データサイエンスのための統計学

確率変数と確率分布 (1)

by 尚晋　(名古屋大学経済学研究科助教)
on 2025 年 5 月 27 日

目次

目次

* 確率変数と確率分布
* 確率変数の期待値と分散
* おもな離散型の確率分布

» 離散型の確率変数と確率分布

離散型確率変数 とりうる各値に対しそれぞれ確率が与えられている変数を確率変数という. 確率変数は, X のように大文字を用いて表す. 例えば、サイコロを振って出る目 X は確率変数である. 正しいさいころなら、$P(X=1)=1/6, P(X=2)=1/6, ..., P(X=6)=1/6$ となる。ここで、一般に書くと
$p_1 \geq 0, p_2 \geq 0, ..., p_6 \geq 0, p_1+p_2+...+p_6=1$ が成り立つ。

離散型確率分布 一般に可算集合 $\{x_1, x_2, ...\}$ の中の値をとる確率変数 X は離散型と言われ, それぞれ値の確率:

$$P(X=x_k)=f(x_k) \quad (k=1,2,...) \tag{1}$$

を X の確率分布 (probability distribution)という, ただし, f は

$$f(x_k) \geq 0, \quad (k=1,2,...) \text{かつ} \sum_{k=1}^{\infty} = 1 \tag{2}$$

の条件を満たす. この f を離散型の確率分布という.

» 離散型確率分布

さらに、2 個のさいころを同時に振って出る目 X_1, X_2 の和 $X_1 + X_2$ も確率変数で, 各値の確率は表 5.3, 図 5.3 となる。

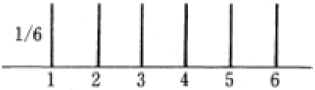

図 5. 2　さいころの目の確率
どの目も等確率で出現する．離散型一様分布と呼ばれる．

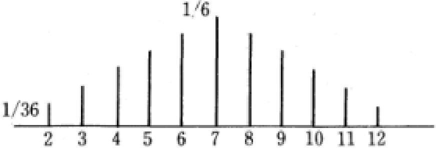

図 5. 3　2 個のさいころの目の和の確率
和の確率分布に峰(モード)が出現する．これは中心極限定理の初期段階である(第 8 章参照)．

表 5. 3　2 個のさいころの目の和の確率の表

和 (和/2)	2 1	3 3/2	4 2	5 5/2	6 3	7 7/2	8 4	9 9/2	10 5	11 11/2	12 6
確　率	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

» 連続型の確率変数の確率分布

連続型確率変数 厳密に測定すると実現値が無限に存在するため、それぞれの実現値に番号が振れない確率変数のこと.

連続型確率変数の確率分布 確率変数 X のとる値が関数 $f(x)$ によって,

$$P(a \leq X \leq b) = \int_a^b f(x)dx \qquad (3)$$

と表される場合, X は連続型の確率分布を持つという (図 5.4). ただし, 全ての x に対して, $f(x) \geq 0$, かつ $\int_{-\infty}^{\infty} f(x)dx = 1$ を満たす. 関数 $f(x)$ をXの確率密度関数(probability density function) あるいは単に密度関数という.

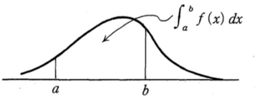

図 5.4 密度関数の図
密度関数は確率論の重要な概念だが，目でもわかりやすい．曲線下の面積がその区間の確率となる．$f(x)$ が 0 である区間の確率は 0 である．

- 連続型の確率分布においては, 一点の確率は 0 となる. $P(X = a) = 0$

» 累積分布関数

ある値以下の確率が必要なことがある．たとえば，飛行機の安全性を考えるとき，現実の積載重量 X が最大安全値を下回る確率は，この飛行機の安全確率となる．

累積分布関数 確率 X に対して，x を実数とするとき x 以下の確率

$$F(x) = P(X \leq x) \tag{4}$$

を X の累積分布関数 (cumulative distribution function) と呼ぶ．

連続型の場合 連続型の場合は $F(x)$ は f の定積分となる．

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(u)\,du \tag{5}$$

離散型の場合 離散型の場合は確率を $-\infty$ からは x まで積み上げたものになる：

$$F(x) = P(X \leq x) = \sum_{u \leq x} f(u) \tag{6}$$

となる．和は x 以下の u にわたる整数個の和である (x 自体は整数とは限らない)．

» 累積分布関数の性質

* 単調増加関数：

$$x_1 < x_2 \quad \text{ならば} \quad F(x_1) \leq F(x_2) \tag{7}$$

* 範囲：

$$F(-\infty) = 0, \qquad F(\infty) = 1 \tag{8}$$

* 右連続：

$$\text{各点 } x \text{ で } \epsilon \downarrow 0 \text{ の時} \quad F(x+\epsilon) \to F(x) \tag{9}$$

表 5.4

x	$f(x)$	$F(x)$
2	1/36	1/36=0.028
3	2/36	3/36=0.083
4	3/36	6/36=0.167
5	4/36	10/36=0.278
6	5/36	15/36=0.417
7	6/36	21/36=0.583
8	5/36	26/36=0.722
9	4/36	30/36=0.833
10	3/36	33/36=0.917
11	2/36	35/36=0.972
12	1/36	1

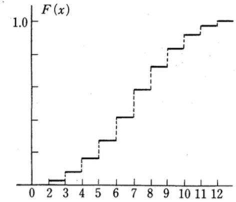

図 5.9

2 個のさいころの目の和の累積分布関数 $F(x)$ とその階段状グラフ (表 5.5, 図 5.9)
2, 3, …, 12 の各整数点で不連続(ただし右からは連続), それ以外では連続である. 増加は不連続点でのジャンプに限る. 典型的な離散型の確率分布である. 階段の蹴上げの高さが $f(x)$ であるが, 表の $F(x)$ は整数値に対してのみ表示した.

目次

* 確率変数と確率分布
* 確率変数の期待値と分散
* おもな離散型の確率分布

» 期待値

確率変数はいろいろの値をとるが，それらの値を代表する平均(正確には，確率の重みつき平均) が考えられる．これが期待値である．

期待値 確率変数 X に対して，それがとる値の重みつき平均を，確率変数の期待値といい，$E(X)$ と書く：

$$E(X) = \begin{cases} \sum_{x} xf(x) & \text{（離散型）} \\ \int_{-\infty}^{\infty} xf(x)\,dx & \text{（連続型）} \end{cases} \tag{10}$$

x の関数 $g(X)$ についても同様に期待値を定義する．

$$E(g(X)) = \begin{cases} \sum_x g(x)f(x) & \text{（離散）} \\ \int_{-\infty}^{\infty} g(x)f(x)\,dx & \text{（連続）} \end{cases}$$

» 期待値 例

例 1 確率変数：

$$X := \begin{cases} 1 & \text{with probability } p \\ 0 & \text{with probability } 1-p \end{cases}$$

の期待値：

$$E(X) = \sum_x x f(x) = 1 \cdot p + 0 \cdot (1-p) = p$$

例 2 指数分布：

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 (\text{ただし、}\lambda > 0) \\ 0 & x < 0 \end{cases}$$

の期待値：

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} x f(x)\, dx \\ &= \int_0^{\infty} x \lambda e^{-\lambda x}\, dx = \frac{1}{\lambda} \end{aligned}$$

» 期待値の演算 E の性質

期待値の演算 E の性質

$$(a) E(c) = c \qquad (\text{c は定数}) \qquad (11a)$$

$$(b) E(X + c) = E(X) + c \qquad (11b)$$

$$(c) E(cX) = cE(X) \qquad (11c)$$

$$(d) E(X + Y) = E(X) + E(Y) \qquad (\text{期待値の加法性}) \qquad (11d)$$

性質11dについて、2 個のさいころの目の和の確率分布は表 5. 3 でもとめてあるので, $E(X_1 + X_2) = 2 \cdot (1/36) + 3 \cdot (1/36) + \ldots = 252/36 = 7$, 他方, $E(X_1) = 7/2$, $E(X_2) = 7/2$ より, たしかに11dが成立している.

» 分散

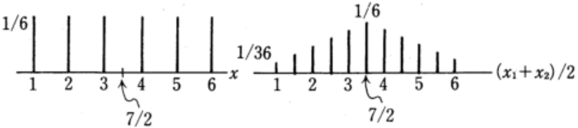

図 5.11　さいころの確率分布の期待値への集中の度合

今後簡単のために, 期待値を $\mu = E(X)$ で表そう. この図のように、期待値が同じでも, 確率分布は異なる。ここで「ばらつき」即ち、期待値 $E(X)$ からのずれの量 $X - E(X)$ 確率変数のもう一つの重要な指標である。

分散

$$V(X) = E\{(X-\mu)^2\} \tag{12}$$

と定義し、これを X の分散という. 分散の平方根を標準偏差という。上の定義から, 必ず, $V(X) \geq 0$ であり, $V(X)$ の値が大きいほど X のばらつきは大きい.

$$V(X) = \begin{cases} \sum_x (x-\mu)^2 f(x) & \text{（離散型）} \\ \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\, dx & \text{（連続型）} \end{cases} \tag{13}$$

» 分散

分散の演算 V の性質

分散の演算 V の性質

$$(a)\, V(c) = 0 \qquad \text{（定数の変動はゼロ）} \tag{14a}$$

$$(b)\, V(X + c) = V(X) \tag{14b}$$

$$(c)\, V(cX) = c^2 V(X) \tag{14c}$$

一般に、式12を直接使うよりは, 下記の等式を用いてもいい。

$$V(X) = E(X^2) - (E(X))^2 \tag{15}$$

$$\begin{aligned} V(X) &= E\{(X-\mu)^2\} \\ &= E(X^2) - 2\mu E(X) + \mu^2 \\ &= E(X^2) - \mu^2 \\ &= E(X^2) - (E(X))^2 \end{aligned}$$

» 標準化

確率変数の標準化 X から期待値 $E(X)$ を引き、標準偏差 $\sqrt{V(X)}$ で割ることで作られる新たな確率変数

$$Z = \frac{X - E(X)}{\sqrt{V(X)}} \tag{16}$$

を標準化された確率変数と呼ぶ.

標準化された確率変数 Z の性質

$$(a) E(Z) = 0 \tag{17a}$$
$$(b) V(Z) = 1 \tag{17b}$$

いかなる確率変数も, その期待値を引き, さらにその標準偏差 (分散の平方根) の尺度で割れば, 期待値は 0 に, 分散は 1 に調整される. この変換を確率変数の標準化といい, Z を標準化変数という.

目次

» 二項分布

ベルヌーイ試行 結果が 2 通り(成功 p ／失敗 $q(= 1 - p)$)しかない試行をベルヌーイ試行という。

二項分布 同じ条件且つ独立にベルヌーイ試行を n 回行って x 回成功する確率は

$$f(x) = {}_nC_x p^x(1 - p)^{n-x}, \quad x = 0, 1, 2, ..., n \tag{18}$$

で与えられる。ただし、${}_nC_x = \frac{n!}{x!(n-x)!}$ である。確率関数が上の $f(x)$ で与えられる確率分布を2 項分布 (binomial distribution) といい, X が 2 項分布に従う事を $X \sim Bi(n, p)$ と表す。

- 二項分布の期待値と分散は:

$$E(X) = np, \quad V(X) = np(1 - p) \tag{19}$$

» 二項分布 例

ある池の魚には，それを獲るとき確率 0.2 でその尾部に赤い色の標識が付いている．いま魚を 5 匹獲ったとき，その中で印の付いた魚がそれぞれ $x = 0, 1, 2, 3, 4, 5$ 回出る確率は？
これは試行回数 $n = 5$、成功確率 $p = 0.2$ の二項分布 $X \sim Bi(5, 0.2)$：

$$P(X = 0) = f(0) = \frac{5!}{0!(5-0)!}(0.2)^0(0.8)^5 = 0.328$$

$$P(X = 1) = f(1) = \frac{5!}{1!(5-1)!}(0.2)^1(0.8)^4 = 0.410$$

$$P(X = 2) = 0.205, P(X = 3) = 0.051, P(X = 4) = 0.006, P(X = 5) = 0.0003$$

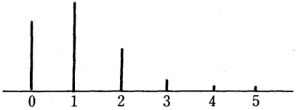

図 6.3 魚の匹数の分布
$n=5$, $p=1/5$ の二項分布であるが，5 匹中 1 匹の確率が最大という点に意味がある．

$E(X) = np$ であることが確認できる

» ポアソン分布

ポアソン分布　二項分布 $Bi(n, p)$ について、n が十分大きく、かつ p が非常に小さい場合、成功回数 X の分布は

$$f(x) = \frac{e^{-\lambda}\lambda^x}{x!}, \quad x = 0, 1, 2, \dots \quad (\lambda > 0) \tag{20}$$

に収束。これをポアソン分布 (Poisson distribution)と呼び、
$X \sim Po(\lambda)$ と表す。λ（ラムダ）はポアソン分布固有のパラメータ。
$e = 2.718\dots$ は自然対数の底、定数。

- ポアソン分布の期待値と分散は：

$$E(X) = \lambda, \quad V(X) = \lambda \tag{21}$$

常に「期待値 = 分散 = λ」という、珍しい性質。

» ポアソン分布 例

ゴール数の例：サッカー J リーグ一部 2012 年第 1 節 ~ 第 3 節、のべ 54 チームのゴール数のデータを入手したとする。データ（サンプル数 54）の平均 $\bar{X} = 1.116$、分散 $s^2 = 1.001$、標準偏差 $s = 1.001$。
ポアソン分布の未知パラメータ λ を $\lambda = 1.116$ と置き、ゴール数実現値 $x = 0, 1, 2, \ldots$ の確率を計算する。この計算した結果とサンプルデータの相対度数と比較しみれば：

ポアソン確率（%）	32.76	36.56	20.40	7.59	2.12	0.47
データ相対度数（%）	29.63	35.19	25.93	7.41	1.85	0.00
x（ゴール数）	0	1	2	3	4	5

ポアソン分布での結果は実際のサンプルデータの分布とよく似ていることが分かる。

» 離散型一様分布

離散型一様分布 さいころを振ったときに出る目 X の確率分布が一様分布の例で，

$$f(x) = 1/N, \quad x = 1, 2, \ldots N \quad (N \text{ は正整数}) \tag{22}$$

を，$1, 2, \ldots, N$ 上の離散型一様分布 (uniform distribution of discrete type)という。
一例として，さいころでは $N = 6$ である．この期待値は $E(X) = \frac{N+1}{2}$，分散は $V(X) = \frac{N^2-1}{12}$。