

經濟統計學講義ノート No.4

離散確率分布

蛭川雅之

2024 年 10 月 4 日

目 次

1. 離散確率変数
2. 複数の離散確率変数
3. ベルヌーイ分布
4. 二項分布
5. ポアソン分布

1 離散確率変数

1.1 確率変数とは...

- 内容は記述統計から推測統計へと徐々にシフトする。
 - － まず手始めに、確率変数とは何かを学ぶ。
- そもそも**確率変数**とは何なのか？
 - － 日本工業規格では、確率変数を「どのような値となるかが、ある確率法則によって決まる変数。確率法則は確率分布で記述される。とることができる値が離散的であるか、連続的であるかによって、それぞれ離散（確率）変数、連続（確率）変数という。離散確率変数で表されるデータを計数値 (discrete variable)、連続確率変数で表されるデータを計量値 (continuous variable) という。」(JIS Z8101-1: 1999) と規定している。

- 確率変数は次のように定義される。
 - － 事象に対して実数値をとる変数である。
 - － そのとりうる実現値に確率が付与されている。
- 確率変数は次の 2 種類に大別される。
 - － **離散確率変数**... 実現値は有限個（例： $\{0, 1\}$, $\{1, 2, 3, 4, 5, 6\}$ ）
もしくは可算無限個（例： $\{0, 1, 2, 3, \dots\}$ ）
 - － **連続確率変数**... 実現値はある区間（例： $[0, 1]$, $(-\infty, \infty)$ ）内の任意の実数

■確率変数・実現値の表記

- 一般に、確率変数を大文字（例： X, Y, Z ）、実現値を小文字（例： x, y, z ）で書き表す。

1.2 離散確率変数と確率関数

- 離散確率変数 X の実現値が小さい順に $x_1, x_2, \dots, x_n$ であり、対応する確率が $p_1, p_2, \dots, p_n$ であるとする。

– ただし、 $p_i = \Pr(X = x_i), i = 1, \dots, n$ である。

- $p_1, p_2, \dots, p_n$ は、

$$0 < p_i < 1, \sum_{i=1}^n p_i = 1$$

を満たす。

- ここで、任意の実数 x に対し、関数 $p_X(x)$ を以下のように定義する。

$$p_X(x) = \begin{cases} p_i & : x = x_i, i = 1, \dots, n \\ 0 & : \text{それ以外} \end{cases}$$

– このような関数を確率関数という。

1.3 分布関数

- 離散確率変数 X の実現値を小さい順に $x_1, x_2, \dots, x_n$ とし、その確率関数を $p_X(x)$ とする。
- 任意の実数 x に対し、添え字 $j (\in \{1, \dots, n\})$ を $x_j \leq x$ となる最大の整数とすると、 X の **(累積) 分布関数** は次のように定義される。

$$F_X(x) = \Pr(X \leq x) = \Pr(X \leq x_j) = p_X(x_1) + \dots + p_X(x_j) = \sum_{i=1}^j p_i$$

– $F_X(x)$ のグラフは階段状である。

- 分布関数は以下の性質を持つ。
 1. $0 \leq F_X(x) \leq 1$
 2. $F_X(-\infty) = 0, F_X(\infty) = 1$
 3. $x < x^*$ を満たす x, x^* に対し、 $F_X(x) \leq F_X(x^*)$

- 確率関数および分布関数は、確率変数が

- (i) どのような値を
- (ii) どのような確率でとるのか

を示す。

- － 言い換えれば、確率変数がどのように分布しているかを示す。
- － 確率変数の分布を**確率分布**という。

問題 1 下表のようなクジがあり、確率変数 X をこのクジを 1 本引いて得られる賞金の額と定義する。このとき、 X の確率分布を求めよ。

1 等	2 等	3 等	4 等	5 等	ハズレ	
10000 円	5000 円	3000 円	1000 円	100 円	0 円	合計
1 本	2 本	7 本	40 本	150 本	800 本	1000 本

1.4 離散確率変数の期待値・分散・標準偏差

- 離散確率変数 X の実現値を $x_1, x_2, \dots, x_n$ とし、その確率関数を $p_X(x)$ とする。

1. X の期待値（平均値）は

$$\begin{aligned} E(X) &= \mu \\ &= \sum_{i=1}^n x_i p_X(x_i) \\ &= \sum_{i=1}^n x_i p_i \end{aligned}$$

と定義される。

2. X の分散は

$$\begin{aligned} \text{Var}(X) &= \sigma^2 \\ &= E(X - \mu)^2 \\ &= \sum_{i=1}^n (x_i - \mu)^2 p_X(x_i) \\ &= \sum_{i=1}^n (x_i - \mu)^2 p_i \end{aligned}$$

と定義される。

3. X の標準偏差はこの分散の正の平方根 $\sigma = \sqrt{\text{Var}(X)}$ である。

問題 2 問題 1 で定義した確率変数 X の期待値、分散、標準偏差をそれぞれ求めよ。

1.5 期待値・分散に関する性質

■期待値

- 定数 a, b に対し以下の性質が成り立つ。

$$E(aX + b) = aE(X) + b$$

$$E(b) = b$$

問題 3

$$Var(X) = E(X^2) - \mu^2 = E(X^2) - \{E(X)\}^2$$

が成り立つことを示せ。

■分散

- 定数 a, b に対し以下の性質が成り立つ。

$$\text{Var}(aX + b) = a^2 \text{Var}(X)$$

$$\text{Var}(b) = 0$$

問題 4 確率変数 X を標準化して得られる新しい確率変数

$$Z = \frac{X - \mu}{\sigma}$$

について、

$$E(Z) = 0$$

$$\text{Var}(Z) = 1$$

が成り立つことを示せ。

2 複数の離散確率変数

2.1 同時確率分布

- 2つの離散確率変数 X と Y について、 X は $x_1, x_2, \dots, x_n$ のいずれか、 Y は $y_1, y_2, \dots, y_m$ のいずれかをとるものとする。
- X が $x_i, i = 1, \dots, n$ かつ Y が $y_j, j = 1, \dots, m$ をとる確率を

$$p_{ij} = p_{XY}(x_i, y_j) = \Pr(X = x_i, Y = y_j)$$

と表記する。

– 1変数の場合と同様、

$$0 \leq p_{ij} < 1, \sum_{i=1}^n \sum_{j=1}^m p_{ij} = 1$$

が成り立つ。

- 2 つの離散確率変数の組 (X, Y) は以下のような同時確率分布を持つ。

$X \backslash Y$	y_1	$\cdots$	y_j	$\cdots$	y_m
x_1	p_{11}	$\cdots$	p_{1j}	$\cdots$	p_{1m}
$\vdots$	$\vdots$		$\vdots$		$\vdots$
x_i	p_{i1}	$\cdots$	p_{ij}	$\cdots$	p_{im}
$\vdots$	$\vdots$		$\vdots$		$\vdots$
x_n	p_{n1}	$\cdots$	p_{nj}	$\cdots$	p_{nm}

例 5 修学年数 X と時給 Y との間には関連があると考えられる。単純化のため、 X は 12, 16 (単位：年) のいずれか、一方、 Y は 800, 2000 (単位：円/時間) のいずれかをとるものとする。また、 (X, Y) の同時確率分布は以下のものであるとする。

$X \backslash Y$	800	2000
12	$3/8$	$1/8$
16	$1/8$	$3/8$

2.2 周辺確率分布

- X が $x_i, i = 1, \dots, n$ をとる確率は

$$p_X(x_i) = \Pr(X = x_i) = \sum_{j=1}^m \Pr(X = x_i, Y = y_j) = \sum_{j=1}^m p_{ij}$$

である。

- 同時確率分布の表では、各行の和として与えられる。

- 同様に、 Y が $y_j, j = 1, \dots, m$ をとる確率は

$$p_Y(y_j) = \Pr(Y = y_j) = \sum_{i=1}^n \Pr(X = x_i, Y = y_j) = \sum_{i=1}^n p_{ij}$$

である。

- 同時確率分布の表では、各列の和として与えられる。

- X の確率分布および Y の確率分布はそれぞれ表の周辺に現れる。
 - これらを (X および Y の) **周辺確率分布**という。

$X \backslash Y$	y_1	$\cdots$	y_j	$\cdots$	y_m	X の周辺確率
x_1	p_{11}	$\cdots$	p_{1j}	$\cdots$	p_{1m}	$p_X(x_1)$
$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
x_i	p_{i1}	$\cdots$	p_{ij}	$\cdots$	p_{im}	$p_X(x_i)$
$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
x_n	p_{n1}	$\cdots$	p_{nj}	$\cdots$	p_{nm}	$p_X(x_n)$
Y の周辺確率	$p_Y(y_1)$	$\cdots$	$p_Y(y_j)$	$\cdots$	$p_Y(y_m)$	1

問題 6 例 5 の同時確率分布について、 X および Y の周辺確率をそれぞれ計算し、下表の空欄を埋めよ。

$X \backslash Y$	800	2000	X の周辺確率
12	$3/8$	$1/8$	
16	$1/8$	$3/8$	
Y の周辺確率			

2.3 共分散

- 2つの離散確率変数 X と Y について、 X は $x_1, x_2, \dots, x_n$ のいずれか、 Y は $y_1, y_2, \dots, y_m$ のいずれかをとる場合、 X と Y の**共分散**は次のように定義される。

$$\begin{aligned} Cov(X, Y) &= E[\{X - E(X)\} \{Y - E(Y)\}] \\ &= E(XY) - E(X)E(Y), \end{aligned}$$

$$E(X) = \sum_{i=1}^n x_i p_X(x_i), E(Y) = \sum_{j=1}^m y_j p_Y(y_j),$$

$$E(XY) = \sum_{i=1}^n \sum_{j=1}^m x_i y_j p_{XY}(x_i, y_j) \left(= \sum_{i=1}^n \sum_{j=1}^m x_i y_j p_{ij} \right).$$

- $Cov(X, Y)$ は確率変数 X と Y の関係の強さを表す尺度である。
 - この尺度は X と Y の単位に依存する。

2.4 相関係数

- 2 つの離散確率変数 X と Y の分散がそれぞれ $Var(X)$ と $Var(Y)$ 、共分散が $Cov(X, Y)$ であるとき、 X と Y の相関係数は

$$Corr(X, Y) = \frac{Cov(X, Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}}$$

と定義される。

- $Corr(X, Y)$ は X と Y の線形関係の強さを表す。
- $-1 \leq Corr(X, Y) \leq 1$ が成り立つ。

問題 7 例 5 の同時確率分布について、 X と Y の共分散および相関係数を計算せよ。

2.5 独立性

- 2つの離散確率変数 X と Y について、すべての i, j に関し

$$p_{XY}(x_i, y_j) = p_X(x_i) p_Y(y_j)$$

が成り立つとき、 X と Y は互いに独立である（もしくは独立に分布している）という。

- X と Y が無相関であっても、必ずしも独立ではない。
 - X と Y が無相関とは、 X と Y に線形関係が見られないという意味である。
 - X と Y が独立ならば、これらは必ず無相関である。

問題 8 例5の同時確率分布について、 X と Y が独立に分布しているかどうか確認せよ。

2.6 2つの確率変数の期待値・分散に関する性質

■期待値に関する性質

- 確率変数 X と Y の和 $X + Y$ および積 XY の期待値に関して、次の性質が成り立つ。

1. 和の期待値は期待値の和に等しい。

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

2. 確率変数 X と Y が無相関である場合（独立である場合を含む）、積の期待値は期待値の積に等しい。

$$E(XY) = E(X)E(Y)$$

■分散に関する性質

- 確率変数 X と Y の和 $X + Y$ の分散は

$$Var(X + Y) = Var(X) + Var(Y) + 2Cov(X, Y)$$

である。

- 特に、確率変数 X と Y が無相関である場合（独立である場合を含む）、

$$Var(X + Y) = Var(X) + Var(Y)$$

が成り立つ。

2.7 条件付き確率分布・条件付き期待値

- 2つの離散確率変数 X と Y について、 $X = x_i, i = 1, \dots, n$ を与えたときの Y の条件付き確率分布は

$$\begin{aligned} p_{Y|X}(y_j|x_i) &= \Pr(Y = y_j | X = x_i) \\ &= \frac{\Pr(X = x_i, Y = y_j)}{\Pr(X = x_i)} \\ &= \frac{p_{XY}(x_i, y_j)}{p_X(x_i)} \\ &= \frac{p_{ij}}{\sum_{j=1}^m p_{ij}}, j = 1, \dots, m. \end{aligned}$$

を計算することにより得られる。

- この条件付き確率分布を利用して、 $X = x_i, i = 1, \dots, n$ を与えたときの Y の条件付き期待値

$$E(Y|X = x_i) = \sum_{j=1}^m y_j p_{Y|X}(y_j|x_i) = \frac{\sum_{j=1}^m y_j p_{ij}}{\sum_{j=1}^m p_{ij}}$$

が得られる。

- $E(Y|X = x_i)$ は X の実現値 x_i によって変わる。

問題 9 例 5 の同時確率分布について、 X を与えたときの Y の条件付き確率分布および条件付き期待値を求めよ。

2.8 期待値繰り返しの法則

- X を与えたときの Y の条件付き期待値と Y の（無条件）期待値との間に

$$E(Y) = E\{E(Y|X)\}$$

が成り立つ。

- これを期待値繰り返しの法則という。

問題 10 例 5 の同時確率分布について、期待値繰り返しの法則が成り立つことを確認せよ。

2.9 全分散の法則

- X を与えたときの Y の条件付き期待値・分散と Y の（無条件）分散との間に

$$\text{Var}(Y) = \text{Var}\{E(Y|X)\} + E\{\text{Var}(Y|X)\}$$

が成り立つ。

- これを全分散の法則（もしくは、分散分解式）という。

問題 11 例 5 の同時確率分布について、全分散の法則が成り立つことを確認せよ。

3 ベルヌーイ分布

- ある事柄が起こる（成功）か起こらない（失敗）か二者択一の状況を考える。
 - － 例：客が来る・来ない、宝くじに当たる・外れる、コインの表が出る・裏が出る...
- 実現値が 0, 1 のみをとる確率変数 X を考える。
 - － 0 は失敗、1 は成功に対応する。
- X は成功確率 $p = p_X(1) = \Pr(X = 1)$ のベルヌーイ分布に従う。
 - － ベルヌーイ分布の期待値と分散は以下の通りである。

$$E(X) = p$$

$$\text{Var}(X) = p(1 - p)$$

4 二項分布

- 今度は成功確率 p のベルヌーイ試行を n 回行う場合を考え、確率変数 X を n 回の試行のうち成功する回数と定義する。
 - X の実現値は $x = 0, 1, \dots, n$ である。
 - $X = x$ となる確率は

$$p_X(x) = \Pr(X = x) = {}_n C_x p^x (1 - p)^{n-x}$$

で与えられ、

$$\sum_{x=0}^n p_X(x) = p_X(0) + p_X(1) + \cdots + p_X(n) = 1$$

が成り立つ。

- このとき、確率変数 X は二項分布 $B(n, p)$ に従うという。
 - n と p のように分布を特徴づける定数を母数（パラメータ）という。
 - 二項分布 $B(n, p)$ の期待値と分散は以下の通りである。

$$E(X) = np$$

$$Var(X) = np(1 - p)$$

問題 12 フリーキック決定率が 20% のサッカー選手が 1 試合に 4 回フリーキックを蹴るとする。このとき次の問いに答えよ。

1. ノーゴールの確率を求めよ。
2. 1 ゴールの確率を求めよ。
3. 2 ゴール以上の確率を求めよ。
4. ゴール数の期待値と標準偏差を求めよ。

5 ポアソン分布

- 二項分布 $B(n, p)$ において試行回数 n が極めて大きい場合、確率分布を正確に求めるのは煩雑である。
 - 計算に n の階乗が含まれることに由来する。
- 成功確率 p が極めて小さく、平均値を一定 (λ と表記する) とみなして差し支えない場合、二項分布を簡単な公式で近似できる。
 - λ の例：天災の頻度、印刷物の誤植の数、不良品の頻度...
 - 具体的には、二項分布 $B(n, p)$ の確率関数

$$p_X(x) = \Pr(X = x) = {}_n C_x p^x (1 - p)^{n-x}$$

に対し、 $np = \lambda$ の条件で $n \rightarrow \infty$ の極限をとる。

- 確率変数 X が確率関数

$$p_X(x) = \Pr(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}, x = 0, 1, 2, \dots$$

で与えられる分布を平均発生回数 λ のポアソン分布という。

- X の取りうる値に上限がない点に注意 ($\Leftrightarrow n \rightarrow \infty$ の場合の‘二項分布’)。
- e はネイピア数とよばれ、具体的には

$$e = \lim_{n \rightarrow \infty} \left(1 + \frac{1}{n}\right)^n = 2.718281828459045235360287471352 \dots$$

である。

- ポアソン分布の期待値と分散はともに

$$E(X) = \text{Var}(X) = \lambda$$

である。

問題 13 ある電器メーカーの工場で生産される液晶テレビには 1000 台に 4 台の割合で不良品が発生するという。いま 500 台の液晶テレビがある家電量販店に納品されたとする。この中に不良品が 0~4 台含まれている確率を二項分布（正確な計算）およびポアソン分布（近似計算）に基づいてそれぞれ求め、結果を比較せよ。【ヒント：Excel 上では、組み合わせの総数は関数 `combin`、 e のべき乗は関数 `exp`、階乗は関数 `fact` を用いて計算できる。】