omega, \mathcal{A}, P)$ を確率空間、$X : \Omega \to \mathbb{R}$ を確率変数とする。X が連続型確率変数であり、X の確率密度関数が

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

で与えられるとき、f_X を**正規分布**といい、X は正規分布に従うという。f_X を $N(\mu, \sigma)$ で表す。

3 平均・分散・共分散

定義 3.1. （平均）$(\Omega, \mathcal{A}, P)$ を確率空間、$X : \Omega \to \mathbb{R}$ を確率変数とする。このとき、X の期待値 $E[X]$ を

$$E[X] := \begin{cases} \sum_{x \in X(\Omega)} P_X(x)x & X \text{ が離散型確率変数であるとき} \\ \int_{-\infty}^{\infty} f_X(x)xdx & X \text{ が連続型確率変数であるとき} \end{cases}$$

で定義する。

定義 3.2. （分散）$(\Omega, \mathcal{A}, P)$ を確率空間、$X : \Omega \to \mathbb{R}$ を確率変数とする。$\mu = E[X]$ とおく。$Y : \Omega \to \mathbb{R}$ を $Y(\omega) = (X(\omega) - \mu)^2$ で定義する。Y は確率関数である。Y を $(X - \mu)^2$ で表す。このとき、X の分散 $V[X]$ は

$$V[X] := E[(X - \mu)^2]$$

で定義される。$V[X] = E[X^2] - E[X]^2$ が成り立つので、計算の際は $E[X^2], E[X]$ を計算したほうが簡単である。

定義 3.3. （共分散）$(\Omega, \mathcal{A}, P)$ を確率空間、$X, Y : \Omega \to \mathbb{R}$ を確率変数とする。$\mu_X = E[X], \mu_Y = E[Y]$ とおく。このとき、$Z : \Omega \to \mathbb{R}$ を $Z(\omega) = (X(\omega) - \mu_X)(Y(\omega) - \mu_Y)$ で定義すれば、Z は確率変数である。Z を $(X - \mu_X)(Y - \mu_Y)$ で表す。このとき、X, Y の共分散 $Cov(X, Y)$ は

$$Cov(X, Y) = E[(X - \mu_X)(Y - \mu_Y)]$$

で定義される。

例 3.1. $(\Omega, \mathcal{A}, P)$ を確率空間、$X : \Omega \to \mathbb{R}$ を確率変数とする。X が二項分布 $B(n, p)$ に従うとする。このとき、$E[X] = np, V[X] = np(1 - p)$ である。

例 3.2. $(\Omega, \mathcal{A}, P)$ を確率空間、$X : \Omega \to \mathbb{R}$ を確率変数とする。X が正規分布 $N(\mu, \sigma)$ に従うとする。このとき、$E[X] = \mu, V[X] = \sigma^2$ である。

参考文献

[1] Andriy Burkov. (2019). The hundred-page machine learning book.

[2] Marc Peter Deisenroth., A. Aldo Faisal., Cheng Soon Ong. (2020). Mathematics for machine learning. Cambridge University Press.

[3] 小縣信也., 斎藤翔汰., 溝口聡., 若杉一幸. (2021). ディープラーニング E 資格エンジニア問題集 インプレス.