統計学ノートNo. 4
離散確率分布

担当教員：蛭川雅之
研究室：紫英館3階345研究室
メール： hirukawa@econ.ryukoku.ac.jp
オフィスアワー：毎週金曜日3講時

1　確率変数

確率変数とは…

- **講義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。**
- **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。**

確率変数とは…（つづき）

- **確率変数は次の2種類に大別される。**
  1. **離散確率変数：実現値は有限個**

     $$\{0,1\}, \{1,2,3,4,5,6\}, \ldots$$

     もしくは可算無限個

     $$\{0,1,2,3, \ldots\}, \ldots$$

  2. **連続確率変数：実現値はある区間**

     $$[0,1], (-\infty, \infty), \ldots$$

     内の任意の実数

確率変数と実現値の一般的な表記

- **確率変数を大文字**

$$X, Y, Z, \ldots$$

で表す。

- **実現値を小文字**

$$x, y, z, \ldots$$

で表す。

2　離散確率変数

確率変数と確率分布

- 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。

確率変数と確率分布（つづき）

- **実現値**x_i**と確率**p_i**との対応関係は次表のようになる。**

X	x_1	x_2	$\cdots$	x_n	**計**
確率	p_1	p_2	$\cdots$	p_n	1

  - **この対応関係を**X**の確率分布という。**

確率変数と確率分布（つづき）

- $p_1, p_2, \dots, p_n$は、

$$(i)\ p_i \geq 0$$

$$(ii)\ \sum_{i=1}^{n} p_i = 1$$

を満たす。

確率分布：例

- 偏りのない正六面体のサイコロ大小2個を同時に投げるとき、出る目の和をXとすると、Xの確率分布は次のようになる。

X	2	3	4	5	6	7	8	9	10	11	12	計
確率	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$	1

分布関数

- 離散確率変数Xの実現値を小さい順に

$$x_1, x_2, \dots, x_n$$

とする。

- さらに、任意の実数xに対し、添え字$j(\in \{1, \dots, n\})$を

$$x_j \le x$$

となる最大の整数と定義する。

分布関数（つづき）

- **任意の実数xに対し、Xの（累積）分布関数は次のように定義される。**

$$F_X(x) = Pr(X \leq x) = Pr\left(X \leq x_j\right)$$

$$= p_1 + \cdots + p_j = \sum_{i=1}^{j} p_i$$

  - **$F_X(x)$のグラフは階段状である。**

分布関数（つづき）

- **分布関数は以下の性質を持つ。**

1. $0 \le F_X(x) \le 1$
2. $F_X(-\infty) = 0, F_X(\infty) = 1$
3. $x \le x^*$**を満たすに対し、**$F_X(x) \le F_X(x^*)$

確率変数の期待値

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

- **Xの期待値は次のように定義される。**

$$\mu = E(X) = x_1 p_1 + \cdots + x_n p_n = \sum_{i=1}^{n} x_i p_i$$

確率変数の分散と標準偏差

- Xの**分散**は次のように定義される。

$$\sigma^2 = Var(X)$$
$$= (x_1 - \mu)^2 p_1 + \cdots + (x_n - \mu)^2 p_n$$
$$= \sum_{i=1}^{n} (x_i - \mu)^2 p_i$$

- Xの**標準偏差**はこの分散の正の平方根

$$\sigma = \sqrt{\sigma^2}$$

である。

期待値・分散に関する性質

- 定数a, bに対し、以下の性質が成り立つ。

1. 期待値：

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

$$E(b) = b$$

2. 分散：

$$Var(aX + b) = a^2 Var(X)$$

$$Var(b) = 0$$

3　複数の離散確率変数

同時確率分布

- 2つの離散確率変数X, Yについて、
  - X の実現値が$x_1, x_2, \dots, x_n$
  - Y の実現値が$y_1, y_2, \dots, y_m$

  であるとする。
- さらに、

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

と表記する。

同時確率分布（つづき）

- (X, Y)の実現値(x_i, y_j)とその確率p_{ij}との対応（＝**同時確率分布**）が得られる。

$X \backslash Y$	y_1	$\cdots$	y_j	$\cdots$	y_m	計
x_1	p_{11}	$\cdots$	p_{1j}	$\cdots$	p_{1m}	p_1
$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
x_i	p_{i1}	$\cdots$	p_{ij}	$\cdots$	p_{im}	p_i
$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
x_n	p_{n1}	$\cdots$	p_{nj}	$\cdots$	p_{nm}	p_n
計	q_1	$\cdots$	q_j	$\cdots$	q_m	1

周辺確率分布

- Xが$x_i, i = 1, \ldots, n$をとる確率は

$$p_i = Pr(X = x_i) = \sum_{j=1}^{m} p_{ij}$$

　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。

周辺確率分布（つづき）

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

$$q_j = Pr(Y = y_j) = \sum_{i=1}^{n} p_{ij}$$

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。**

周辺確率分布（つづき）

- **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。**

共分散

- 確率変数XとYの**共分散**は次のように定義される。

$$Cov(X,Y) = E[\{X - E(X)\}\{Y - E(Y)\}]$$
$$= E(XY) - E(X)E(Y),$$
$$E(X) = \sum_{i=1}^{n} x_i p_i,$$
$$E(Y) = \sum_{j=1}^{m} y_j q_j,$$
$$E(XY) = \sum_{i=1}^{n} \sum_{j=1}^{m} x_i y_j p_{ij}$$

- $Cov(X,Y)$は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。

相関係数

- **確率変数**X**と**Y**の相関係数は次のように定義される。**

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

$$Var(X) = \sum_{i=1}^{n}\{x_i - E(X)\}^2 p_i,$$

$$Var(Y) = \sum_{j=1}^{m}\{y_j - E(Y)\}^2 q_j$$

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

独立性

- 離散確率変数XとYに関し、全てのi,jについて$p_{ij}=p_i q_j$が成り立つとき、XとYは互いに独立であるという。
- 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。

2つの確率変数の期待値・分散に関する性質

1. **確率変数X, Yと定数a, bに対し、和の期待値に関して以下のことが成り立つ。**

$$E(aX+bY)=aE(X)+bE(Y)$$

2. **確率変数X, Yが無相関（または独立）である場合、以下が成り立つ。**

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

2つの確率変数の期待値・分散に関する性質（つづき）

3. **確率変数X,Yと定数a,bに対し、和の分散に関して以下が成り立つ。**

$$Var(aX+bY)$$
$$= a^2Var(X) + b^2Var(Y) + 2abCov(X,Y)$$

- X,Y**が無相関（または独立）である場合、上の結果は以下のようになる。**

$$Var(aX+bY) = a^2Var(X) + b^2Var(Y)$$

4　代表的な離散確率分布

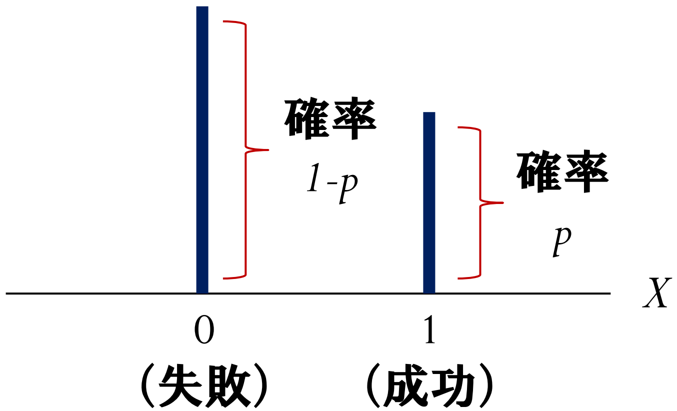
ベルヌーイ分布

- 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。

1. 客が来る・来ない
2. 宝くじに当たる・外れる
3. コインの表が出る・裏が出る
4. 視聴者がある番組を見る・見ない…

ベルヌーイ分布（つづき）

- **実現値が**$\{0,1\}$**のみをとる確率変数**X**を考える。**
  - $0 =$ **失敗；**$1 =$ **成功**
- X**は成功確率**$p = Pr(X = 1)$**のベルヌーイ分布に従う。**



ベルヌーイ分布（つづき）

- **ベルヌーイ分布に従う確率変数Xの期待値と分散は以下の通りである。**

$$E(X) = p$$

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

二項分布

- **偏りのない正六面体のさいころ1個を3回投げ、1の目が出る回数をXとする。**
  - **成功確率$p = \frac{1}{6}$のベルヌーイ試行を3回行うと考える。**

X	0	1	2	3	計
確率	${}_3C_0\left(\frac{1}{6}\right)^0\left(\frac{5}{6}\right)^3$	${}_3C_1\left(\frac{1}{6}\right)^1\left(\frac{5}{6}\right)^2$	${}_3C_2\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^1$	${}_3C_3\left(\frac{1}{6}\right)^3\left(\frac{5}{6}\right)^0$	1

二項分布（つづき）

- **一般に、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$x = 0,1, \ldots, n$である。**
- **$X = x$となる確率は**

$$Pr(X = x) = {}_nC_x p^x (1-p)^{n-x}$$

である。

二項分布（つづき）

- **このような確率分布を二項分布といい、$B(n,p)$で表す。**

- **二項分布$B(n,p)$に従う確率変数Xの期待値と分散は以下の通りである。**

$$E(X) = np$$

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

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。
- 発生確率pが極めて小さく、平均値npを一定（$np=\lambda$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…

ポアソン分布（つづき）

- **確率変数Xをそのような事象の発生回数とする。**
- **平均発生回数λに対し、$X = x$となる確率は**

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

$e = \lim_{n\to\infty}\left(1 + \frac{1}{n}\right)^{n} = 2.7182818 \ldots$ **：ネイピア数**

で与えられる。

ポアソン分布（つづき）

- **このとき、確率変数Xは平均発生回数λのポアソン分布に従うという。**
- **平均発生回数λのポアソン分布に従う確率変数Xの期待値と分散は以下の通りである。**

$$E(X) = Var(X) = \lambda$$

ポアソン分布：例題

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2台含まれている確率を求めよ。

ポアソン分布：例題（つづき）

- **不良品の発生確率**：$p = \frac{4}{1000}$
- **不良品の平均台数：**

$$\lambda = np = 500 \times \frac{4}{1000} = 2$$

- **確率変数**X**を不良品の台数とすると、**$X = x$**となる確率は次のようになる。**

$$Pr(X = x) = \frac{e^{-2}2^x}{x!} = \frac{1}{e^2} \times \frac{2^x}{x!}$$

ポアソン分布：例題（つづき）

- **求める確率は、**

$$Pr(X=2)=\frac{1}{e^2}\times\frac{2^2}{2!}$$

$$\approx\frac{1}{(2.7182818)^2}\times\frac{4}{2}$$

$$\approx\frac{2}{7.389056}$$

$$\approx 0.2707$$