

確率と確率分布

1 偶然な実験

統計学の理論展開を数学モデルとして取り扱うためには、確率の概念が必要となります。不確実性の数値化（モデル化）には、コイン投げ、さいころ投げを例として議論されますが、これには概念の定義に（1）再現性（繰り返すことができる）と（2）客観性（多くの人々が認める）をもつ対象を考えるからです。数学的には「測度論」とよばれる面積や体積を定めるための道具をつかいます。公理的な確率論とよばれる由縁です。抽象的にモデル化し、つぎのように確率空間とよび、偶然の現象を数学的に定義します。

偶然によって起こると考えられる全ての結果全体 Ω を考えます。結果のあつまり、集合、を意味します。つぎにそのうちのいくつかの結果を事象とよび、これを集めたものを A と表します。厳密にはすこし厄介になるので、一般には大きさを測る、あるいは数え上げができるものと解釈します。実際には、 P という記号で、事象の大きさ（可能性）を数値で表現します。まず確率の大きさをはかる対象となる集合を事象とよび、全事象と空事象、和事象、積事象、余事象（補事象）を定義されているとします。全事象 Ω は、すべての可能な結果全体で、当たり前なこと、この確率を 1 と定める。空事象 \emptyset は、起こり得ないことで、確率を 0 とします。和事象 $A \cup B$ とは少なくともひとつ、いづれかが起こること、積事象 AB , $A \cap B$ とは、すべてのものが同時に起こること、余事象 \bar{A} , A^c とは、これが起こらない、否定を意味します。

確率の定義；3つの組 (Ω, \mathcal{F}, P) が確率空間とは加法性と連続性をみたす集合関数（集合から数値を対応）のことです。つぎを満たすとします；

(a) 基準化 $P(\emptyset) = 0, P(\Omega) = 1$

(b) 加法性 $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

(c) 連続性 $A_i, i = 1, 2, \dots$ が $A_1 \subset A_2 \subset \dots, \lim_i A_i = \bigcup_{i=1}^{\infty} A_i$ は

$$P(\lim_i A_i) = P\left(\bigcup_{i=1}^{\infty} A_i\right) = \lim_i P(A_i)$$

確率の基本定理：

(1) $A \subset B$ ならば、 $P(B) = P(A) + P(B \setminus A) = P(A) + P(B \cap A^c) \geq P(A)$

(2) $P(A^c) = 1 - P(A)$

(3) $P\left(\bigcup_{i=1}^n A_i\right) = \sum_i P(A_i) - \sum_{i < j} P(A_i \cup A_j) + \sum_{i < j < k} P(A_i \cup A_j \cup A_k) \dots + (-1)^{n+1} P(A_1 \cup A_2 \cup \dots \cup A_n)$, ここで $\sum_{i < j}$ は $i < j$ を満たすすべての (i, j) の組にわたる和をとる。他も同様。

(4) $A_i, i = 1, 2, \dots$ が互いに素, すなわち $A_i \cap A_j = \emptyset (i \neq j)$ ならば $P(\bigcup_i A_i) = \sum_i P(A_i)$

(5) de Morgan の法則

$$P(\bigcup_i A_i) = 1 - P(\bigcap_i A_i^c)$$

少なくとも一つが起こる確率の計算は, 1 からすべてがまったく起こらない確率を引けばよい。

$$P(\bigcap_i A_i) = 1 - P(\bigcup_i A_i^c)$$

2 事象の独立

まず条件付き確率を定義する。ある事象が起こったときに, もう一つの事象にどの程度影響を及ぼしているかの尺度を表す。記号では $P(B|A)$ と表す。

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

ただし $P(A) > 0$ とする。事象 A は多少とも起こるものと前提条件をする。これを A が起こったときの B の起こる条件付き確率という。

(1) $P(A) = P(A|B)P(B) + P(A|B^c)P(B^c)$

(2) $P(A \cap B \cap C) = P(A)P(B|A)P(C|A \cap B), \quad P(A \cap B) > 0$

(3) ベイズの定理; 事象列 A_1, A_2, \dots, A_n を Ω の分割, つまり互いに素 (どの2つをとっても積事象が空事象) であり, これらすべての和集合が全事象になるとき, つぎの条件付き確率についての関係式が成り立つ。

$$P(A_j|B) = \frac{P(B|A_j)P(A_j)}{\sum_k P(B|A_k)P(A_k)}$$

事象 $A_k, k = 1, 2, \dots, n$ を原因と考え, 結果の事象 B の条件付き確率がわかっていれば, 結果からどの原因により起こっているかを考える。事前確率 $P(A_k), k = 1, 2, \dots, n$ から事後確率 $P(A_j|B)$ を計算する式を意味する。

独立事象の定義; 2つの事象の間について, 独立とは一方の起こる可能性に影響を与えないこと。

$P(B|A) = P(B)$ を書き換えて

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

が成り立つとき。一般に $A_i, i = 1, 2, \dots, n$ が独立であるとは,

$$P(\bigcap_{i=1}^n A_i) = P(A_1)P(A_2) \cdots P(A_n)$$

で n 個のすべての部分系 (1組のかたまりと考えて) が独立であるとき, で定義する。

3 確率試行と確率変数

確率試行 コイン投げ確率変数とは実験の結果 Ω から数値 (実数) を対応させるもの一般に $\{X \leq a\} = \{\omega \in \Omega | X(\omega) \leq a\}$ と表し, これが事象となるもの。すなわち $\{X \leq a\} \in \mathcal{F}$ とします。

5枚のコインを投げて, 表のでた枚数を X とおけば, 各結果を H (表, Head), T (裏, Tail) を使うと, $\{X = 0\} = \{TTTTT\}$, $\{X = 2\} = \{HHTTT, HTHTT, \dots, TTTHH\}$, $\{X = 3\} = \{HHHTT, HHTHT, \dots, TTTTH\}$ となります。また $\{X \leq 2\} = \{X = 0\} \cup \{X = 1\} \cup \{X = 2\}$, $\{X \leq \sqrt{2}\} = \{X = 0\} \cup \{X = 1\}$ などが成り立ちます。これは集合の包含関係で要素が同じであることを表しています。

4 確率分布

コイン投げやさいころ振りを繰り返して, この実験結果を整理まとめていくと, ある傾向が求められます。たとえば, 5枚のコインを投げて, 何枚表が出たかを記録してします。大体このような実験では, 確率的に変動しますから, 起こりやすいところ, 起こらないところの大小で, 高さを描いたグラフが記述統計で調べた度数分布表です。これに対して, 理論的にあるいは仮想的にそれぞれそのくらいの可能性があるかを計算します。これが, コイン投げの確率分布です。コイン投げの分布では, 可能性をできるだけ調べ上げて, 当てはまる事象の大きさを測ります。つまり確率を計算します。このときによく用いる概念が2項係数などで, 場合の数を効率よく数え上げようとするものです。上の例でいえば, 2^5 通りの結果から, 表が1枚出る事象は5通りあるので, $P(X = 1) = \frac{5}{2^5}$ となります。

和の分布; 数枚のコインを投げる場合では, 各コインの結果 (表が1, 裏がゼロ) をまとめて, 和を取ると, 表の出た枚数となります。一般にこれは和の分布を計算することです。独立な確率変数について,

$$P(X_1 + X_2 = a) = \sum_k P(X_1 = k)P(X_2 = a - k)$$

同時分布 (結合分布); 2変量 (X, Y) の結果について, 積事象 $\{X = a\} \cap \{Y = b\}$ を $(X, Y) = (a, b)$ と表し, $f_{X,Y}(a, b) = P(X = a, Y = b)$ を同時密度とよびます。周辺分布とは $f_X(a) = \sum_b f_{X,Y}(a, b)$, $f_Y(b) = \sum_a f_{X,Y}(a, b)$ をいいます。行列の表形式で与えられた2次元の度数分布表から周辺度数 (縦計や横計) を求めたことに相当します。また和の分布は

$$P(X_1 + X_2 = a) = \sum_{(x,y)|x+y=a} f_{X,Y}(x, y)$$

という意味で, ここで和 \sum は $x + y = a$ となるすべての (x, y) にわたる和をとります。最後の項が独立であれば, 確率の積になります。

以上は離散型確率変数ですが, 連続型の場合もほぼ同様です。

離散型確率変数;

$$P(a \leq X \leq b) = \sum_{\{k|a \leq k \leq b\}} P(X = k)$$

連続型確率変数;

$$P(a \leq X \leq b) = \int_a^b f_X(x) dx$$

ここで $F_X(x) = P(X \leq x)$ とおき, 微分ができるばあい, その微分係数 (導関数) を $f_X(x) = \frac{dF_X(x)}{dx} |_{x=c}$ とします。

5 期待値

平均の定義式；

$$E(X) = \begin{cases} \sum_i x_i f_X(x_i) & \text{離散型} \\ \int_{-\infty}^{\infty} x f_X(x) dx & \text{連続型} \end{cases}$$

離散型では $E[(X - a)^2] = \sum_i (x_i - a)^2 f_X(x_i) = \sum_i x_i^2 f_X(x_i) - 2a \sum_i x_i f_X(x_i) + a^2$ 連続型では $E[(X - a)^2] = \int_{-\infty}^{\infty} (x - a)^2 f_X(x) dx = \int_{-\infty}^{\infty} x^2 f_X(x) dx - 2a \int_{-\infty}^{\infty} x f_X(x) dx + a^2$ となります。

分散 $V(X)$ の定義は，上の式で， $a = E(X)$ とおいたもので，

$$V(X) = \begin{cases} \sum_i x_i^2 f_X(x_i) - \{E(X)\}^2 & \text{離散型} \\ \int_{-\infty}^{\infty} x^2 f_X(x) dx - \{E(X)\}^2 & \text{連続型} \end{cases}$$

一般に 2 変量確率変数 (X, Y) から定めた実数値関数の確率変数 $h(X, Y)$ の期待値は

$$E(h(X, Y)) = \begin{cases} \sum_i \sum_j h(x_i, y_j) f_{X,Y}(x_i, y_j) & \text{離散型} \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x, y) f_{X,Y}(x, y) dx dy & \text{連続型} \end{cases}$$

2 重積分やら，2 重和をとることが入って，すこし難しいかも知れませんが，1 変量が面積の計算，2 変量が体積の計算をしているだけです。平均は重心の計算と同じです。和の分散には， $h(x, y) = \{(x + y) - a\}^2$ として計算します。ここで $a = E(X + Y)$ は既に定数となっていることに注意。

6 練習問題

1 つぎの4つの命題は同値であることを示しなさい。

- (a) 事象 A と B は独立 (b) 事象 A^c と B は独立
(c) 事象 A と B^c は独立 (d) 事象 A^c と B^c は独立

2 もし $P(A|B) > P(A)$ ならば, $P(B|A) > P(B)$ を示せ。(ヒント: $\frac{P(A|B)}{P(B|A)} = \frac{P(A)}{P(B)}$ を示せ。)

3 X が平均 μ , 分散 σ^2 をもつとき, つぎの値を μ, σ^2 で表しなさい。

- (1) $E(3X + 2)$ (2) $E(X^2)$
(3) $E(2X^2 + 2X - 1)$ (4) $E[(X - 2)^2]$

4 分散に関するつぎの式を導きなさい。

- (1) $V(aX + b) = a^2V(X)$ (b にはよらない)
(2) X, Y が独立ならば, $V(X - Y) = V(X) + V(Y)$ (符号に注意, 差の分散でも分散の和になる)

5 つぎの表は離散型確率変数の結合密度を表すとします。

y の値		1	2	3
x の値	0	0.15	0.10	0.15
	1	0.10	0.0	0.10
	2	0.15	0.10	0.15

このとき,

- (1) $E(XY) = E(X)E(Y)$ が成り立つが, 独立ではないことを示せ。
(2) $V(X + Y)$ を計算しなさい。