確率論における基礎知識

平成 20 年 11 月 16 日
東京大学　大学院経済学研究科
河野　愛一朗

１．基本的な用語の定義

１．１　試行（**Trial**）

どんな結果が起こるか，偶然に（一定の確率的に）決まるような出来事を起こすことを試行という．実験や観察，くじ，将来の出来事などが例として挙げられる．

１．２　事象（**Event**）

上の試行によって，発生した出来事のことを意味する．例えば，「宝くじが当たるということ」とか,「テストで合格すること」などが例として挙げられる．事象は，よく英数字の大文字で表される．例えば，Ａ＝(テニスの試合に勝利)などのようにおくことがある．この事象の記号は，単なる記号に過ぎず，変数や定数のように値をとるものとは全くことなるので注意すること．逆に，集合論で使用される論証は，この事象においても同様に使用できる．

１．３　同時（**Joint**）

２つの事象がどちらも起こるとしよう. 例えば, 山崎くんと中村くんがいて, ２人がそれぞれ所得を得る．このとき，山崎くんと中村くんの両方が１０００万円の所得を得た．このような場合を，確率論では，「**同時に**」起きたという．つまり，確率論の言う「同時」とは，時間軸的に“ぴったし同時に”などという意味とは全く異なる．単に，両方の出来事が起きたということを指す．

１．４　背反

背反は逆に，同時に起こりえないことを指す．例えば，「明日のあるテストで川浦さんが７０点をとる」ということと，同じ「明日のあるテストで川浦さんが４０点をとる」ということは，背反と言える．

１．５　独立（**Independent**）

ある事象の発生（試行）が別の事象の発生（試行）に影響を与えないとしよう．そのようなとき，それらの事象（試行）は「独立である」と呼ぶ．

例えば，箱があって，その中に当たりのボールとハズレのボールがあるとしよう．ボールをとるという事象を行い，当たりかハズレか確認するわけだが，まず，取ったボールは箱に戻して次の試行を続けて行うとしよう．このとき，ｎ回目の試行の際に当たりかハズレかによって，ｎ＋１回目の試行で当たり・ハズレの確率は変化する．これは独立とは言えない（従属：dependent という）．逆に，取ったボールは箱に戻さずに次の試行を続けて行うとしよう．この場合，ｎ回目の試行もｍ回目の試行も，ボールの中にある当たり・ハズレのボールの数は不変なので，どちらも当たり・ハズレの確率は変化しない．この場合，ｎ回目に当たりが出る事象とｍ回目に当たりが出るという事象は独立であると言える．

２．論理記号の使用

２．１　確率への事象の導入（コルモゴロフの公理[1]）

- $P(A)$：事象 A が発生する確率（Probability）を指す．他は，Prob(A)や Pr(A)など表す場合もある．必ず，$0 \leq P(A) \leq 1 \quad (\forall A)$　となる．

- 例１：全事象（起こりうる全ての事象）を U，空事象（何も起こらない事象）を，φ とする．このとき，P(U)=1，P(φ)=0 となる．

[1] Kolmogorov axioms

- 例２：$P(A\cap B)$ は,「事象 A と事象 B が同時に発生する確率」を指す．逆に, $P(A\cup B)$ は,「事象 A と事象 B の少なくとも一方（両方起きてもよい）が発生する確率」を指す.

２．２　確率の加法性と背反

例えば，集合論では,

$$A\cup B = A + B - A\cap B \quad (\forall A, B).$$

同様な話は，確率論においても言え,

$$\underline{P(A\cup B) = P(A) + P(B) - P(A\cap B) \quad (\forall A, B).} \tag{2.2.1}$$

ここで背反であるときを考えよう.

$$\text{事象 A と事象 B が背反} \quad \Leftrightarrow \quad P(A\cap B) = P(\phi) = 0$$

より，これに(2.2.1)を代入すれば,

$$\underline{P(A\cup B) = P(A) + P(B) \quad (\forall A, B)}$$ [2]

となる．事象が３つ以上の場合でも同様である．つまり， $P(\bigcup_i A_i) = \sum_i P(A_i)$

２．３　余事象とその応用

- 定義：ある事象に対して，その事象が起きないという事象を考える．その事象を余事象といい，例えば，事象 A の余事象は， $\overline{A}$ と表される．ちなみに, A と $\overline{A}$ は明らかに背反である.

一般に集合論では,

$$A + \overline{A} = U \quad (\forall A).$$

よって，

$$P(A) + P(\overline{A}) = P(U) = 1 \quad (\forall A). \tag{2.3.1}$$

---

[2] （別解）A と B が背反　$\Leftrightarrow$　$A + B = A\cup B$　$\Leftrightarrow$　$P(A) + P(B) = P(A\cup B)$

- 例３：事象 A と事象 B の少なくとも一方が発生する確率

(2.3.1) ⇒ $$P(A \cup B) + P(\overline{A \cup B}) = 1 \tag{2.3.2}$$

ド・モルガンの法則[3]： $$\overline{A \cup B} = \overline{A} \cup \overline{B} = \overline{A} \cap \overline{B} \tag{2.3.3}$$

(2.3.2)に(2.3.3)を代入すると，

$$P(A \cup B) + P(\overline{A} \cap \overline{B}) = 1$$

$$\Leftrightarrow \quad P(A \cup B) = 1 - P(\overline{A} \cap \overline{B}).$$

「少なくとも一方が発生する確率」というのは求めにくいが，その否定を考えることによって，若干容易に求められるようになる場合がある．

２．４　独立

例えば，事象 A と事象 B が独立であれば，

$$P(A \cap B) = P(A)P(B)$$

となる．事象が３つ以上の場合でも同様である．つまり，$P(\bigcap_i A_i) = \prod_i P(A_i)$.

２．５　条件付確率（**Conditional Probability**）

- 定義：事象 A と事象 B が存在する．このとき，事象 A がすでに起きている時点での事象 B が起こる確率を，A が起こったときの B の条件付確率といい，$P_A(B)$ や $P(B \mid A)$ などと表す．

- 例４：岡田くんと渡辺くんの出席率

		渡辺		合計
		出席	欠席	
岡田	出席	0.05	0.55	0.6
	欠席	0.25	0.15	0.4
合計		0.3	0.7	1.0

[3] De Morgan’s law

　ここで，岡田くんが出席する事象をA，渡辺くんが出席する事象をBとする．つまり，それぞれが欠席する事象は，$\overline{A},\overline{B}$となる．上の表から，$P(A)=0.6$，$P(\overline{A})=0.4$，$P(B)=0.3$，$P(\overline{B})=0.7$となる．しかし，表をよく見ると，$A$と$B$は独立でない．独立であれば，$P(A\cap B)=P(A)P(B)$だが，$P(A\cap B)=0.05$であり，$P(A)P(B)=0.18$と一致しない．つまり，従属しているのである．岡田くんは本来，6 : 4 の確率で出席：欠席だが，渡辺くんが出席しているときに限れば，1 : 5 割合でしか，出席していない．つまり，渡辺くんが出席するという条件で，岡田くんが出席する確率：$P(B|A)$は 1／6 だと言える．渡辺くんにおいても同様に考えることができる．1／6 は，0.05／0.3，つまり，$P(A\cap B)$／$P(A)$で求められた．

- 定理：条件付確率の通常の確率（Unconditional Probability）への変換

$$\underline{P(B|A)=\frac{P(A\cap B)}{P(A)}} \tag{2.5.1}$$

もし，AとBが独立であれば，$P(A\cap B)=P(A)P(B)$（∵２．４）より，

$$P(B|A)=\frac{P(A)P(B)}{P(A)} \quad \Leftrightarrow \quad \underline{P(B|A)=P(B)}.$$

- 例５：$P(A\cap B)$の求め方

以下の４通りが考えられる．

一般的には，以下が成立する．

(2.3.1)　$\underline{\Rightarrow \quad P(A\cap B)=1-P(\overline{A}\cup\overline{B})}$

(2.2.1)　$\underline{\Leftrightarrow \quad P(A\cap B)=P(A)+P(B)-P(A\cup B)}$

(2.5.1)　$\underline{\Leftrightarrow \quad P(A\cap B)=P(B|A)P(A)}$

もし，AとBが独立であれば，

$$\underline{P(A\cap B)=P(A)P(B)} \text{（∵２．４）}.$$

３．期待値

３．１　確率変数

- 定義：確率変数（Random variable）とは，ある値をとることが事象になるような変数のことをいう．

つまり，実際にどんな値をとるか，偶然に（一定の確率的に）決まる変数のことを指す．例えば，「明日行われる試験の点数」や「じゃけんを６回行い，その中で勝利する数」などが該当する．ここで注意すべきことは，確率変数は値をとりうる変数であるのであって，確率自体ではない．確率変数と確率の対応関係を確率分布（Probability distribution）[4]という．

- 例６：あるダーツゲームにおいて，藤代さんと城さんの各得点

藤代			
得点	0	5	20
確率	0.5	0.2	0.3

城			
得点	0	5	20
確率	0.3	0.6	0.1

この表における得点は，確率的に定まっているので，確率変数と言える．また，この表自体も，確率変数と確率の対応関係を示すので，確率分布と言える．

確率変数は文字でおくことが多いので，実践してみよう．藤代さんの得点を，X とする．X=0,5,20 となる．ここで，X かそれぞれの値（実現値という）をとるときの確率はどのように表されるか？確率変数の定義に基づけば，値をとること，すなわち，X=0，X=5，X=20 それぞれが事象になる．よって，２．１における確率の表記の仕方によれば，$P(X=0)=0.5$，$P(X=5)=0.2$，$P(X=20)=0.3$ が確率である．

次に，城さんの確率変数を X とおくことはできるだろうか？藤代さんと城さんのケースを一つの問題として考える場合，基本的には文字を変えるべきである．なぜならば，藤代さんと城さんは異なる確率分布（つまり表）をとる

[4] 複数の確率変数（事象）の組と確率の関係を表したものを同時分布（Joint distribution）という．例えば，例４の表がそれに当たる．

からである. よって, 例えば, 城さんの確率変数を Y とおいて, Y=0,5,20 となる. X=0,5,20 で実現値は同じだが，$P(Y=0)=0.3$，$P(Y=5)=0.6$，$P(Y=20)=0.1$ となり確率分布が異なるため, 意味合いが違ってしまうので, 違う文字にすべきである.

３．２　期待値の定義

- 以下，確率変数 X の期待値を考える．実現値はｎ個存在し，$X = x_1, x_2, \cdots, x_n$ とする．通常，確率変数は大文字，実現値は小文字で表す.

- 定義：確率変数 $X = x_1, x_2, \cdots, x_n$ の期待値（Expectation）を $E(X)$ と表し，

$$\underline{E(X) = P(X = x_1)x_1 + P(X = x_2)x_2 + \cdots + P(X = x_n)x_n}$$

となる．つまり，$E(X) = \sum_{i=1}^{n} P(X = x_i)x_i$ と書ける．このとき，確率の総和の $\underline{P(X = x_1) + P(X = x_2) + \cdots + P(X = x_n) = 1}$，つまり，$\sum_{i=1}^{n} P(X = x_i) = 1$ であることに注意すること.

- 例７：あるダーツゲームにおいて，藤代さんと城さんの各得点

藤代			
得点	0	5	20
確率	0.5	0.2	0.3

城			
得点	0	5	20
確率	0.3	0.6	0.1

藤代さんと城さんの確率変数をそれぞれ X,Y とする.

$$\begin{aligned} E(X) &= P(X = 0)\cdot 0 + P(X = 5)\cdot 5 + P(X = 20)\cdot 20 \\ &= 0.5\cdot 0 + 0.2\cdot 5 + 0.3\cdot 20 \\ &= 7 \end{aligned}$$

$$\begin{aligned} E(Y) &= P(Y = 0)\cdot 0 + P(Y = 5)\cdot 5 + P(Y = 20)\cdot 20 \\ &= 0.3\cdot 0 + 0.6\cdot 5 + 0.1\cdot 20 \\ &= 5 \end{aligned}$$

となる.

４．ベイズの定理

４．１　２事象のケース

(2.5.1)より，条件付確率の定義は,

$$P(B\mid A)=\frac{P(A\cap B)}{P(A)} \quad \Leftrightarrow \quad P(A\cap B)=P(B\mid A)P(A).$$

同様に，
$$P(A\mid B)=\frac{P(A\cap B)}{P(B)} \quad \Leftrightarrow \quad P(A\cap B)=P(A\mid B)P(B).$$

よって，
$$P(B\mid A)P(A)=P(A\mid B)P(B)$$
となるので,

つまり，
$$P(A\mid B)=\frac{P(B\mid A)P(A)}{P(B)}. \qquad (4.1.1)$$

これをベイズの定理[5]という.

● 例８：予防接種

ある国民病に対する予防接種（例えば，結核に対するＢＣＧ接種とツベルクリン反応のようなもの）を考える．この予防接種をしなければ子供のうちに必ず感染してしまうとする．そこで，生まれてすぐ行われる予防接種の後に実施される検査では，陽性か陰性の反応を示す．予防接種された人が成年したのちに行われる調査によって，病気にかからなかった人には高い割合で，その検査で陽性が出ていたことが分かった．具体的に，病気にかからなかった人が全体の 78%，検査で陽性だった人が全体の 72%，病気にかからなかった人の中から陽性だった人の割合が 90%となっている．この結果から，この検査で陽性であるときに予防接種で免疫がついている確率を求めよう.

ここで，病気にかからない，つまり，免疫がついているという事象を A，検査で陽性である事象を B とする．すなわち，$P(A)=0.78$，$P(B)=0.72$，$\mathrm{P}(B|A)=0.90$ となる．ベイズの定理より,

$$P(A\mid B)=\frac{P(B\mid A)P(A)}{P(B)}=\frac{0.90\times 0.78}{0.72}=0.975$$

なので, 求める確率, すなわち, 陽性であるときに免疫がついている確率は 97.5%だと判断できる.

[5] Bayes’ theorem, Bayes’ rule

ちなみに，今回のケース（２事象の場合）では以下のような同時分布の表を描くことでも求めることができる[6].

	陽性	陰性	合計
成功	0.70	0.08	0.78
失敗	0.02	0.20	0.22
合計	0.72	0.28	1.00

４．２　３事象のケース

全事象 U を互いに背反な２つの事象である A_1, A_2 に分割する．

つまり，$U = A_1 + A_2 \quad \Leftrightarrow \quad U \cap B = A_1 \cap B + A_2 \cap B$.

$$\Leftrightarrow \quad P(U \cap B) = P(A_1 \cap B) + P(A_2 \cap B)$$

$$\Leftrightarrow \quad P(B) = P(A_1 \cap B) + P(A_2 \cap B). \tag{4.2.1}$$

(4.1.1)のベイズの定理を $P(A_1 \mid B)$ に適用し，

$$P(A_1 \mid B) = \frac{P(A_1 \cap B)}{P(B)}. \tag{4.2.2}$$

(2.5.1)の条件付確率の定義から，

$$P(B \mid A_1) = \frac{P(A_1 \cap B)}{P(A_1)} \quad \Leftrightarrow \quad P(A_1 \cap B) = P(B \mid A_1)P(A_1). \tag{4.2.3}$$

同様に，　$P(A_2 \cap B) = P(B \mid A_2)P(A_2)$. (4.2.4)

(4.2.1)を(4.2.2)に代入すると，

$$P(A_1 \mid B) = \frac{P(A_1 \cap B)}{P(A_1 \cap B) + P(A_2 \cap B)}.$$

これに，(4.2.3)と(4.2.4)を代入すると，

$$P(A_1 \mid B) = \frac{P(B \mid A_1)P(A_1)}{P(B \mid A_1)P(A_1) + P(B \mid A_2)P(A_2)} \quad （但し，U = A_1 + A_2）.$$

よって，$\underline{P(A_j \mid B) = \dfrac{P(B \mid A_j)P(A_j)}{P(B \mid A_1)P(A_1) + P(B \mid A_2)P(A_2)} \; (j = 1,2, U = A_1 + A_2)}$ (4.2.5)

となり，または，$P(A_j \mid B) = \dfrac{P(B \mid A_j)P(A_j)}{\sum_{i=1}^{2} P(B \mid A_i)P(A_i)}$ $(j = 1,2)$ とも書ける．(4.2.6)

これが，３事象の場合のベイズの定理となる．

---

[6] $P(A \mid B) = \dfrac{P(A \cap B)}{P(B)} = \dfrac{0.70}{0.72} = 0.972 \fallingdotseq 0.975$

● 例９：ウソ発見器

世の中に犯罪者が 5%存在しているとする．いま，世の中の人を無作為抽出し，ウソを言っているという兆候が現れたらアラームが鳴るように設計されているウソ発見器をかけたところ，犯罪者では 90%，一般人（犯罪者以外の全ての人）では 10%の確率でアラームが鳴るものとする．一見すると，このウソ発見器の精度は高そうである．このとき，この検査でアラームが鳴った者が犯罪者である確率を求めよう．

そこで，犯罪者がウソ発見器にかけられる事象を A_1，一般人がウソ発見器にかけられる事象を A_2，アラームが鳴る事象を B とする．よって，求める確率は，$P(A_1 \mid B)$ であり， $P(A_1)=0.05, P(A_1)=0.95, P(B \mid A_1)=0.9, P(B \mid A_2)=0.1$ となる．

(4.2.5)のベイズの定理に j=1 を代入すると，

$$P(A_1 \mid B)=\frac{P(B \mid A_1)P(A_1)}{P(B \mid A_1)P(A_1)+P(B \mid A_2)P(A_2)}$$

左辺に必要な確率を代入して，

$$P(A_1 \mid B)=\frac{0.9\times 0.05}{0.9\times 0.05+0.1\times 0.95}=\frac{0.045}{0.045+0.095}=\frac{0.045}{0.140}\fallingdotseq 0.32$$

よって，アラームが鳴った者が犯罪者である確率は 32%である．このように一見すると，精度が高そうだったが，圧倒的多数が一般人であるため，ウソ発見器が，犯罪者を見つける精度は低くなってしまうことが分かる．

４．３　多事象のケース

全事象 U を互いに背反な事象である $A_1, A_2, \cdots, A_n$ に分割すれば，(4.2.6)より，

$$P(A_j \mid B)=\frac{P(B \mid A_j)P(A_j)}{\sum_{i=1}^{n} P(B \mid A_i)P(A_i)} \quad \left(j=1,2,\cdots,n \mid U=\sum_{i=1}^{n} A_i \right)$$

となる．証明は，４．２と同様に成される．