

確率論まとめ

このまとめは個人的にまとめたものであり厳密性を考えていない他、間違いがあるかもしれない。

1 確率空間

1.1 標本空間と事象

1.1.1 導入

はじめに言葉を定義する。ある試行を行ったときの事象を**標本点**と呼ぶ。標本点全体を**標本空間**と呼ぶ。ここで標本空間を Ω で示す。試行を行った時の発生する事象を A で表す。 A は標本空間の部分集合であり、 $A = \Omega$ の時 A を**全事象**、 A が標本点を持たないとき A は**空事象**と呼ばれる。なお空事象は \emptyset で表される。事象 A が起きない事象を**補事象**と呼び、 A^C で表される。

さらに2つ以上の事象を考えてみる。2つの事象 A と B において少なくとも一方が起きるという事象を A と B の**和事象**と呼ばれる。これは $A \cup B$ で表される。また A 、 B の両方が起こる事象について**積事象**と呼び、 $A \cap B$ で表される。 A 、 B が同時に起こりえないことを**排反**と呼ぶ。

実例を示してみよう。サイコロを1回振ったときに結果は1,2,3,4,5,6のいずれかである。このときの結果1つ1つが標本点である。標本空間は

$$\Omega = \{1, 2, 3, 4, 5, 6\} \quad (1,1)$$

サイコロを振った時に6以下の自然数が出るという事象は Ω のため全事象、7以上の自然数が出るという事象は空事象となる。サイコロを振った時に奇数が出る事象を A とすると、補事象 A^C は偶数が出る事象となる。サイコロを振った時に3以下が出る事象を B とすると、 A と B の和事象は $\{1, 2, 3, 5\}$ 、積事象は $\{1, 3\}$ となる。サイコロを振った時に5が出る事象を C とすると、 A と C は排反である。

1.1.2 ド・モルガンの法則

和事象と積事象の補集合に関する法則である。具体的には次式で表される。

$$(A \cup B)^C = A^C \cap B^C \quad (1,2)$$

$$(A \cap B)^C = A^C \cup B^C \quad (1,3)$$

ここで導入のサイコロの例をもとに実例を挙げてみよう。 A はサイコロを振った時に奇数が出る事象、 B はサイコロを振った時に3以下が出る事象である。

$$(A \cup B)^C = \{1, 2, 3, 5\}^C = \{4, 6\} \quad (1,4)$$

$$A^C \cap B^C = \{1, 3, 5\}^C \cap \{1, 2, 3\}^C = \{2, 4, 6\} \cap \{4, 5, 6\} = \{4, 6\} \quad (1,5)$$

$$(A \cap B)^C = \{1, 3\}^C = \{2, 4, 5, 6\} \quad (1,6)$$

$$A^C \cup B^C = \{1, 3, 5\}^C \cup \{1, 2, 3\}^C = \{2, 4, 6\} \cup \{4, 5, 6\} = \{2, 4, 5, 6\} \quad (1,7)$$

1.1.3 結合法則及び分配法則

確率の事象にも結合法則及び分配法則が適応可能。具体的には次式で表される。

$$(A \cup B) \cup C = A \cup B \cup C \quad (1,8)$$

$$(A \cap B) \cap C = A \cap B \cap C \quad (1,9)$$

$$(A \cup B) \cap C = (A \cap C) \cup (B \cap C) \quad (1,10)$$

$$(A \cap B) \cup C = (A \cup C) \cap (B \cup C) \quad (1,11)$$

1.2 確率の定義

確率とは**事象の起きやすさを表す量**である。事象 A が起きる確率を $P(A)$ と表す。具体的には以下に示す**確率の公理**を満たす写像 $P(A)$ を確立と呼ぶ。

- 事象 A に対して $P(A)$ は実数である。そして $0 \leq P(A) \leq 1$
- $P(\Omega) = 1$
- 互いに排反な事象 A_1, \dots, A_n に対して、 $P(A_1 \cup \dots \cup A_n) = P(A_1) + \dots + P(A_n)$

1.3 確率の性質

確率の公理より簡単に導くことができる性質を以下に示す。

$$P(\emptyset) = 0 \quad (1,12)$$

$$P(A^C) = 1 - P(A) \quad (1,13)$$

$$A \subset B \Rightarrow P(A) \leq P(B) \quad (1,14)$$

1.3.1 加法定理

さらに確率の公理より次の等式が成立する。

$$P(A \cup B) = P(A \cap B^C) + P(A \cap B) + P(A^C \cap B) \quad (1,15)$$

$$P(A) = P(A \cap B^C) + P(A \cap B) \quad (1,16)$$

$$P(A) = P(A^C \cap B) + P(A \cap B) \quad (1,17)$$

これら 3 式を足すことで**加法定理**が得られる。

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \quad (1,18)$$

1.4 条件付確率

2つの事象 A と B に対して、事象 A が起きたという条件の下で事象 B が起きるといった確率を条件付確率と呼び $P(B|A)$ と表す。条件付確率は次式で求めることができる。

$$P(B|A) = \frac{P(A \cap B)}{P(A)} \quad (1,19)$$

また次式を変形すると

$$P(A \cap B) = P(A)P(B|A) \quad (1,20)$$

なおこの式を乗法定理と呼ぶ。

具体例を考えてみる。袋 X が1つと袋 Y が2つ存在し、袋 X には赤玉3個白玉1個、袋 Y には赤玉2個白玉1個あるとする。袋を1つ選びその中から1つ球をとるということを考える。袋 X を引いたときに白玉を引く確率を考える。袋 X を引く事象を A 、白玉を引く事象を B とする。 $P(A) = \frac{1}{3}$ であり、 $P(A \cap B) = \frac{1}{12}$ よって条件付確率 $P(B|A) = \frac{1}{4}$ である。

また袋 Y を引く事象を C 、赤玉を引く事象を D とすると、 $P(C) = \frac{2}{3}$ 、 $P(D|C) = \frac{1}{2}$ のため、乗法定理より $P(C \cap D) = \frac{1}{3}$ となる。もしこの具体例がわからない場合は図示して考えてみるとよい。

1.5 独立性

2つの事象 A と B に対して、 $P(B|A) = P(B)$ となるとき、事象 A と事象 B は独立という。独立である条件は A が B に依存しないことである。このとき乗法定理は次のように変形できる。

$$P(A \cap B) = P(A)P(B) \quad (1,21)$$

実例としてサイコロを2回振った時の出目の組み合わせを考えてみる。1回目に振り奇数が出る事象を A 、2回目に振り奇数が出る事象を B とすると、 A と B はお互い依存していない。よって $P(A \cap B) = P(A)P(B) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$ となる。

1.6 ベイズの定理

事象 A が事象 B_1, \dots, B_k の事象が発生したときのみに起こりうるということを考える。そして B_1, \dots, B_k 以外の事象が発生したときに A は起こらないこととする。つまり事象 B_i が発生し、事象 A が起こるような条件を考える。このとき事象 A が起きたときにそれが B_i という事象であったという条件付確率を求める。この時求める確率は $P(B_i|A)$ となる。(1,19) 式より、

$$P(B_i|A) = \frac{P(B_i \cap A)}{P(A)} \quad (1,22)$$

ここで A というのは条件 B_1, \dots, B_k が起こった上で発生するため、

$$P(A) = \sum_{j=1}^k P(A \cap B_j) = \sum_{j=1}^k P(B_j)P(A|B_j) \quad (1,23)$$

よって次式が成立する。

$$P(B_i|A) = \frac{P(B_i \cap A)}{\sum_{j=1}^k P(B_j)P(A|B_j)} \quad (1,23)$$

なおこの式を**ベイズの定理**と呼ぶ。

やはり式だけ見てもわかりずらいため実例を出してみる。袋 B_1 に赤玉 3 つ白玉 1 つ、袋 B_2 に赤玉 2 つ白玉 2 つ、袋 B_3 に赤玉 3 つ白玉 2 つあるとする。この時袋 B_i を選ぶ事象を B_i とし、白玉を選ぶ事象を A とする。ここで白玉を引いたときにそれが袋 B_2 から引いたものである確率をベイズの定理を用いて計算する。まず B_i だが 3 種類の袋を偏りなく選ぶため確率は $\frac{1}{3}$ 。ここで $P(A|B_1) = \frac{1}{3}, P(A|B_2) = \frac{1}{2}, P(A|B_3) = \frac{2}{5}$ よってベイズの定理より、

$$P(B_2|A) = \frac{\frac{1}{3} \cdot \frac{1}{2}}{\frac{1}{3} \cdot \frac{1}{4} + \frac{1}{3} \cdot \frac{1}{2} + \frac{1}{3} \cdot \frac{2}{5}} = \frac{10}{23} \quad (1,24)$$

2 確率変数と確率分布

2.1 確率変数と確率分布

2.1.1 確率変数

ある変数の値をとる確率が存在する変数を**確率変数**と呼ぶ。例えばサイコロならば 1 から 6 までの整数値をとるため 1,2,3,4,5,6 が確率変数となり、 $P(X) = \frac{1}{6}$ となる。

2.1.2 離散型と連続型

確率変数 X が可算個の離散値の場合、確率変数 X は**離散型**となる。また確率変数 X が連続値をとる場合は**連続型**となる。以降**離散型か連続型によって使用できる式が異なる**ためしっかりと判別しよう。

2.1.3 確率関数

まず、**離散型**確率変数の確率を表現する関数として次の式を考える。なお X は確率変数を、 x は取りうる値を示している。今後の説明のため $x = x_1, x_2 \cdots x_i \cdots$ とする。

$$f(x) = P(X = x) \quad (2,1)$$

この $f(x)$ を**確率関数**と呼ぶ。ここで確率関数は次の 2 つの性質を当然満たす。

$$f(x_i) > 0 \quad (2,2)$$

$$\sum_{i=1}^{\infty} f(x_i) = 1 \quad (2,3)$$

2.1.4 確率密度関数

次に**連続型**の場合の確率関数のようなものを考える。連続型確率変数 X が区間 $(a, b]$ にあるとする。この時の確率を

$$P(a < X \leq b) = \int_a^b f(t)dt \quad (2,4)$$

で考える。この確率関数のようなものは次の 2 つの性質を当然満たす。

$$f(x) \geq 0 \quad (2,5)$$

$$\int_{-\infty}^{\infty} f(x)dx = 1 \quad (2,6)$$

この関数 $f(x)$ を**確率密度関数**と呼ぶ。

2.1.5 分布関数

確率関数および確率密度関数ではある 1 点やある区間での確率であった。しかし場合によっては確率が積もっていく様子を捉えたいこともある。よって確率が積もっていく様子を捉える量として**分布関数**というものを用意する。分布関数は**大文字の F で表す**。分布関数は次で示す。

$$F(X) = P(X \leq x) \quad (2,7)$$

さらに分布関数は次の性質を持っている

- $F(a) \leq F(b) \quad (a < b)$
- $0 \leq F(x) \leq 1$
- $F(x)$ は右連続

さてここまで書いたことを考えると当然だが分布関数が決まると確率 (密度) 関数は同時に決定する。同様に確率 (密度) 関数が決定すると分布関数が決定する。ここで**離散型**の確率関数と分布関数の関係をまとめる。

$$F(x) = P(X \leq x) = \sum_{x_i \leq x} f(x_i) \quad (2,8)$$

次に**連続型**の確率密度関数と分布関数の関係をまとめる。

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(t)dt \quad (2,9)$$

なお密度関数や分布関数による確率的挙動を**確率分布**と呼んでいるが、確率分布については 3 節にて触れる。

2.2 期待値と平均と分散

2.2.1 平均そして期待値

離散型の平均 μ について以下のように定義する。

$$\mu = \sum_{i=1,2,\dots} x_i f(x_i) \quad (2,10)$$

実例を挙げてみよう。サイコロの出目の平均を計算すると 3.5 となる。

連続型の平均 μ について以下のように定義する。なお確率変数 X が取りうる領域を \mathcal{X} とする。

$$\mu = \int_{\mathcal{X}} x f(x) dx \quad (2,11)$$

そしてこれら平均を総じて確率変数 X の**期待値**という。期待値は $E[x]$ で表される。なお $E[X]$ は線形性を持っており場合によっては計算が楽である。

2.2.2 分散

確率変数のばらつきを捉えるものに**分散**がある。これは確率変数 X の平均 μ からの離れ具合を 2 乗に基づいて平均的に測るものである。分散 $\sigma^2 = V(x)$ は次式で定義する。

$$\sigma^2 = V[x] = E[(X - \mu)^2] \quad (2,12)$$

これは**離散型**, **連続型**どちらでも使うことができる。ただし期待値はそれぞれ対応する式で用いること。

2.2.3 標準偏差

分散はばらつきの尺度として優秀な一方で 2 乗しているため元の確率変数と単位が異なる。よって単位を合わせるために分散の平方根を標準偏差とする。

2.2.4 k 次モーメント

一般的に $E[X^k]$ を k 次モーメントという。そして $E[(X - \mu)^k]$ を k 次の中心モーメントという。そして今まで出たものをまとめると、平均は 1 次のモーメント、分散は 2 次の中心モーメントである。

ここで期待値が線形であり、 $E[X] = \mu$ を利用することで平均と分散に関して次の関係性を導くことができる。

$$\begin{aligned}\sigma^2 &= E[(X - \mu)^2] \\ &= E[X^2 - 2\mu X + \mu^2] \\ &= E[X^2] - 2\mu E[X] + E[\mu^2] \\ &= E[X^2] - \mu^2\end{aligned}$$

2.2.5 標準化

世の中の無数の確率変数が同じ平均と標準偏差を持つように確率変数を変換することを標準化と呼ぶ。具体的に以下のように変換する。なお Z は確率変数である。

$$Z = \frac{X - \mu}{\sigma} \quad (2,13)$$

このとき平均 $E[Z]$ は 0、分散 $V[Z]$ は 1 になる。

逆に標準化されている確率変数 Z に対して、 $X = \mu + \sigma Z$ と変換すると、 $E[X] = \mu$ と $V[x] = \sigma^2$ が簡単に得られる。

2.3 多次元確率変数と同時確率分布と周辺確率分布

2.3.1 多次元確率変数

まず 1 次元の確率変数 X_1, \dots, X_k をまとめた確率変数 $X = (X_1, \dots, X_k)$ を多次元確率変数という。これはベクトルである。

2.3.2 同時確率分布

例えば袋の中に赤、青、黄の 3 色のボールが入っていることを考える。そしてそのボールには 1 から 4 までの数字が書かれている。この時それぞれのボールの確率を求めるにはボールの色という確率変数とボールの数字という確率変数の 2 種類を考慮しなければならない。

このようにいくつかの確率変数を同時に扱うときの確率分布を同時確率分布という。

離散型の場合、次式で示される。

$$f(x_i, y_j, \dots) = P(X = x_i, Y = y_j, \dots) \quad (2,14)$$

連続型の場合、次式で示される。

$$f(a < X \leq b, c < Y \leq d, \dots) = \int_a^b \int_c^d \dots f(x, y, \dots) dx dy \dots \quad (2,15)$$

2.3.3 周辺確率分布

周辺確率とは、ただ1つだけの事象が起こる確率である。つまり1つの変数を固定し、もう1つの変数のパターンをすべて足せばいい。**離散型**の場合、次式を例とする。

$$P(X_1 = x_1) = \sum_{x_2} \sum_{X_3} \cdots P(X_1 = x_1, X_2 = x_2, X_3 = x_3 \cdots) \quad (2,16)$$

要は1つの確率変数を固定して後ひたすら足しているだけなのだ。**連続型**の場合、次式を例とする。

$$P(X_1 \leq x_1, -\infty < X_2 < \infty \cdots) = \int_{-\infty}^x \int_{-\infty}^{\infty} \cdots f(x_1, c_2, \cdots) dx_1 dx_2 \cdots \quad (2,17)$$

2.4 共分散

2つの確率変数 X と Y の関係を表す量として**共分散**を次式で定義する。

$$\sigma_{xy} = Cov[X, Y] = E[(X - \mu_x)(Y - \mu_y)] \quad (2,18)$$

また共分散の代わりに**相関係数**に着目することもある。相関係数は次式で定義される。

$$\rho_{xy} = Corr[X, Y] = \frac{Cov[X, Y]}{\sqrt{V[X]V[Y]}} \quad (2,19)$$

2.5 確率変数の和の平均と分散

確率変数の和 $X + Y$ における平均と分散は次式で表すことができる。

- 和: $E[X + Y] = E[X] + E[Y]$
- 分散: $V[X + Y] = V[X] + V[Y] + 2Cov[X, Y]$

なお分散については以下に導出を記載しておく。

$$\begin{aligned} V[X + Y] &= E[\{(X + Y) - (\mu_x + \mu_y)\}^2] \\ &= E[\{(X - \mu_x) + (Y - \mu_y)\}^2] \\ &= E[(X - \mu_x)^2] + E[(Y - \mu_y)^2] + 2E[(X - \mu_x)(Y - \mu_y)] \\ &= V[X] + V[Y] + 2Cov[X, Y] \end{aligned}$$

2.6 確率不等式

2.6.1 チェビシエフの不等式

チェビシエフの不等式を紹介する。確率変数 X における平均と分散を用いる。このとき任意の $\varepsilon > 0$ に対して、次が成り立つ。

$$P[|X - \mu| \geq \varepsilon] \leq \frac{\sigma^2}{\varepsilon^2} \quad (2,20)$$

この不等式は、特定の値以下もしくは以上の累積分布関数の最大値を求めることができる。

2.6.2 コーシー・シュバルツの不等式

次に**コーシー・シュバルツの不等式**を紹介する。コーシー・シュバルツの不等式は確率変数 X と Y の分散と共分散との間に次の不等式が成立するというものである。

$$\text{Cov}[X, Y]^2 \leq V[X]V[Y] \quad (2,21)$$

2.7 条件付密度関数

確率変数 $X = x$ 、 $Y = y$ の時の**条件付密度関数** $f_{Y|X}(y|x)$ を次式で定義する。

$$f_{Y|X}(y|x) = \frac{f(x, y)}{f_X(x)} \quad (2,22)$$

3 確率分布

3.1 離散型確率分布

3.1.1 一様分布

確率変数の値に関わらず一定の確率をとるような確率モデルの場合、**離散型**確率変数 X は**一様分布**に従うといわれる。

$$f(x_i) = P(X = x_i) = \frac{1}{n} \quad (i = 1, \dots, n) \quad (3,1)$$

実例を見てみよう。サイコロを振ったときに出る目の確率を考える。確率変数 X は 1 から 6 までの値をとる。そして確率変数にかかわらず確率は常に $\frac{1}{6}$ となる。

3.1.2 ベルヌーイ分布

まず確率変数が 2 つの値 0 と 1 しかとらないという状況を考える。ここで $P(1) = \theta$ と定義すると、 $P(0) = 1 - \theta$ となる。もちろん $0 < \theta < 1$ となる。このときこの確率は次のように表現できる。

$$P(x) = \theta^x (1 - \theta)^{1-x} \quad (3,2)$$

この密度関数をもつ**離散型**確率変数 x は、**ベルヌーイ分布**に従うという。

実例を考えてみよう。10%の確率で当選するくじがあったとする。くじに当選する事象を $X = 1$ 、外れる事象を $X = 0$ とすると、 $\theta = 0.1$ のため

$$P(x) = 0.1^x (0.9)^{1-x}$$

となる。

3.1.3 2 項分布

ここでベルヌーイ分布をとるような試行、つまりベルヌーイ試行を複数回繰り返すような状況を考える。互いに独立したベルヌーイ試行を n 回繰り返したとき、ベルヌーイ試行の確率変数 X が 1 となるような状況が y 回起こった時を考える。 Y はもちろん確率変数となる。この時、この事象の組み合わせは $nC_y = \frac{n!}{(n-y)!y!}$ であり、それぞれが起きる確率は $\theta^y (1 - \theta)^{n-y}$ となるため、

$$f(y) = P(Y = y) = \frac{n!}{(n-y)!y!} \theta^y (1 - \theta)^{n-y} \quad (y = 0, 1, \dots, n) \quad (3,3)$$

3.1.4 ポアソン分布

二項分布をさらに発展させて考えてみる。ベルヌーイ試行の確率変数 X が 1 となる確率がとても小さい状況。しかしながら試行回数 n は膨大であるという状況を考える。この時 y 回確率変数 $X = 1$ となる試行が発生する確率を求める。

確率変数 $X = 1$ となる回数が Y となるときの期待値 $E[Y] = n\theta = \lambda$ とする。つまり

$$0 < n\theta = \lambda < \infty \quad (n \rightarrow \infty, \theta \rightarrow 0) \quad (3,4)$$

この時の 2 項分布の密度関数の極限として次が求めることができる。

$$f(Y) = P(Y = y) = \frac{\lambda^y}{y!} e^{-\lambda} \quad (3,5)$$

この密度関数を持つ離散型確率変数 Y は、ポアソン分布に従うといわれる。

3.1.5 幾何分布

ここで一度話をベルヌーイ試行まで戻す。初めて $X = 1$ となるまでの試行回数 Y の確率分布を幾何分布と呼ぶ。このとき次式で密度関数は示される。

$$P(Y = y) = (1 - \theta)^{y-1} \theta \quad (k = 1, 2, \dots) \quad (3,6)$$

3.2 連続型確率分布

3.2.1 一様分布

連続型確率変数 X が区間 (a, b) の間で、一様にどれかの値をとる可能性があるとき、

$$f(x) = \frac{1}{b - a} \quad (3,7)$$

この密度関数を持つ連続型確率変数 X は区間 (a, b) 上の一様分布に従うといわれる。これを $X \sim U(a, b)$ と表記する。なお期待値は、

$$E(X) = \frac{b + a}{2} \quad (3,8)$$

分散は、

$$V(X) = \frac{(b - a)^2}{12} \quad (3,9)$$

となる。

3.2.2 指数分布

ある期間に平均して λ 回起こる現象が、次起こるまでの期間 X が指数分布に従うとき、 $X = x$ となる確率密度関数は次式で示される。

$$f(x) = \lambda e^{-\lambda x} \quad (3,10)$$

この時この連続型確率分布を指数分布と呼ぶ。そしてこの状況を $X \sim Ex(\lambda)$ と表記する。なお期待値と分散は次のようになります。

$$E(X) = \frac{1}{\lambda} \quad (3,11)$$

$$V(x) = \frac{1}{\lambda^2} \quad (3,12)$$

3.2.3 正規分布

確率変数 X が次の密度関数を持つとき、**正規分布**に従うと言われ $X \sim N(\mu, \sigma^2)$ とあらわされる。

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (3,13)$$

急に式が出てきて混乱しかねないこの正規分布はどんな特徴があるの以下に示す。

- 平均値/最頻値/中央値が一致する。
- 平均値を中心に左右対称である。
- x 軸が漸近線となる。
- 分散が大きくなると、曲線の山は低くなり、左右に広がって平らになる。

3.3 モーメント母数関数