確率論と確率変数の復習
(教科書1.4.1-1.4.4 の内容)

担当: 長倉 大輔
(ながくらだいすけ)

確率論の基礎

- 確率とは?

確率とは 0 から 1 までの値をとり (%で表すなら 0 から 100)、物事が起きる**確からしさ**を表すものである。

確率が 1 に近いほどその物事が起こりやすい事を示している。

起きうる全ての物事の確率を足すと 1 (100%) になる。

確率論の基礎

- 確率に関連した用語

(**標本空間**) 起こりうる物事全てのあつまり。

(**根元事象**) 起こりうる物事。

(**事象**) 標本空間の中で任意の根元事象の組み合わせ(任意の部分集合)。その根元事象のうちの**どれかが起こる**事を意味する。

(**全事象**) 標本空間全体の事。**全事象を S と書く**。

（**空事象**） 起こりえない物事、もしくは根元事象が何も入っていない事象。
空事象は Φ という記号で表す。

確率論の基礎

- (例) サイコロの目

サイコロには{1}、{2}、{3}、{4}、{5}、{6} の 6 つの目がある。

(標本空間) {1, 2, 3, 4, 5, 6}の事。

(根元事象) {1}や{2}や{3}の事。

(事象) {1, 3} や {2, 5} や{2, 4, 6}の事。

（全事象）{1, 2, 3, 4, 5, 6}の事。

（空事象） {7} や{0}の事 (つまり起こりえない事象)。

(Φ ={7}, {0})

確率論の基礎

- 確率に関連した用語

(**和事象**) ある事象とある事象の和集合。

サイコロの例では例えば事象{2, 4} と事象{ 1, 3}
の和事象は

$$\{ 2, 4\} \cup \{1, 3\} = \{ 1, 2, 3, 4 \}$$

となる (**“∪” という記号で和事象を表す**)。

確率論の基礎

- 確率に関連した用語

(**積事象**) ある事象とある事象の共通集合。

サイコロの例では例えば事象{1, 2, 4}と事象{1, 3}の積事象は

$$\{1, 2, 4\}\cap\{1,3\} = \{1\}$$

となる (**“∩” という記号で積事象を表す**)。

確率論の基礎

- 確率に関連した用語

(**排反**)
ある事象とある事象が排反であるとはそれらが**共通の根元事象をもたない**という事である。サイコロを例にとると、事象{2, 4}と事象{1, 3}は排反しているという。これを
$\{2, 4\} \cap \{1, 3\} = \Phi$ と表す。

(**余事象**)
ある事象 A の余事象 A^C とはその事象と**排反しており**、その事象との**和集合が全事象**となる事象の事である
($A \cup A^C = S$ という事)。
例えば $A=\{1\}$ の余事象は $A^C=\{2, 3, 4, 5, 6\}$ である。

確率論の基礎

例題 1

サイコロの例において {2, 4, 6} の余事象を答えよ。

コインを投げると表か裏が出るとする。
以下の例題に答えなさい。

例題 2

コインを1度だけ投げる時の根元事象はどのようになるか?

例題 3

コインを 2 回投げる時の根元事象はどのようになるか?

確率論の基礎

- 確率の公理

確率とはそれぞれの**事象に対して付与され**、次の3 つの**公理**(公理とは常に満たされるとあらかじめ仮定する決まりごとの事)を満たす数字の事である。

(P1) 任意の事象が起きる確率は **0 以上 1 以下**。
(P2) 全事象が起こる**確率は 1** (つまり「全ての根元事象のうちどれかが起こる」という事象の起こる確率は 1)。
(P3) **排反な** 2 つの事象について、どちらかが起こる確率はそれぞれの事象が起こる確率の和。

確率論の基礎

- 確率の公理の数学的表現

ある事象を E と表すとする(サイコロの例では $E=\{1, 3\}$ など)。その事象の確率を $\Pr(E)$ で表すとすると、先ほどの確率の3 つの公理は

(P1) $0 \leq \Pr(E) \leq 1$、
(P2) $\Pr(S) = 1$、
(P3) $E_1 \cap E_2 = \Phi$ の時、
$\Pr(E_1 \cup E_2) = \Pr(E_1) + \Pr(E_2)$ 、

と表せる。

確率論の基礎

- 確率の性質

これら 3 つの確率の公理よりさまざまな性質を導く事ができる。

(1) 事象 E の余事象の起こる確率は 1 から事象 E が起こる確率をひいたものと等しい。

$$\mathbf{Pr}(E^c) = 1 - \mathbf{Pr}(E)$$

(2) 事象 E_1 が事象 E_2 の部分集合であれば (これを $E_1 \subseteq E_2$と書く)、

$$\mathbf{Pr}(E_1) \leq \mathbf{Pr}(E_2)$$

確率論の基礎

- 確率の性質

(3) (**排反ではない**) 2 つの事象 E_1 、E_2について

$$\mathbf{Pr}(E_1 \cup E_2) = \mathbf{Pr}(E_1) + \mathbf{Pr}(E_2) - \mathbf{Pr}(E_1 \cap E_2)$$

が成り立つ。

(4) 空事象Φの起きる確率は 0 である ($\mathbf{Pr}(\Phi) = 0$)。

(例)　$\mathrm{Pr}(\{7\}) = 0$。

確率論の基礎

例題 4

袋の中に赤玉 1 個と白玉 2個が入っているとする。
玉を2個取り出す時に

(1) 1つ目に白玉、2つ目に赤玉がでる確率は？

(2) 少なくとも、1つ目か2つ目のどちらかが赤玉である確率は？

どの玉が取り出される確率も全て等しいとする。

例題 5

2 つの事象　$E_1=\{1, 3, 5\}$, $E_2=\{1, 2\}$について、$\Pr(E_1) = 0.5$, $\Pr(E_2) = 0.2$ であり, さらに $\Pr(\{1\}) = 0.1$ であるとする。この時、事象 $E_3 = \{1, 2, 3, 5\}$ の確率 $\Pr(E_3)$ を求めよ。

確率論の基礎

- 条件付確率

事象 E_1 が生じたという条件のもとで事象 E_2 が生じる確率を E_1 のもとでの E_2 の**条件付確率**と呼び、

$$\Pr(E_2 \mid E_1)$$

と書く。

(例)

袋の中に赤玉 1 個と白玉 2 個が入っているとする。
玉を2個取り出す時に、1つめに白玉が出たという条件付きで2つめに白玉が出る確率はいくつか？

確率論の基礎

- 条件付確率の公式 1

$\Pr(E_2 | E_1)$ は

$$\Pr(E_2 \mid E_1) = \frac{\Pr(E_2 \cap E_1)}{\Pr(E_1)}$$

で与えられる。これを先ほどの(例)に当てはめると、
$E_1 = \{1回目に白玉がでる\}$、$E_2 = \{2回目に白玉が出る\}$

$$\Pr(E_2 \mid E_1) = \frac{\Pr(E_2 \cap E_1)}{\Pr(E_1)} = \frac{1/3}{2/3} = \frac{1}{2}$$

となる。

確率論の基礎

例題 6

さきほどの(例)において1つ目に赤玉が出たという条件付きで、2つ目に白玉がでる確率はいくつか？

例題 7

お隣さんが飼っている犬が 2 匹の子犬を生みました。お隣さんはあなたに 2 匹の犬を飼ってもらいたいと考えています。あなたは2匹ともオスであれば飼いたいので、2匹ともオスであるか質問したところ、お隣さんは1匹だけ性別を確認しており、「少なくとも1匹はオスでした」と答えました。さてこの時、「少なくとも1匹がオス」という条件付きで、残りのもう1匹もオスである確率はいくつでしょう？

確率論の基礎

- 条件付確率の公式 2

条件付確率の公式 1 を書き直して

$$\Pr(E_2 \cap E_1) = \Pr(E_2 | E_1) \Pr(E_1)$$

を得る。(同様に $\Pr(E_2 \cap E_1) = \Pr(E_1 | E_2) \Pr(E_2)$ もいえる)。

例題 8

1 組のトランプ(ジョーカーなし)から 2 枚を抜き出す時、2 枚ともエースとなる確率はいくつか？ただし1 枚引いたあとに、それを戻さないで2枚目を引くする。またどのカードも等しい確率で抜き出されるとする。

確率論の基礎

- 独立

ある2つの事象が $\Pr(E_1|E_2)=\Pr(E_1)$ を満たす時(もしくは $\Pr(E_2|E_1)=\Pr(E_2)$ の時)、この2つの事象は(確率的に)**独立**であるという。これはこの2つの事象がお互いの**確率に影響を及ぼさない**という事を表している。この時、

$$\Pr(E_2 \cap E_1)=\Pr(E_2)\Pr(E_1)$$

が成り立つ。

- 独立な事象の例

コインを2回投げた時、2回目に表が出る確率は1回目の結果に依存しないのあれば、1回目の結果と2回目の結果は独立である。

確率論の基礎

- ベイズの定理

事象 E_1 が起こった時の、事象 E_2 の条件付確率を求める公式である。18世紀にイギリスのベイズ(Bayes)が考案した。

事象 E_1 を結果、事象 E_2 と $E_2{}^C$ をその考えうる原因であるとした場合、ベイズの定理はある結果 E_1 が起こった時に、それぞれの考えうる原因が実際にその原因である確率を示してくれる。

Pr(結果|原因)から Pr(原因|結果)がわかる。

確率論の基礎

- ベイズの定理

今、$\Pr(E_2)$、$\Pr(E_2^C)$、$\Pr(E_1|E_2)$、$\Pr(E_1 | E_2^C)$ がわかっているとすると、ベイズの定理(公式)は

$$\Pr(E_2 | E_1) = \frac{\Pr(E_1 | E_2)\Pr(E_2)}{\Pr(E_1 | E_2)\Pr(E_2) + \Pr(E_1 | E_2^C)\Pr(E_2^C)}$$

で与えられる。

ベイズの定理は「事象 E_2 が起きる確率が E_1 という情報を得る事によって更新された」と考える事もできる。このような解釈の場合 $\Pr(E_2)$ を**事前確率**、$\Pr(E_2|E_1)$ を**事後確率**という。

確率論の基礎

- ベイズの定理の応用

(病気の検査)
ある病気を検査するための検査薬があるとしよう。その検査薬を用いれば、もし実際に病気であれば99%の確率(0.99)で陽性反応が出るが、実際には病気でなくても2%の確率(0.02)で間違って陽性(偽陽性)となってしまうとしよう。

	陽性反応	陰性反応
病気である時	99%	1%
病気でない時	2%	98%

確率論の基礎

- ベイズの定理の応用

全人口の 0.1% (0.001) がこの病気にかかっているとしよう。この時、もしこの検査薬によって陽性反応が出た場合に実際に病気である確率はいくつになるか？
求めたいのは Pr(病気|陽性)である。ベイズの定理によると

$$
\Pr(\text{病気}\,|\,\text{陽性}) = \frac{\Pr(\text{陽性}\,|\,\text{病気})\Pr(\text{病気})}{\Pr(\text{陽性}\,|\,\text{病気})\Pr(\text{病気}) + \Pr(\text{陽性}\,|\,\text{健康})\Pr(\text{健康})}
$$

$$
= \frac{0.99 \times 0.001}{0.99 \times 0.001 + 0.02 \times 0.999} = \frac{99}{99 + 1998} \approx 0.047
$$

となり、この検査薬で陽性反応が出ても実際に病気である確率は**5%しかない!!**。

確率論の基礎

- ベイズの定理 (2つ以上の事象の場合)

一般的に事象 D の原因としてありうる事象が $E_1, E_2, \ldots, E_n$ と複数ある場合は E_j $(j = 1, .., n)$ の D という条件付確率は

$$\Pr(E_j \mid D) = \frac{\Pr(D \mid E_j)\Pr(E_j)}{\sum_{i=1}^{n}\Pr(D \mid E_i)\Pr(E_i)}$$

となる。ただしここで E_1 から E_n は互いに排反で $(E_i \cap E_j = \Phi,\ i, j = 1, \ldots, n)$, $E_1 \cup E_2 \cup \ldots \cup E_n = S$ と仮定する。

確率論の基礎

例題 9

ある製品のメーカーとして、A、B、C の3 社がある。この製品は外見上まったく同じで区別がつかないとする。3社の市場シェアは、A社 50 %、B 社 30 %、C社20%である。過去の経験では、各社製品の不良率は A 社10 %、B 社 20 %、C 社 25% である。いま、どのメーカーのものかわからずに、この製品を 1 個買ったところ、それが不良品であった。それが、

(1) A社である確率、および (2) B社である確率、

求めよ。

問題

銀行は融資をする時に審査を行い、審査結果が A (融資可)の場合は融資が行われ、A^C(融資不可)の場合は融資が行われないとしよう。審査対象の全企業中、A の割合は 30%、A^C の割合は70%であった。データより審査結果が A であればそのうち10%、A^Cであれば(もし融資していたとしたら)そのうちの90% が返済不可能になる事がわかっているとしよう。返済不可能になる事象を D としよう。

$$\Pr(A) = 0.3,\ \ \Pr(A^C) = 0.7,\ \ \Pr(D \mid A) = 0.1,\ \Pr(D \mid A^C) = 0.9$$

である。

(1) ベイズの定理を用いて、返済不能になる人がAと判断される確率 $\Pr(A \mid D)$ を求めなさい。

(2) 返済可能であるのに融資不可と審査されてしまう確率 $\Pr(A^C \mid D^C)$ を求めなさい。

離散型確率変数

- 確率変数とは？

ある現象が確率的に起こったときにその現象に対応する数値の事を**確率変数**という。

株価収益率などは数値そのものだが、お天気などの場合にも、晴れの時には「1」、曇りの時には「2」などのようにそれぞれの現象に数値を対応させる事ができる。

そのような変数を考えれば、それは確率変数である。

離散型確率変数

- 確率変数の種類

(離散型確率変数)
サイコロの目や、晴れの時には「1」、曇りの時には「2」のように離散的に値をとる確率変数を**離散型確率変数**という。

(連続型確率変数)
株価収益率、身長、体重などのように連続的に値をとる確率変数を**連続型確率変数**という。

離散型確率変数

- 確率関数

離散型確率変数 X は m 個の実数 x_i , $i=1,\ldots, m$ を実現値として取り、それぞれの実現値の確率は p_i , $i=1,\ldots, m$ であるとしよう。この時、この離散型確率変数 X の**確率関数** $p_X(.)$ は

$$p_X(x_i) = p_i, \qquad 0 < p_i \leq 1,$$

を満たし、また $y \neq x_i$ に対しては

$$p_X(y) = 0$$

を満たす関数である(慣例として大文字 X で確率変数、小文字 x で確率変数 X が取りうる値 を表す)。

離散型確率変数

つまり**確率関数**とは離散型確率変数の取りうる値にその**確率**を対応させる**関数**の事である。

p_i , i =1,.., m は確率であるから

$$\sum_{i=1}^{m} p_i = p_1 + \cdots + p_m = 1$$

を満たす。

離散型確率変数

(例1)　コイン投げ
確率変数 X はコインを投げて表が出たら $X = 1$、裏が出たら $X = 0$ という実現値を取るとしよう。

表が出る確率も裏が出る確率も 1/2 とする。この時、X の確率関数 $p_X(x)$ は

$$p_X(1) = 1/2, \quad p_X(0) = 1/2, \quad p_X(y) = 0 \text{ for } y \neq 1, 0$$

となる。

離散型確率変数

(例2) サイコロの目
サイコロの目を離散型確率変数とみなし、それを X としよう。X の取りうる値は $X = 1, 2, 3, 4, 5, 6$ である。

全ての目は同じ確率で出るとすると、この X の確率関数 $p_X(x)$ は

$$p_X(1) = 1/6,\quad p_X(2) = 1/6,\quad p_X(3) = 1/6,$$
$$p_X(4) = 1/6,\quad p_X(5) = 1/6,\quad p_X(6) = 1/6,$$
$$p_X(x) = 0 \quad \text{for } x \neq 1, 2, 3, 4, 5, 6$$

となる。

離散型確率変数

- 累積分布関数

ある実数 x に対して確率変数 X が x 以下の値である確率 $\Pr(X \leq x)$を対応させる関数を**累積分布関数**という。

この定義は連続型確率変数に対してもあてはまる。

離散型確率変数

離散型確率変数の場合、具体的には以下のようになる。

ある離散型確率変数 X が m 個の実数: $x_1 < x_2 < \ldots < x_m$ を取るとしよう。この時、X の**累積分布関数** $F_X(x)$ は

$$F_X(x) = \Pr(X \le x) = \sum_{i=1}^{j} p_X(x_i)$$

と定義される。ここで $p_X(x)$ は X の確率関数、x_j は $x_1, \ldots, x_m$ のうち **x 以下で最も大きい数**である。

離散型確率変数

(例3)　サイコロの目

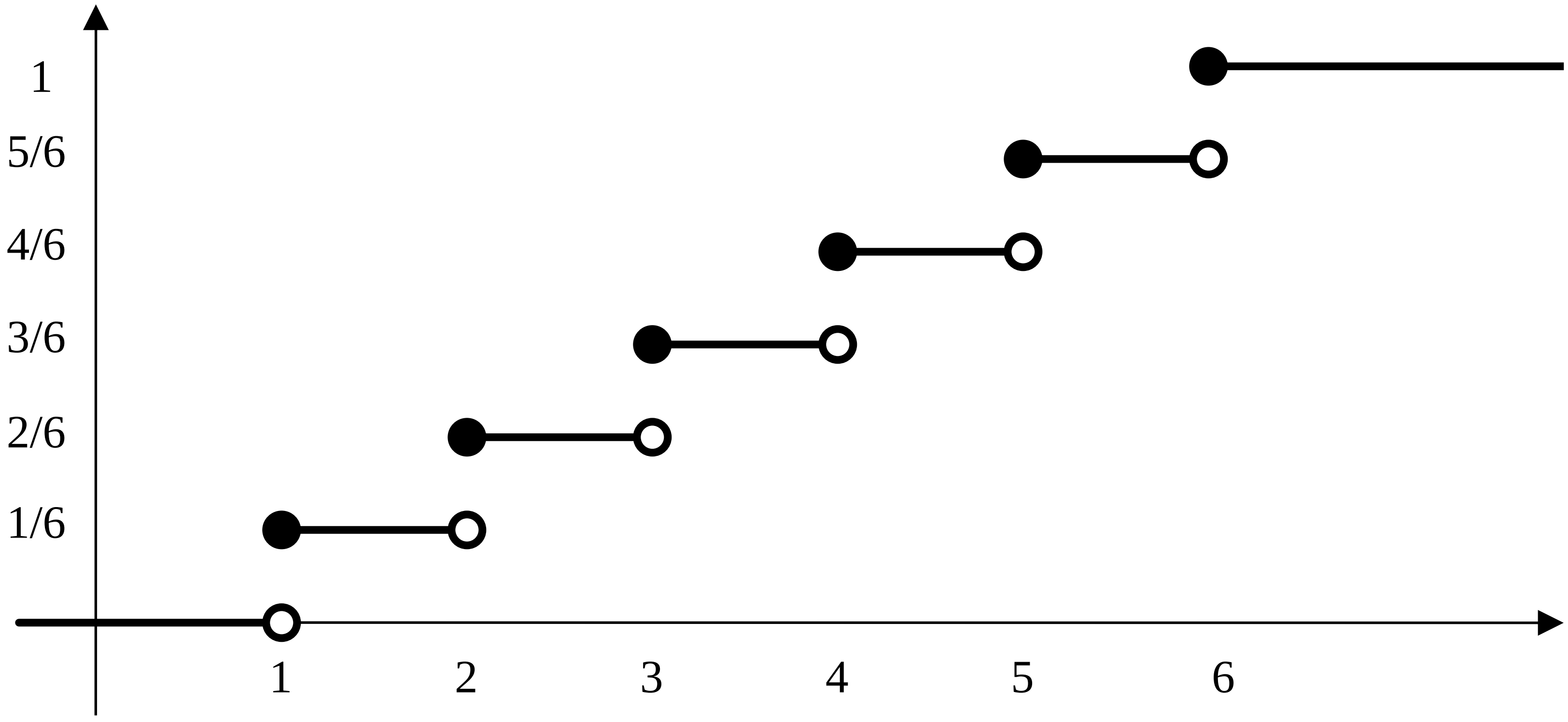
先ほどのサイコロの例では

$$F(x)=\Pr(X\leq x)=\begin{cases}0, & x<1\\ 1/6, & 1\leq x<2\\ 2/6, & 2\leq x<3\\ 3/6, & 3\leq x<4\\ 4/6, & 4\leq x<5\\ 5/6, & 5\leq x<6\\ 1, & 6\leq x\end{cases}$$

となる。

離散型確率変数

(サイコロの目の分布関数のグラフ)



離散型確率変数

- 平均と分散

データを要約する際に標本平均と標本分散を計算した。

同様に、**確率分布を要約する**ために確率分布に対しても**平均**と**分散**を定義する。

離散型確率変数の平均と分散は以下のように定義される。

確率変数の平均は**期待値**とも呼ばれる。今後はこの2つの単語を併用する。

離散型確率変数

離散型確率変数 X が m 個の実数 $x_1, \ldots, x_m$ を取り、その確率関数を $p_X(x_i)$ としよう。この時、X の平均は

$$E(X) = \mu = x_1 p_X(x_1) + \ldots + x_m p_X(x_m)$$
$$= \sum_{i=1}^{m} x_i p_X(x_i)$$

分散は

$$\mathrm{var}(X) = \sum_{i=1}^{m} (x_i - \mu)^2 p_X(x_i)$$

と定義される。$E(X)$ と $\mathrm{var}(X)$ はそれぞれ平均と分散を表す(英語の Expectation (期待値) と Variance (分散)の略)。

離散型確率変数

観測数 n の標本平均はそれぞれの観測値に **$1/n$ という重みを付けた加重平均**であるのに対して、確率分布の平均は取りえる値のそれぞれをその値の**確率で重みをつけた加重平均**である。

確率分布の平均と分散についても、平均は**分布の中心**、分散(およびその平方根をとった標準偏差)は**分布のばらつき具合**を表すという解釈は標本平均、標本分散の場合と同じである。

離散型確率変数

- 平均と分散の性質

標本平均や標本分散で成り立つ性質は確率分布の平均や分散でもそのまま成り立つ。以下は代表的なものである。

(平均)

(1) 定数 c について $E(c) = c$,

(2) $E(X - \mu) = 0$,

(3) $E(aX + b) = a\mu + b$,

(4) $E(X + Y) = E(X) + E(Y)$

(分散)

(1) $\mathrm{var}(X) = E(X^2) - \mu^2$,

(2) $\mathrm{var}(aX + b) = a^2\mathrm{var}(X)$,

離散型確率変数

- ベルヌーイ分布

コインを投げて表か裏が出るような、

(1) 起こりうる出来事(根元事象)の数が 2 つしかなく、
(2) 表が出れば $X = 1$、裏が出れば $X = 0$ のようにそれぞれの起こりうる出来事に 0 と 1 が対応し、
(3) それぞれの確率が

$$\Pr(X = 1) = p,\ \Pr(X = 0) = 1 - p,\ 0 < p < 1$$

で与えられる確率変数 X の分布を**ベルヌーイ分布**という。

離散型確率変数

例題 2 (ベルヌーイ分布の期待値と分散)

ベルヌーイ分布の期待値と分散が

$$E(X) = p,\ \ \mathrm{var}(X) = (1 - p)\,p$$

となる事を示しなさい。

離散型確率変数

- 同時確率と周辺確率

表が出る確率も裏が出る確率も共に1/2 の 2 枚のコインをそれぞれ コイン1, コイン2 とし、X_i をコイン i (i =1, 2) の目に対応する確率変数とする。

X_i はコインi を投げて、表が出たとき 1 を取り、裏が出たときに 0 を取るとする。この時、X_1, X_2 が共に 1 を取るような確率、すなわち

$$\Pr(X_1 = 1, X_2 = 1)$$

のような確率を X_1 と X_2 の**同時確率**という。

離散型確率変数

同時確率に対して

$$\Pr(X_i = 1) = 1/2\ ,\quad \Pr(X_i = 0) = 1/2,\quad i = 1, 2$$

を**周辺確率**とよぶ。

離散型確率変数

- 独立

2 つの(離散型)確率変数 X_1 と X_2 に対して、その同時確率がそれぞれの周辺確率の積と等しいとき、すなわち

$$\Pr(X_1 = k, X_2 = j) = \Pr(X_1 = k) \Pr(X_2 = j) ,$$

(k, j は X_1, X_2 が取りうる値) が成り立つとき、この 2 つの(離散型)確率変数は**独立**であるという。

離散型確率変数

- 同時確率と周辺確率の関係

さきほどのコインの例で同時確率と周辺確率には

$$\Pr(X_1=1) = \Pr(X_1=1, X_2=1) + \Pr(X_1=1, X_2=0)$$

のような関係がある事がわかる。これは、根元事象が

{コイン1, コイン2} = {表表}, {表裏}, {裏表}, {裏裏}

の4つであり、これらは互いに排反であるので、

$$\begin{aligned}\Pr(X_1=1) &= \Pr(\{表表, 表裏\}) \\ &= \Pr(\{表表\}) + \Pr(\{表裏\}) \\ &= \Pr(X_1=1, X_2=1) + \Pr(X_1=1, X_2=0),\end{aligned}$$

となる事よりわかる。

離散型確率変数

- 同時確率関数と周辺確率関数

先ほどの例において、例えば X_1 と X_2 がそれぞれ k と j になる確率を与える関数

$$p_{12}(k, j) = \Pr(X_1 = k, X_2 = j)$$

をこの2つの確率変数の**同時確率関数**という。

またこの時、X_1 の確率関数 $p_1(x)$は**周辺確率関数**と呼ばれる場合がある。

X_1 と X_2 が独立であれば、$p_{12}(k, j) = p_1(k)\, p_2(j)$ となる。

離散型確率変数

- 一般の場合

2 つの確率変数 X と Y があり、X は $x_1, x_2, \ldots, x_m$ の m 個の値を正の確率で取り、Y は $y_1, y_2, \ldots, y_n$ の n 個の値を正の確率で取るとする。この時、X と Y の同時確率関数は

$$p_{XY}(x_i, y_j) = \Pr(X = x_i, Y = y_j),\ i = 1, 2, \ldots, m,\ \ j = 1, 2, \ldots n,$$

であり、その周辺確率関数は同時確率関数の和と等しく

$$p_X(x_i) = \Pr(X = x_i) = \sum_{j=1}^{n} p_{XY}(x_i, y_j), \quad i = 1, 2, \ldots, m$$

$$p_Y(y_j) = \Pr(Y = y_j) = \sum_{i=1}^{m} p_{XY}(x_i, y_j), \quad j = 1, 2, \ldots, n$$

を満たす。

離散型確率変数

同時確率関数は、定義より

$$\sum_{i=1}^{m}\sum_{j=1}^{n} p_{XY}(x_i, y_j) = 1$$

を満たす

離散型確率変数

- 条件付確率関数

条件付確率に対しては**条件付確率関数**が定義される。

条件付確率関数の値を求めるには同時確率関数の値を周辺確率関数の値でわってやればよい。先ほどのコインの例では、例えば $X_1 = 1$ という条件付の $X_2 = 0$ の条件付確率関数 $p_{2|1}(0|1)$ の値は

$$p_{2|1}(0 \mid 1) = \Pr(X_2 = 0 \mid X_1 = 1)$$

$$= \frac{\Pr(X_1 = 1,\ X_2 = 0)}{\Pr(X_1 = 1)} = \frac{p_{12}(1,\ 0)}{p_1(1)} = \frac{1/4}{1/2} = \frac{1}{2}$$

となる。X_1と X_2 は独立なのでこれは $p_2(0)$に等しい。

離散型確率変数

- 条件付平均と条件付分散

離散型確率変数 X は $x_1, \ldots, x_m$ の m 個の値を取り $X=x_i$ の $Y=y_j$ という条件付確率関数の値は $p_{X|Y}(x_i \mid y_j) = p_{i\mid j}$ で与えられているとする。この時 X の**条件付平均** と **条件付分散**は

(条件付平均)

$$E(X \mid Y = y_j) = \mu_{X|Y=y_j} = \sum_{i=1}^{m} x_i p_{i\mid j}$$

(条件付分散)

$$\mathrm{var}(X \mid Y = y_j) = \sum_{i=1}^{m} (x_i - \mu_{X|Y=y_j})^2 p_{i\mid j}$$

のように定義される。

離散型確率変数

- 共分散

離散型確率変数 X と Y の**共分散** cov(X, Y) は

$$\mathrm{cov}(X,Y) = E\{(X-\mu_X)(Y-\mu_Y)\}$$
$$= \sum_{i=1}^{m}\sum_{j=1}^{n} p_{ij}(x_i-\mu_X)(y_j-\mu_Y)$$

によって定義される(cov は英語の covariance (共分散) の略)。ここで μ_X と μ_Y はそれぞれ X と Y の平均であり、p_{ij} は $p_{ij} = p_{XY}(x_i, y_i) = \Pr(X = x_i, Y = y_j)$ である。

定義より $\mathrm{cov}(X, X) = \mathrm{var}(X)$ となることに注意。

離散型確率変数

- 共分散の性質

(1) $\mathrm{cov}(X, Y) = E(XY) - \mu_X \mu_Y$

ここで $\mu_X = E(X), \mu_Y = E(Y)$, および

$$E(XY) = x_1 y_1 p_{11} + x_1 y_2 p_{12} + \dots + x_m y_n p_{mn}$$

$$= \sum_{i=1}^{m} \sum_{j=1}^{n} x_i y_j p_{ij}$$

である。

(2) 2つの確率変数 X と Y が独立であれば、

$$\mathrm{cov}(X, Y) = 0,$$

(3) $\mathrm{var}(X + Y) = \mathrm{var}(X) + \mathrm{var}(Y) + 2\mathrm{cov}(X, Y)$

(4) $\mathrm{cov}(aX, bY) = ab\,\mathrm{cov}(X, Y)$

離散型確率変数

- 相関係数

確率変数 X と確率変数 Y の**相関係数**は

$$\rho_{XY} = \frac{\mathrm{cov}(X,Y)}{\sqrt{\mathrm{var}(X)}\sqrt{\mathrm{var}(Y)}}$$

によって定義される。標本相関係数の場合と同様、

$$-1 \leq \rho_{XY} \leq 1$$

であり、単位の変化に影響されず**線形関係の強さを測る**ものである。

離散型確率変数

相関係数の定義より $\rho_{XY} = 0$ となるのは $\mathrm{cov}(X, Y) = 0$ となる時、およびその時に限る。

相関係数は X と Y の間の線形関係の強さを測るものである。よってもし X と Y が独立 (つまり何の関係もない)であれば $\rho_{XY} = 0$ (つまり $\mathrm{cov}(X, Y) = 0$) である。しかしながら、$\mathrm{cov}(X, Y) = 0$ であっても ***X* と *Y* は独立とは言えない**事に注意が必要である($Y = X^2$ のようは非線形の関係があるかもしれない）。

離散型確率変数

- 離散型確率変数のベイズの定理

今2つの離散型確率変数 X と Y がありそれぞれ x_i, i=1,…,M, y_j, j =1,…, N の値を取るとする(M, N は ∞でもよい)。今、$X = x_i$ という条件付きの $Y = y_j$ の確率を $\Pr(y_j \mid x_i)$ と表すとすると、これは

$$\Pr(y_j \mid x_i) = \frac{\Pr(x_i \mid y_j)\Pr(y_j)}{\sum_{j=1}^{N} \Pr(x_i \mid y_j)\Pr(y_j)}$$

と書き換えられる。

右辺の分母は実際には $\Pr(Y = y_j)$ と等しい事に注意。

連続型確率変数

- 連続型確率変数

株価収益率、身長、体重、などのように連続的に値をとる確率変数を**連続型確率変数**という。

連続的に値をとるとは、どんなに短い区間にも無限に取りうる値があるという事である。例えば 0.5 と 0.51 の間には 0.501が、0.501 と 0.502の間には 0.5011が…というように取りうる値がびっしりと詰まっている。

連続型確率変数に対しては離散型確率変数のように 1 つ 1 つの値に確率を付与するという事ができない。
よって点ではなく**幅に対して**確率を定義する。

連続型確率変数

- 密度関数

離散型確率変数の確率関数に対応するものとして、連続型確率変数に対して、**密度関数**を定義する。

連続型確率変数 X に対しては X が区間$[a , b]$ の中の値を取る確率: $\Pr(a \leq X \leq b)$ を考える。この確率が

$$\Pr(a \leq X \leq b) = \int_a^b f(x)dx$$

のようにある関数 $f(x)$ の積分で与えられる時、この $f(x)$ をこの確率変数 X の**密度関数**という。

連続型確率変数

- 密度関数の性質

密度関数は以下の性質を持つ。

(1) $\int_{-\infty}^{\infty} f(x)dx = \Pr(-\infty < x < \infty) = 1.$

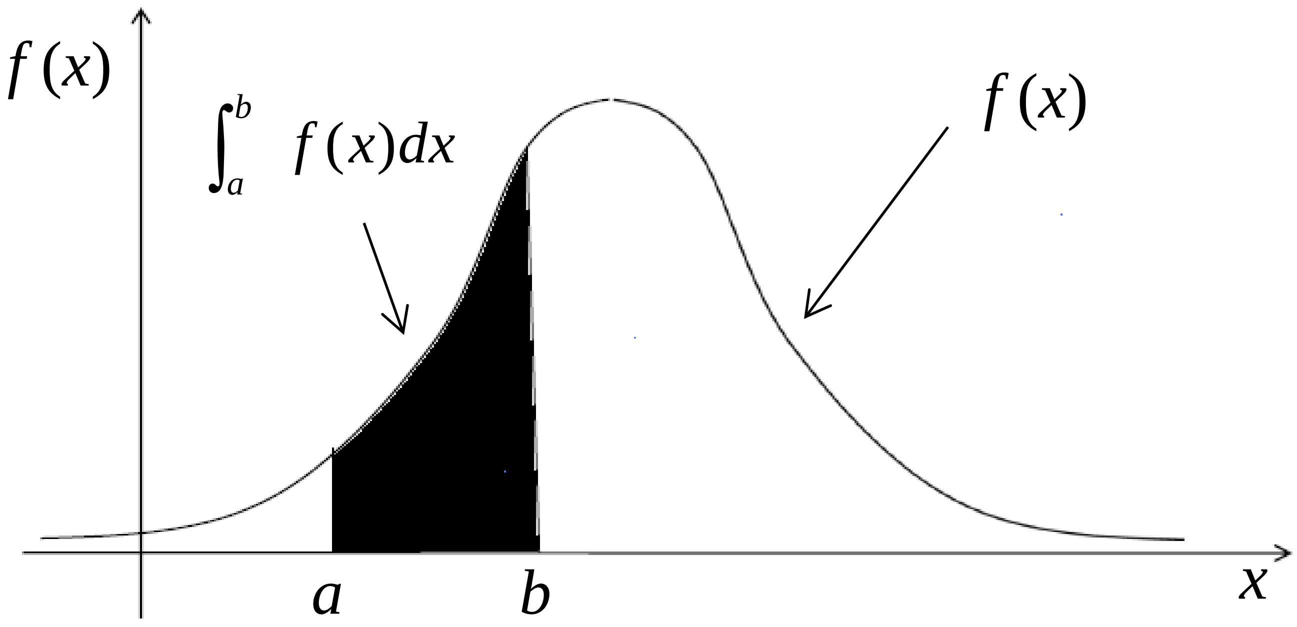
(2) 実現値として取りうる x に対して $f(x) > 0$.

ある連続型確率分布の特徴は全て密度関数の形状によって決まる。密度関数が同じであれば同じ分布である。

連続型確率変数

- 密度関数の図形的意味

確率変数 X に対して $\Pr(a \le X \le b)$ は密度関数 $f(x)$ の区間 $[a, b]$ での積分、$\int_a^b f(x)\mathrm{d}x$ で与えられる。図形的には下図の黒い部分の**面積の値**の事である。



連続型確率変数

- 1 点の確率

連続型確率変数 X に対して$\Pr(X = a)$の確率はどのような値をとるだろうか？

連続型確率変数の確率は上のように面積で定義されるが、$\Pr(X = a)$ は線であり面積を持たない。つまり

連続型確率変数が 1 点を取る確率は 0 である

と考えるのである（$\Pr(X = a) = 0$）。

連続型確率変数

連続型確率変数のこの性質より

$$\Pr(a \le X) = \Pr(a < X) + \Pr(X = a) = \Pr(a < X),$$

であり、また同様に

$$\Pr(a \le X \le b) = \Pr(a < X < b)$$

となる。

連続型確率変数

- 連続型確率変数の平均と分散

連続型確率変数の期待値と分散は密度関数を用いて、以下のように定義される。

(平均)

$$\mu = E(X) = \int_{-\infty}^{\infty} x\, f(x)dx$$

(分散)

$$\mathrm{var}(X) = E[(X-\mu)^2] = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx$$

連続型確率変数の平均と分散も 離散型確率変数のスライドで述べられた性質を全て満たす。

連続型確率変数

- 連続型確率変数の共分散

連続型確率変数 X, Y の共分散は

$$\mathrm{cov}(X,Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-\mu_X)(y-\mu_Y)f_{XY}(x,y)\,\mathrm{d}x\,\mathrm{d}y$$

と定義される 。ここで $f_{XY}(x, y)$ は**同時密度関数**と呼ばれるもので

(1) $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x,y)\,\mathrm{d}x\,\mathrm{d}y = 1,$

(2) 実現値として取りうる x, y に対して $f_{XY}(x, y) > 0$,

(3) $\Pr(a < X < b,\ c < Y < d) = \int_{c}^{d}\int_{a}^{b} f_{XY}(x,y)\,\mathrm{d}x\,\mathrm{d}y,$

を満たす x, y の2変量関数である。

連続型確率変数

連続型確率変数の共分散も離散型確率変数のスライドで述べられた性質を全て満たす。ただし性質(1) における $E(XY)$ は

$$E(XY) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\, f_{XY}(x, y)\,\mathrm{d}x\,\mathrm{d}y$$

で置き換えられる。

連続型確率変数

- 連続型確率変数の相関係数

離散型の場合と同じように

$$\rho_{XY} = \frac{\mathrm{cov}(X, Y)}{\sqrt{\mathrm{var}(X)}\sqrt{\mathrm{var}(Y)}}$$

と定義される。$|\rho_{XY}| \leq 1$ を満たす。

連続型確率変数

- 連続型確率変数の独立性

連続型確率変数 X と Y が**独立**であるとは同時密度関数が X と Y の**密度関数の積として表せる**事である。すなわち

$$f_{XY}(x, y) = f_X(x)\, f_Y(y)$$

となる事である。この時 $f_X(x)$と $f_Y(y)$を特に**周辺密度関数**と呼ぶ事もある。

連続型確率変数

- 条件付密度関数

連続型確率変数 X, Y について X の $Y=y$ という**条件付の密度関数**は

$$f_{X|Y}(x \mid y) = \frac{f_{XY}(x, y)}{f_Y(y)}$$

と定義される。

- 同時密度関数の分解

上の定義式の両辺に $f_Y(y)$ をかけると同時密度関数は

$$f_{X|Y}(x \mid y) f_Y(y) = f_{XY}(x, y)$$

と**条件付密度関数と周辺密度関数の積**として表す事ができる。

連続型確率変数

■ 条件付期待値、分散

連続型確率変数 X, Y について X の $Y=y$ という**条件付期待値と分散**はそれぞれ

$$\mu_{X|Y} = E(x\,|\,y) = \int_{-\infty}^{\infty} x\, f_{X|Y}(x\,|\,y)dx$$

および

$$\sigma^2_{X|Y} = \mathrm{var}(x\,|\,y) = \int_{-\infty}^{\infty} (x-\mu_{X|Y})^2\, f_{X|Y}(x\,|\,y)dx$$

と定義される。周辺密度関数が条件付き密度関数に置き換わっただけである。

連続型確率変数

- 連続型関数のベイズの定理

今2つの連続型確率変数 X と Y がありそれぞれ $a < X < b$, $c < Y < d$ の範囲の値を取るとする(a, c は $-\infty$, b, d は∞でもよい)。今、$X = x_i$ という条件付きの Y の条件付き密度関数の y_j における値を $f(y_j \mid x_i)$ と表すとすると、これは

$$f_{Y|X}(y_j \mid x_i) = \frac{f_{X|Y}(x_i \mid y_j) f_Y(y_j)}{\int_c^d f_{X|Y}(x_i \mid y) f_Y(y) dy}$$

と書き表せる。
ここで右辺の分母は実際には $f_X(x_i)$ と等しい事に注意。