確率変数とその分布

目次

本スライドの内容

このスライドは，次の書籍の第 2 章「確率変数とその分布」の内容に基づく．

- 『ガイダンス 確率統計：基礎から学び本質の理解へ』，発行：サイエンス社，ISBN：978-4-7819-1526-5.

書籍に関する最新の情報は，以下の URL から入手することができます．

`https://www.saiensu.co.jp`

この URL は，サイエンス社が運営しているホームページです．

概要

このスライドでは，確率変数と分布の概念，および分布を特徴付ける分布関数や，平均，分散について解説し，確率変数の標準化の考え方を紹介する．このスライドでは，(Ω, P) は確率空間を表すものとする．

定義 2.1.1 (確率変数)

定義 2.1.1 (確率変数)

標本空間 Ω 上で定義された実数値関数

$$X : \Omega \ni \omega \mapsto X(\omega) \in \mathbb{R} \quad (\text{実数全体の集合})$$

を Ω 上の確率変数とよぶ．また，Ω 上の確率 P を考えるとき，Ω 上の確率変数 X のことを，確率空間 (Ω, P) 上の確率変数とよぶこともある．なお，Ω 上の確率変数 X を，$\{X(\omega)\}_{\omega\in\Omega}$ と表すこともある．

記号 2.1.1

X は Ω 上の確率変数とする．

- 1 変数関数 $\varphi(x)$ に対して，Ω 上の確率変数 $\varphi(X)$ は

$$\varphi(X)(\omega) = \varphi(X(\omega)) \qquad (\omega \in \Omega)$$

と定義される．

E を実数 $\mathbb{R}$ の部分集合とする．

- $X(\omega)$ が E に属する根元事象 ω の集合

$$\{\omega \in \Omega \mid X(\omega) \in E\}$$

を，$\{X \in E\}$ と略記することもある．さらに，確率

$$P(\{\omega \in \Omega \mid X(\omega) \in E\})$$

は $P(X \in E)$ と略記することもある．他にも，たとえば $P(\{\omega \in \Omega \mid a \leq X(\omega) \leq b\})$ は $P(a \leq X \leq b)$ と略記することもある．

記号 2.1.1 (続)

次に，X に加えて Y も Ω 上の確率変数とする．このとき，2 変数関数 $v(x, y)$ に対して，Ω 上の確率変数 $v(X, Y)$ は

$$v(X, Y)(\omega) = v(X(\omega), Y(\omega)) \qquad (\omega \in \Omega)$$

と定義される．また，E に加えて F も実数 $\mathbb{R}$ の部分集合とする．このとき，$X(\omega)$ が E に属し，同時に $Y(\omega)$ が F に属する根元事象 ω の集合

$$\{\omega \in \Omega \mid X(\omega) \in E \text{ かつ } Y(\omega) \in F\}$$

を，$\{X \in E,\ Y \in F\}$ や $\{X \in E\} \cap \{Y \in F\}$ と略記したり，ベクトル表記を用いて $\{(X, Y) \in E \times F\}$ と略記することもある．また，「任意の $\omega \in \Omega$ で成立する確率変数に関する関係式」については，変数部分の (ω) を略して表記することが多い．たとえば，$X(\omega)Y(\omega)^2 > 3Y(\omega) + 2$ $(\omega \in \Omega)$ という関係式を略記する場合は，$XY^2 > 3Y + 2$ と表記する．

定義 2.1.2 (定義関数) と問 2.1.2

次の「定義関数」は，確率統計の様々な場面で活用する．

定義 2.1.2 (定義関数)

集合 A 上で 1，A の外で 0 である関数

$$1_A(x) = \begin{cases} 1, & x \in A \\ 0, & x \notin A \end{cases}$$

を A の定義関数とよぶ．

問 2.1.2　$\Omega = [0,1]$ とし，事象 $A = [0,1/2]$ と $B = [1/4,3/4]$ を考える．横軸に $\omega \in [0,1]$ を取り，縦軸に y を取り，次の関数の概形を描け．

$$(1) \quad y = 1_A(\omega) + 1_B(\omega) \quad (\omega \in [0,1]),$$

$$(2) \quad y = 3 \cdot 1_A(\omega) - 1_B(\omega) \quad (\omega \in [0,1]).$$

定義 2.1.3 (離散分布，確率関数)

離散型確率変数を定義するためには，次の「離散分布と確率関数の概念」が必要となる．

定義 2.1.3 (離散分布，確率関数)

実数からなる集合 $\{x_1, x_2, \cdots\}$ ($i \neq j$ なら $x_i \neq x_j$) の上で定義された関数 $p(x)$ に対して $p_k = p(x_k)$ ($k \geq 1$) とおく．$p_k \geq 0$ ($k \geq 1$) であり，かつ $\sum_{k\geq 1} p_k = 1$ をみたすとき，

$$\begin{pmatrix} x_k \\ p_k \end{pmatrix}_{k\geq 1} = \begin{pmatrix} x_1 & x_2 & \cdots \\ p_1 & p_2 & \cdots \end{pmatrix} \tag{2.3}$$

を**離散分布**または単に**分布**とよび，関数 $p(x)$ を**確率関数**とよぶ．

注意 2.1.1

定義 2.1.3 において，$p_{n+1} = p_{n+2} = \cdots = 0$ (n は自然数) であるとき，

$$\begin{pmatrix} x_1 & x_2 & \cdots & x_n & x_{n+1} & x_{n+2} & \cdots \\ p_1 & p_2 & \cdots & p_n & 0 & 0 & \cdots \end{pmatrix}$$

で与えられる離散分布 (2.3) は，確率が 0 の部分は省略して，記号

$$\begin{pmatrix} x_k \\ p_k \end{pmatrix}_{1 \le k \le n} = \begin{pmatrix} x_1 & x_2 & \cdots & x_n \\ p_1 & p_2 & \cdots & p_n \end{pmatrix}$$

で表すことが多い．

定義 2.1.4 (離散型確率変数)

予め決められた有限個 (または可算無限個) の値のみを取り得る確率変数は離散型確率変数とよばれる.

定義 2.1.4 (離散型確率変数)

(Ω, P) 上の確率変数 X に対して，(2.3) で与えられる離散分布が存在して，$P(X = x_k) = p_k$ $(k \geq 1)$ が成り立つとき，X は離散分布 (2.3) に従うといい，このことを記号

$$X \sim \begin{pmatrix} x_1 & x_2 & \cdots \\ p_1 & p_2 & \cdots \end{pmatrix} \qquad (2.5)$$

で表す．この X のように，離散分布に従う確率変数を離散型確率変数とよぶ.

例題 2.1.1

例題 2.1.1

男の子 5 人, 女の子 3 人の中から無作為に 3 人を選び, この 3 人の中に含まれる男の子の人数を X とする．このとき，X が従う分布を求め，事象 $A=\{\omega\in\Omega \mid 1\le X(\omega)\le 2\}$ の確率 $P(A)$ を求めよ．

[解答]　8 人から 3 人を選ぶ選び方は ${}_8\mathrm{C}_3=56$ 通り．男の子 k 人，女の子 $3-k$ 人を選ぶ選び方は ${}_5\mathrm{C}_k\cdot{}_3\mathrm{C}_{3-k}$ であるから，X は離散分布

$$\begin{pmatrix} k \\ {}_5\mathrm{C}_k\cdot{}_3\mathrm{C}_{3-k}/{}_8\mathrm{C}_3 \end{pmatrix}_{k=0}^{3}=\begin{pmatrix} 0 & 1 & 2 & 3 \\ 1/56 & 15/56 & 30/56 & 10/56 \end{pmatrix}$$

に従う．よって，

$P(A)=P(X=1)+P(X=2)=(15+30)/56=45/56.$

例 2.1.4 (ベルヌーイ分布)

$0 < p < 1$, $q = 1 - p$ に対して，離散分布

$$\begin{pmatrix} 0 & 1 \\ q & p \end{pmatrix}$$

をパラメータ p のベルヌーイ分布とよび，記号 $Be(p)$ で表す．
成功か失敗か 2 つの結果しかない試行において，成功したら $X = 1$ と定め，失敗したら $X = 0$ と定める．このとき，X は確率変数であり，成功確率を p とおくと X は $Be(p)$ に従う．
このような試行と確率変数 X を，それぞれベルヌーイ試行とベルヌーイ確率変数とよぶ．

例 2.1.5 (二項分布)

自然数 n, $0 < p < 1$, $q = 1 - p$ に対して，離散分布

$$\begin{pmatrix} k \\ {}_nC_k p^k q^{n-k} \end{pmatrix}_{0 \le k \le n} = \begin{pmatrix} 0 & 1 & \cdots & n \\ q^n & npq^{n-1} & \cdots & p^n \end{pmatrix} \tag{2.9}$$

をパラメータ (n, p) の二項分布とよび，記号 $B(n, p)$ で表す．
二項定理より，関係式

$$\sum_{k=0}^{n} {}_nC_k p^k q^{n-k} = (p+q)^n = 1$$

が成り立つため，(2.9) は離散分布の条件をみたす．$n = 1$ のとき，二項分布 $B(1, p)$ とベルヌーイ分布 $Be(p)$ は一致する．
なお，「事象 A が起こる確率が p である試行」を独立に n 回繰り返すとき，この n 回の試行のうち A が起こる回数を X とおくと，X は二項分布 $B(n, p)$ に従うことが知られている．このことは系 3.1.1 で詳しく解説する．

例 2.1.6 (幾何分布)

$0 < p < 1$, $q = 1 - p$ に対して，離散分布

$$\begin{pmatrix} k \\ pq^k \end{pmatrix}_{k \geq 0} = \begin{pmatrix} 0 & 1 & 2 & \cdots \\ p & pq & pq^2 & \cdots \end{pmatrix} \tag{2.10}$$

をパラメータ p の幾何分布とよび，記号 $Ge(p)$ で表す．なお，$\sum_{k=0}^{\infty} q^k = 1/(1-q) = 1/p$ より，(2.10) は離散分布の条件をみたす．「1 回の試行で成功する確率が p である試行」を独立に何回も繰り返すとき，はじめて成功するまでに失敗した試行の回数を X とおくと，$P(X = k) = pq^k$ $(k = 0, 1, \cdots)$ が成り立つため，X は幾何分布 $Ge(p)$ に従う．

例 2.1.7 (ポアソン分布)

$\lambda > 0$ に対して，離散分布

$$\begin{pmatrix} k \\ e^{-\lambda}\dfrac{\lambda^k}{k!} \end{pmatrix}_{k\geq 0} = \begin{pmatrix} 0 & 1 & 2 & \cdots \\ e^{-\lambda} & e^{-\lambda}\lambda & e^{-\lambda}\frac{\lambda^2}{2} & \cdots \end{pmatrix} \tag{2.13}$$

をポアソン分布とよび，記号 $Po(\lambda)$ で表す．ここで，e は自然対数の底であり，その値は $e = 2.718\cdots$ である．

$$e^{\lambda} = \sum_{k=0}^{\infty} \lambda^k / k!$$

であるため，(2.13) は離散分布の条件をみたす．1 年間当たりの交通事故の件数のように，「まれにしか起こらない事象が所定の時間内に発生する件数」を表す確率変数の分布は，ポアソン分布によく当てはまることが知られている．

- 定義 2.1.4 の離散型確率変数と対になる概念として連続型確率変数がある.
- 連続型確率変数とは，生徒の身長や体重のように，取り得る値を無限に細かくできる確率変数のことを指す.
- 連続型確率変数 X においては，X が一定の区間内に入る確率を計算することが重要である．なぜなら，X の取り得る値のすべてに正の確率を与えてしまうと，確率の総和が無限大となってしまい，確率の公理と矛盾が生じるからである.
- 連続型確率変数が一定の区間内に入る確率を数学的に表現するためには,「密度関数の概念」が必要となる.

定義 2.1.5 (密度関数)

定義 2.1.5 (密度関数)

$\mathbb{R}$ 上の実数値関数 $f(x)$ が 2 条件

$$f(x) \geq 0, \quad \int_{-\infty}^{\infty} f(x)dx = 1$$

をみたすとき，$f(x)$ を密度関数とよぶ．$f(x)$ が密度関数のとき，任意の区間 $[a, b]$ に対して

$$\mu_f([a, b]) = \int_a^b f(x)dx$$

と定義し，この μ_f を「$f(x)$ から定まる分布」とよぶ．このとき，$f(x)$ を「μ_f の密度関数」ともよぶ．

定義 2.1.6 (連続型確率変数)

連続型確率変数が区間 $[a, b]$ 内の値を取る確率は，密度関数の区間 $[a, b]$ 上の積分で定義する．

定義 2.1.6 (連続型確率変数)

(Ω, P) 上の確率変数 X に対して，ある密度関数 $f(x)$ が存在して，任意の $a < b$ に対して

$$P(a \leq X \leq b) = \mu_f([a, b])$$

が成り立つとき，X は分布 μ_f に従うといい，このことを $X \overset{\mathrm{pdf}}{\sim} f(x)$ や $X \sim \mu_f$ と表す．$X \sim \mu_f$ のとき，$f(x)$ を「X の密度関数」とよび，X を連続型確率変数とよぶ．

注意 2.1.5

定義 2.1.6 の設定のもとで考察する．このとき，任意の実数 c に対して

$$P(X = c) = P(c \leq X \leq c) = \int_c^c f(x)dx = 0$$

が成り立つため，X が一点に値を取る確率は 0 である．したがって，$a < b$ をみたす任意の実数 a, b に対して，次の関係式が成り立つ．

$$\begin{aligned}\int_a^b f(x)dx &= P(a \leq X \leq b) \\ &= P(a < X \leq b) = P(a \leq X < b) = P(a < X < b).\end{aligned}$$

例 2.1.8 (一様分布)

$a < b$ に対して，密度関数

$$f(x) = \frac{1}{b-a} 1_{[a,b]}(x) = \begin{cases} \dfrac{1}{b-a} & (a \leq x \leq b) \\ 0 & (x < a \text{ または } x > b) \end{cases}$$

から定まる分布 μ_f を $[a, b]$ 上の一様分布とよび，記号 $U(a, b)$ で表す．$c > 0$ は定数とし，A 駅では c 分おきに電車が発車しているとする．B さんが無作為に A 駅に到着するとき，到着後に次の電車が来るまでの B さんの待ち時間を表す確率変数は $U(0, c)$ に従う．

例 2.1.9 (指数分布)

$\lambda > 0$ に対して，密度関数

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

から定まる分布 μ_f をパラメータ λ の指数分布とよび，記号 $\mathrm{Exp}(\lambda)$ で表す．$f(x)$ が密度関数の条件をみたすことは，

$$\int_{-\infty}^{\infty} f(x)dx = \int_0^{\infty} \lambda e^{-\lambda x} dx = -\int_0^{\infty} \frac{d}{dx} e^{-\lambda x} dx = 1$$

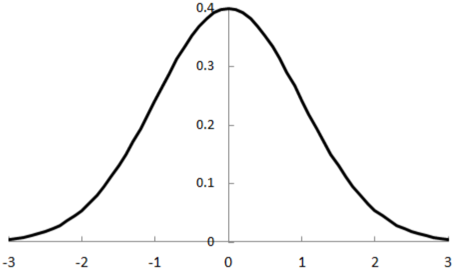
からわかる．指数分布は，地震が起きる間隔や，製品が製造されたときから壊れるまでの時間などを表す確率変数が従う分布としてよく利用される．数学的には，指数分布は「単位時間あたりに起こる確率が常に一定である無作為なイベントの発生間隔」を表す確率変数が従う分布である．

例 2.1.10 (正規分布)

実数 μ および $\sigma > 0$ に対して，密度関数

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} \quad (x \in \mathbb{R})$$

から定まる分布 μ_f を平均 μ, 分散 σ^2 の正規分布とよび，記号 $N(\mu, \sigma^2)$ で表す．特に $N(0, 1)$ は標準正規分布とよばれる．



補題 2.1.2 (正規分布と線形変換)

次の補題 2.1.2 から，正規分布に従う確率変数 X の線形変換 $cX + d$ も正規分布に従うことがわかる．

補題 2.1.2 (正規分布と線形変換)

(Ω, P) 上の確率変数 X は正規分布 $N(\mu, \sigma^2)$ に従うとする．このとき，実数 $c \neq 0$ と実数 d に対して，(Ω, P) 上の確率変数 $Y = cX + d$ は正規分布 $N(c\mu + d, c^2\sigma^2)$ に従う．

注意 2.1.6

(Ω, P) 上の確率変数 X が $N(\mu, \sigma^2)$ に従うとする．このとき，補題 2.1.2 より，確率変数 $Z = (X - \mu)/\sigma$ は標準正規分布 $N(0, 1)$ に従うことがわかり，この Z を X の標準化とよぶ．なお，$N(0, 1)$ に従う確率変数 Z に対する確率

$$p(u) = P(0 \leq Z \leq u) = \frac{1}{\sqrt{2\pi}} \int_0^u e^{-\frac{z^2}{2}} dz \qquad (2.18)$$

の値は表 C.1 を利用して計算できる．

1. $p(0.12)$ の値を調べるとき，まず $0.12 = 0.1 + 0.02$ と分解する．次に表 C.1 の「縦の 0.1 に対応する行」(上から 2 行目) の「横の .02 に対応する列」(左から 3 列目) の数値を探し，この数値が $p(0.12) = 0.0478$ である．

例題 2.1.3

例題 2.1.3

ある入学試験の点数 X は正規分布 $N(51, 625)$ に従うとする．

1. 52 点以上かつ 54 点以下の割合を求めよ．
2. 49 点以上かつ 52 点以下の割合を求めよ．
3. 61 点以上を合格としたときの合格率を求めよ．

ただし，表 C.1 を利用すること．

[解答]　X の標準化を $Z = (X - 51)/\sqrt{625} = (X - 51)/25$ とおくと，Z は標準正規分布 $N(0, 1)$ に従う．$N(0, 1)$ は密度関数 $f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ から定まる分布である．$y = f(x)$ は y 軸に関して左右対称な形をしていることや，「確率の加法性」を用いて次の計算を行う．

例題 2.1.3

1. $P(52 \leq X \leq 54) = P(0.04 \leq Z \leq 0.12)$ であるため，
$$P(52 \leq X \leq 54) = P(0 \leq Z \leq 0.12) - P(0 \leq Z \leq 0.04)$$
$$= 0.0478 - 0.0160 = 0.0318$$
と計算でき，求める割合は 3.18%である．
2. $P(49 \leq X \leq 52) = P(-0.08 \leq Z \leq 0.04)$ であるため，
$$P(49 \leq X \leq 52) = P(-0.08 \leq Z \leq 0) + P(0 \leq Z \leq 0.04)$$
$$= P(0 \leq Z \leq 0.08) + P(0 \leq Z \leq 0.04)$$
$$= 0.0319 + 0.0160 = 0.0479$$
と計算でき，求める割合は 4.79%である．
3. $P(X \geq 61) = P(Z \geq 0.40) = P(Z \geq 0) - P(0 \leq Z \leq 0.40) = 0.5 - 0.1554$ と計算できるため，求める合格率は 34.46%である．

定義 2.2.1 (分布関数)

確率変数 X が x 以下である確率 $P(X \leq x)$ を，x の関数とみなすとき，この関数は X の分布関数とよばれ，次のように定義される．

定義 2.2.1 (分布関数)

(Ω, P) 上の確率変数 X と実数 x に対して，事象

$$\{X \leq x\} = \{\omega \in \Omega \mid X(\omega) \leq x\}$$

の確率を

$$F(x) = P(X \leq x) = P(\{\omega \in \Omega \mid X(\omega) \leq x\})$$

と定め，x の単調増加関数 $F(x)$ を，X の分布関数とよぶ．
なお，$F(x)$ は $F_X(x)$ とも表す．

例 2.2.1

まず，X が離散分布 (2.3) に従うとき，次式が成り立つ．

$$F(x) = P(X \leq x) = \sum_{\substack{k \geq 1 \\ x_k \leq x}} p(x_k) = \sum_{\substack{k \geq 1 \\ x_k \leq x}} p_k.$$

次に，X の密度関数が $f(x)$ であるとき，

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt$$

が成り立つ．よって，この $f(x)$ が x で連続であれば，微分積分学の基本定理より，$\frac{d}{dx}F(x) = f(x)$ が成り立つ．
なお，分布関数 $F(x)$ がわかれば，確率関数 $p(x)$ や密度関数 $f(x)$ を特定できることが知られている．

確率関数・密度関数・分布関数

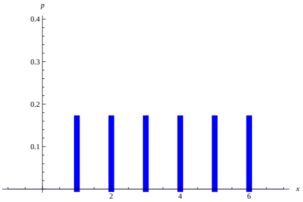


Figure: 離散型確率変数の確率関数と分布関数.

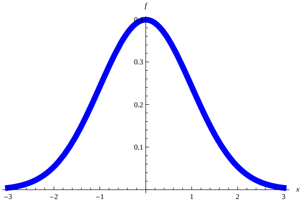


Figure: 連続型確率変数の密度関数と分布関数.

例題 2.2.1

例題 2.2.1

白玉が 7 個，黒玉が 3 個の計 10 個が入った袋がある．この袋の中から玉を 1 個取り出し，その玉を袋に戻さずに，さらに袋の中から玉を 1 個取り出すとき，取り出した 2 個の玉のうち白玉の個数を X とおく．このとき，X の分布関数 $F(x)$ を求めよ．

[解答]　まず $x < 0$ のとき，関係式 $\{X \le x\} = \emptyset$ より，$F(x) = P(\emptyset) = 0$ である．次に $0 \le x < 1$ のとき，関係式

$$\{X \le x\} = \{X = 0\}$$

より，$F(x) = P(X = 0) = (3 \cdot 2)/(10 \cdot 9) = 1/15$ である．

例題 2.2.1

[解答 (続き)]　また $1 \leq x < 2$ のとき，関係式

$$\{X \leq x\} = \{X = 0\} \cup \{X = 1\}$$

より，

$$F(x) = P(X = 0) + P(X = 1) = \frac{3 \cdot 2}{10 \cdot 9} + 2\frac{7 \cdot 3}{10 \cdot 9} = \frac{8}{15}$$

である．最後に $x \geq 2$ のとき，関係式

$$\{X \leq x\} = \{X = 0\} \cup \{X = 1\} \cup \{X = 2\}$$

より，$F(x) = P(X = 0) + P(X = 1) + P(X = 2) = 1$ である．

定義 2.3.1 (期待値)

(Ω, P) 上の確率変数 X と関数 $h(x)$ に対し，確率変数 $h(X) = \{h(X(\omega))\}_{\omega\in\Omega}$ の**期待値** $E(h(X))$ を定義する．

定義 2.3.1 (期待値)

まず，X が離散分布

$$\begin{pmatrix} x_1 & x_2 & \cdots \\ p_1 & p_2 & \cdots \end{pmatrix} \tag{2.22}$$

に従うとき，$E(h(X))$ を次式で定義する．

$$E(h(X)) = \sum_{k=1}^{\infty} h(x_k)p_k = \sum_{k=1}^{\infty} h(x_k)P(X = x_k). \tag{2.23}$$

定義 2.3.1 (期待値)

(Ω, P) 上の確率変数 X と関数 $h(x)$ に対し，確率変数 $h(X) = \{h(X(\omega))\}_{\omega\in\Omega}$ の**期待値** $E(h(X))$ を定義する.

定義 2.3.1 (期待値)

次に，X の密度関数が $f(x)$ であるとき，$E(h(X))$ を次式で定義する.

$$E(h(X)) = \int_{-\infty}^{\infty} h(x)f(x)dx. \qquad (2.24)$$

例 2.3.1 (平均・分散・標準偏差)

X は (Ω, P) 上の確率変数とする．このとき，期待値 $E(X)$ は X の平均ともよばれる．また，

$$V(X) = E((X - E(X))^2)$$

は X の分散とよばれ，$\sigma(X) = \sqrt{V(X)}$ は X の標準偏差とよばれる．$V(X)$ と $\sigma(X)$ はともに，「平均 $E(X)$ のまわりの X の散らばり度合い」を数値で表したものである．X が平均 $E(X)$ に近い値を取る確率が大きいとき，$V(X)$ は小さい．逆に，X が平均 $E(X)$ より離れた値を取る確率が大きいとき，$V(X)$ は大きい．$V(X)$ の単位は「X の単位の 2 乗」であり，$\sigma(X)$ の単位は X の単位と同じである．

定理 2.3.1 (期待値の線形性)

期待値の定義 (定義 2.3.1) より，次の「期待値の線形性」を証明することができる．

定理 2.3.1 (期待値の線形性)

X は (Ω, P) 上の確率変数とする．このとき，関数 $g(x), h(x)$ と定数 a, b, c に対して

$$E(ag(X) + bh(X) + c) = aE(g(X)) + bE(h(X)) + c$$

が成り立つ．このことから，特に次式も成り立つ．

$$E(aX^2 + bX + c) = aE(X^2) + bE(X) + c. \tag{2.31}$$

系 2.3.1 (期待値の基本不等式)

系 2.3.1 (期待値の基本不等式)

X は (Ω, P) 上の確率変数とする．このとき，次の $(1), (2), (3)$ が成り立つ．

$$
\begin{aligned}
&(1) \quad |E(X)| \leq E(|X|), \\
&(2) \quad P(|X| > \varepsilon) \leq \frac{1}{\varepsilon^2} E(X^2) \quad (\varepsilon > 0), \\
&(3) \quad X(\omega) \geq 0 \ (\omega \in \Omega) \text{ であれば } E(X) \geq 0.
\end{aligned}
$$

(2) において，X を $X - \mu$ $(\mu = E(X))$ に置き換えると，

$$P(|X - \mu| > \varepsilon) \leq \frac{1}{\varepsilon^2} E((X - \mu)^2) = \frac{1}{\varepsilon^2} V(X) \quad (\varepsilon > 0) \quad (2.35)$$

が得られ，チェビシェフの不等式とよばれる．

定理 2.3.3 (分散の基本公式)

定理 2.3.1 を用いると，次の「分散の基本公式」を証明することができる．

定理 2.3.3 (分散の基本公式)

X は (Ω, P) 上の確率変数とし，a, b は定数とする．このとき，次が成り立つ．

$$V(aX+b) = a^2 V(X), \quad V(X) = E(X^2) - (E(X))^2.$$

例 2.3.3

(Ω, P) 上の確率変数 X がベルヌーイ分布 $Be(p)$ に従うとき，次が成り立つ.

$$E(X) = 1 \cdot p + 0 \cdot (1-p) = p,$$
$$E(X^2) = 1^2 \cdot p + 0^2 \cdot (1-p) = p,$$
$$V(X) = E(X^2) - (E(X))^2 = p(1-p).$$

例 2.3.4

(Ω, P) 上の確率変数 X が二項分布 $B(n, p)$ に従うとし，$q = 1 - p$ とおく．このとき，二項定理を用いると，$E(X)$ は次のように計算できる．

$$E(X) = \sum_{k=0}^{n} k\,{}_nC_k p^k q^{n-k} = np \sum_{k=1}^{n} \frac{(n-1)!}{(k-1)!(n-k)!} p^{k-1} q^{n-k}$$

$$= np \sum_{l=0}^{n-1} {}_{n-1}C_l p^l q^{(n-1)-l} = np(p+q)^{n-1} = np \quad (l = k-1).$$

例 2.3.4

次に $E(X(X-1))$ についても，二項定理を用いることで，

$$
\begin{aligned}
E(X(X-1)) &= \sum_{k=0}^{n} k(k-1)\cdot\frac{n!}{k!(n-k)!}\cdot p^k q^{n-k} \\
&= n(n-1)p^2 \sum_{k=2}^{n} \frac{(n-2)!}{(k-2)!(n-k)!} p^{k-2} q^{n-k} \\
&= n(n-1)p^2 \sum_{l=0}^{n-2} {}_{n-2}\mathrm{C}_l\, p^l q^{(n-2)-l} \qquad (l = k-2) \\
&= n(n-1)p^2 (p+q)^{n-2} = n(n-1)p^2
\end{aligned}
$$

と計算できる．したがって，次が得られる．

$$
\begin{aligned}
E(X^2) &= E(X(X-1)) + E(X) = n(n-1)p^2 + np, \\
V(X) &= E(X^2) - (E(X))^2 = np(1-p).
\end{aligned}
$$

例 2.3.5

(Ω, P) 上の確率変数 X が幾何分布 $Ge(p)$ に従うとし，
$q = 1 - p$ とおく．このとき，式変形

$$
\begin{aligned}
E(X) &= \sum_{k=0}^{\infty} kpq^k = \sum_{k=0}^{\infty}(1 + k - 1)pq^k = \sum_{k=0}^{\infty} pq^k + \sum_{k=0}^{\infty}(k-1)pq^k \\
&= 1 - p + \sum_{k=1}^{\infty}(k-1)pq^k = 1 - p + q\sum_{k=1}^{\infty}(k-1)pq^{k-1} \\
&= 1 - p + q\sum_{k'=0}^{\infty} k'pq^{k'} = 1 - p + qE(X) \qquad (k' = k - 1)
\end{aligned}
$$

が成り立つ．よって，この式変形で得られた方程式

$$E(X) = 1 - p + qE(X)$$

を $E(X)$ について解くと，$E(X) = (1-p)/p$ が得られる．

例 2.3.5

次に，$(k-1)^2 = k^2 - 2k + 1$ より，次の式変形が成り立つ．

$$
\begin{aligned}
E(X^2) &= \sum_{k=0}^{\infty} k^2 pq^k = \sum_{k=0}^{\infty}((k-1)^2 + 2k - 1)pq^k \\
&= (-1)^2 p + \sum_{k=1}^{\infty}(k-1)^2 pq^k + 2\sum_{k=0}^{\infty} kpq^k - \sum_{k=0}^{\infty} pq^k \\
&= p + q\sum_{k=1}^{\infty}(k-1)^2 pq^{k-1} + 2E(X) - 1 \\
&= p + q\sum_{k'=0}^{\infty}(k')^2 pq^{k'} + \frac{2(1-p)}{p} - 1 \quad (k' = k-1) \\
&= qE(X^2) + \frac{(p-1)(p-2)}{p}.
\end{aligned}
$$

例 2.3.5

この式変形で得られた方程式

$$E(X^2) = qE(X^2) + (p-1)(p-2)/p$$

を $E(X^2)$ について解くと，次が得られる．

$$E(X^2) = \frac{(p-1)(p-2)}{p^2}, \quad V(X) = E(X^2) - (E(X))^2 = \frac{1-p}{p^2}.$$

以上の計算結果をまとめると，次のとおりである．

$$E(X) = \frac{1-p}{p}, \quad V(X) = \frac{1-p}{p^2}. \qquad (2.36)$$

例 2.3.6

(Ω, P) 上の確率変数 X がポアソン分布 $Po(\lambda)$ に従うとき，次が得られる．

$$E(X) = \sum_{k=0}^{\infty} k e^{-\lambda} \frac{\lambda^k}{k!} = \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = \lambda,$$

$$E(X(X-1)) = \sum_{k=0}^{\infty} k(k-1) e^{-\lambda} \frac{\lambda^k}{k!} = \lambda^2 e^{-\lambda} \sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{(k-2)!} = \lambda^2,$$

$$\begin{aligned} V(X) &= E(X^2) - (E(X))^2 \\ &= E(X(X-1)) + E(X) - (E(X))^2 = \lambda. \end{aligned}$$

例 2.3.7

(Ω, P) 上の確率変数 X が一様分布 $U(a, b)$ に従うとき，多項式の定積分を計算することで，次が得られる．

$$E(X) = \frac{1}{b-a}\int_a^b x dx = \frac{a+b}{2},$$

$$V(X) = \frac{1}{b-a}\int_a^b x^2 dx - (E(X))^2 = \frac{(a-b)^2}{12}.$$

例 2.3.8

(Ω, P) 上の確率変数 X は指数分布 $\mathrm{Exp}(\lambda)$ に従うとする．ロピタルの定理より，次が成り立つ．

$$\lim_{x\to\infty}\frac{x}{e^{\lambda x}} = \lim_{x\to\infty}\frac{1}{\lambda e^{\lambda x}} = 0,$$

$$\lim_{x\to\infty}\frac{x^2}{e^{\lambda x}} = \lim_{x\to\infty}\frac{2x}{\lambda e^{\lambda x}} = \lim_{x\to\infty}\frac{2}{\lambda^2 e^{\lambda x}} = 0.$$

このことと，部分積分公式より，次が得られる．

$$E(X) = \int_0^\infty x\lambda e^{-\lambda x}dx = \left[-xe^{-\lambda x}\right]_{x=0}^{x=\infty} + \int_0^\infty e^{-\lambda x}dx = \frac{1}{\lambda},$$

$$E(X^2) = \int_0^\infty x^2\lambda e^{-\lambda x}dx = \left[-x^2e^{-\lambda x}\right]_{x=0}^{x=\infty} + \int_0^\infty 2xe^{-\lambda x}dx = \frac{2}{\lambda^2},$$

$$V(X) = E(X^2) - (E(X))^2 = \frac{1}{\lambda^2}.$$

例 2.3.9

(Ω, P) 上の確率変数 Z が標準正規分布 $N(0,1)$ に従うとする. 0 以上の整数 k に対して $x^{2k+1}e^{-x^2/2}$ は x の奇関数であるため，次式が成り立つ.

$$E(Z^{2k+1}) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^{2k+1} e^{-\frac{x^2}{2}} dx = 0. \qquad (2.37)$$

また，ロピタルの定理より，次式が成り立つ.

$$\lim_{x\to\infty} \frac{x}{e^{\frac{x^2}{2}}} = \lim_{x\to\infty} \frac{1}{xe^{\frac{x^2}{2}}} = 0. \qquad (2.38)$$

例 2.3.9

したがって，(2.37), (2.38)，および部分積分公式より，

$$V(Z) = E(Z^2) = \int_{-\infty}^{\infty} x^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = \frac{2}{\sqrt{2\pi}} \int_0^{\infty} x^2 e^{-\frac{x^2}{2}} dx$$
$$= \frac{2}{\sqrt{2\pi}} \left\{ \int_0^{\infty} e^{-\frac{x^2}{2}} dx - \left[xe^{-\frac{x^2}{2}} \right]_{x=0}^{x=\infty} \right\} = \frac{2}{\sqrt{2\pi}} \left\{ \frac{\sqrt{2\pi}}{2} - 0 \right\} = 1$$

と計算できる．なお，X が正規分布 $N(\mu, \sigma^2)$ に従うとき，X の標準化 $Z = (X - \mu)/\sigma$ は $N(0, 1)$ に従う．したがって，次が得られる．

$$E(X) = E(\mu + \sigma Z) = \mu + \sigma E(Z) = \mu,$$
$$V(X) = V(\mu + \sigma Z) = \sigma^2 V(Z) = \sigma^2.$$

定義 2.3.2 (標準化)

確率変数 X の線形変換 $Z = cX + d$ であり，平均が $E(Z) = 0$ で，分散が $V(Z) = 1$ をみたす変換は，X の標準化とよばれ，次のように定義される．

定義 2.3.2 (標準化)

(Ω, P) 上の確率変数 X に対して，X の**標準化**を

$$Z(\omega) = \frac{X(\omega) - E(X)}{\sqrt{V(X)}} = \frac{X(\omega) - E(X)}{\sigma(X)} \qquad (\omega \in \Omega)$$

と定義する．

注意 2.3.5

(Ω, P) 上の確率変数 X の標準化を $Z = (X - E(X))/\sigma(X)$ とおく．このとき，次が成り立つ．

$$E(Z) = \frac{1}{\sigma(X)} E(X - E(X)) = \frac{1}{\sigma(X)} (E(X) - E(X)) = 0,$$

$$V(Z) = \left(\frac{1}{\sqrt{V(X)}} \right)^2 V(X - E(X)) = \frac{1}{V(X)} V(X) = 1.$$

したがって，Z は平均が 0 であり，分散が 1 である．

注意 2.3.5

例えば，X が一様分布 $U(a,b)$ に従うとき，$E(X)=\frac{a+b}{2}$ および $\sigma(X)=\frac{b-a}{2\sqrt{3}}$ であるため，任意の $c<d$ に対して，

$$
\begin{aligned}
P(c \le Z \le d) &= P\left(c \le \frac{X-(a+b)/2}{(b-a)/(2\sqrt{3})} \le d\right) \\
&= P\left(\frac{c(b-a)}{2\sqrt{3}}+\frac{a+b}{2} \le X \le \frac{d(b-a)}{2\sqrt{3}}+\frac{a+b}{2}\right) \\
&= \frac{1}{b-a}\int_{\frac{c(b-a)}{2\sqrt{3}}+\frac{a+b}{2}}^{\frac{d(b-a)}{2\sqrt{3}}+\frac{a+b}{2}} 1_{[a,b]}(x)\,dx = \frac{1}{2\sqrt{3}}\int_c^d 1_{[-\sqrt{3},\sqrt{3}]}(z)\,dz
\end{aligned}
$$

と計算できる．この計算結果より，この Z は一様分布 $U(-\sqrt{3},\sqrt{3})$ に従う．