確率統計 A

第 2 章 確率変数と分布 (Part 2)

(2012 年 5 月 8 日)

講師:若木宏文
E-mail: wakaki@math.sci.hiroshima-u.ac.jp
Web: http://home.hiroshima-u.ac.jp/~wakaki/lecture/probstatA12/

確率変数の定義

定義 2.1. $(\Omega, \mathcal{B}, \mathrm{P})$を確率空間とする. このとき, Ω上の実数値関数$X(\omega)$が任意の実数aに対して

$$\{\omega; X(\omega) \leq a\} \in \mathcal{B}$$

を満たすとき, 関数$X(\omega)$を$(\Omega, \mathcal{B}, \mathrm{P})$上, または$(\Omega, \mathcal{B})$上の**確率変数** (**Random Variable**: **rv**) という.

2.2. 分布関数

$(\Omega, \mathcal{B})$ 上の確率変数 X によって, $(\mathbb{R}, \mathbb{B}_1)$ 上の確率P_X が誘導されることを見てきた. この確率 P_X を扱う代わりに, 多くの場合, 本章で定義される分布関数を考える.

分布関数の定義

定義2.2. Xを $(\Omega, \mathcal{B})$ 上の確率変数とする. $\mathbb{R}$上で定義された関数

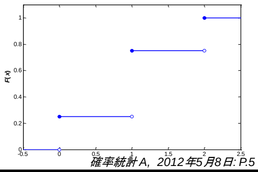
$$F_X(x) = \mathrm{P}(\{\omega \,;\, X(\omega) \leq x\})$$

を X の**分布関数**という. また, $F_X(x)$ を単に $F(x)$ とかき, 事象 $\{\omega \,;\, X(\omega) \leq x\}$ を単に "$X \leq x$" とかく.

確率変数と分布関数の例 (1/2)

ex 2.1 (1) 銅貨を 2 回投げる試行において，表の出た回数を表す確率変数を X とする．

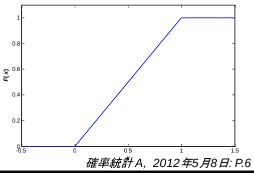
$$F(x)=\begin{cases}\mathrm{P}(\{\varnothing\})=0, & x<0,\\ \mathrm{P}(\{TT\})=1/4, & 0\le x<1,\\ \mathrm{P}(\{TT,TH,HT\})=3/4, & 1\le x<2,\\ \mathrm{P}(\{TT,TH,HT,HH\})=1, & 2\le x,\end{cases}$$



確率変数と分布関数の例 (2/2)

ex 2.1 (2) 区間 $(0,1]$ からランダムに1つの実数を選ぶ試行において，選ばれた実数を表す確率変数を X とする．

$$F(x)=\begin{cases}\mathrm{P}(\varnothing)=0, & x\le 0,\\ \mathrm{P}(0<X\le x)=x, & 0<x<1,\\ \mathrm{P}(0<X\le 1)=1, & 1\le x,\end{cases}$$



定理 2.3

任意の実数 a, b $(a<b)$ に対して，

$$\mathrm{P}(a<X\le b)=F(b)-F(a)$$

が成り立つ．

定理 2.3 の証明

定理 2.4

定理 2.4. 分布関数 $F(x)$ は次の性質をもつ.

(F0) $0 \le F(x) \le 1$.

(F1) (単調性) $x_1 < x_2$ ならば $F(x_1) \le F(x_2)$.

(F2) (右連続性) $F(x+0) = F(x)$.

(F3) $F(-\infty) = 0$, $F(\infty) = 1$.

定理 2.4 の証明

注意事項 1

注 2.4. 確率分布 P と確率変数 X が与えられると, その分布関数が定義 2.2 で定義された. 逆に, 分布関数 $F(x)$ が与えられると, 各区間 $(a, b]$ に対し,

$$\mathrm{P}((a, b]) = F(b) - F(a)$$

となる $(\mathbb{R}, \mathbb{B}_1)$ 上の確率 P_X が一意的に定義されることが知られている.

2.3. 多次元確率ベクトルと分布

多次元データ (Multivariate Data)

互いに**相関** (関係) がある観測値の組が n 個あるようなデータセット

	テストの点数				
番号	国語	社会	数学	理科	英語
1	80	96	81	100	93
2	60	61	39	57	57
3	63	50	41	45	59
4	100	98	75	93	98
5	56	41	26	40	32
6	74	77	65	84	71
7	86	83	62	74	80
8	65	52	43	62	57
9	75	66	45	60	69
10	57	64	47	58	57

それぞれの個人の点数を観測値ベクトル $\boldsymbol{x}_i$ とする:
ex) $\boldsymbol{x}_6=(74,\ 77,\ 65,\ 84,\ 71)'$

観測値行列 $\begin{pmatrix} \boldsymbol{x}_1' \\ \vdots \\ \boldsymbol{x}_n' \end{pmatrix}$ とする.

多次元確率変数について

$(\Omega,\mathcal{B},\mathrm{P})$ 上で定義された k 個の確率変数 $X_1,\ldots,X_k$ を考える. これら k 個の確率変数に関連した事象を同時に扱うことがしばしば生ずる. 例えば, 各 X_i が "$X_i \le x_i$" $(i=1,\ldots,k)$ となる事象, すなわち

$$
\begin{aligned}
"X_1 \le x_1,\ldots,X_k \le x_k" &= \{\omega ; X_1(\omega) \le x_1,\ldots,X_k(\omega) \le x_k\} \\
&= \bigcap_{i=1}^{k} \{\omega ; X_i(\omega) \le x_i\}
\end{aligned}
$$

の確率に関心がある場合である. このような場合, $X_1,\ldots,X_k$ の組 $\boldsymbol{X}=(X_1,\ldots,X_k)$ を考える.

多次元確率変数の分布関数

k次元**確率ベクトル** (**多次元確率変数**): $X_1,\ldots,X_k$ の組 $\boldsymbol{X}=(X_1,\ldots,X_k)$.

任意の $\boldsymbol{x}=(x_1,\ldots,x_k)\in\mathbb{R}^k$ に対して

$$F(x_1,\ldots,x_k)=\mathrm{P}(X_1 \le x_1,\ldots,\ X_k \le x_k)$$

によって定義される関数 $F(\boldsymbol{x})$ を $\boldsymbol{X}$ の分布関数あるいは $X_1,\ldots,X_k$ の**同時**(**joint**)**分布関数**あるいは**結合分布関数**という.

Example 2.2

$\mathbb{R}^2$ 上の区間 $\Omega=(0,1]\times(0,1]$ からランダムに点を選ぶ試行を考える. この試行によって選ばれた点を (X,Y) とする. このとき, (X,Y) の分布関数は次のように与えられる.

$$
F(x,y)=\mathrm{P}(X \le x, Y \le y)=\begin{cases} 0, & x<0 \text{ or } y<0, \\ xy, & 0 \le x < 1,\ 0 \le y < 1, \\ x, & 0 \le x < 1,\ \ y \ge 1, \\ y, & x \ge 1,\ 0 \le y < 1, \\ 1 & x \ge 1,\ y \ge 1. \end{cases}
$$

多次元の確率分布

1次元版 (定理 2.2). 確率変数 X によって任意の $A\in\mathbb{B}_1$ に対して$(\mathbb{R},\mathbb{B}_1)$ 上に

$$\mathrm{P}_X(A)=\mathrm{P}(X^{-1}(A))$$

となる確率分布 P_X が導入される.

多次元版: k 次元確率ベクトル $\boldsymbol{X}$ によって任意の $A_i\in\mathbb{B}_1$ $(i=1,\ldots,k)$ に対して $(\mathbb{R}^k,\mathbb{B}_k)$ 上に

$$\begin{aligned}\mathrm{P}_{\boldsymbol{X}}(A_1\times\cdots\times A_k)&=\mathrm{P}(\boldsymbol{X}^{-1}(A_1\times\cdots\times A_k))\\&=\mathrm{P}(\{\omega;\ X_1(\omega)\in A_1,\ldots,\ X_k(\omega)\in A_k\})\end{aligned}$$

となる確率分布 $\mathrm{P}_{\boldsymbol{X}}$ が導入される.

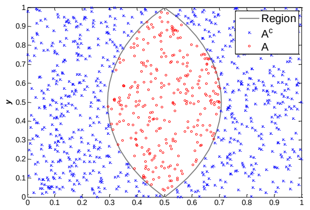
モンテカルロ法

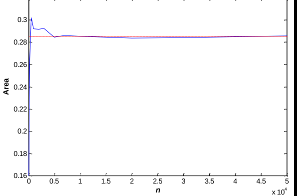
ex 2.2 の試行をn回繰り返すことによって, Ωに含まれる複雑な図形Aの面積を

$$\#(A)/n$$

として近似できる. ここに, $\#(A)$はn回の試行において図形Aに含まれる点の数を表す. この値はnを大きくすると図形Aの面積に近づくことが証明される (大数の法則). このような試行は乱数表を利用して行うことができる. また, このような方法はモンテカルロ法あるいはシミュレーション法と呼ばれる.

モンテカルロ法の例





定理 2.5

任意の実数 $a_i,\ b_i\ (a_i<b_i, i=1,\ldots,k)$ に対して

$$\begin{aligned}&\mathrm{P}(a_1<X_1\le b_1,\ldots,a_k<X_k\le b_k)\\&\qquad=\Delta_{x_1}(a_1,b_1)\cdots\Delta_{x_k}(a_k,b_k)F(x_1,\ldots,x_k)\end{aligned}$$

が成り立つ. ただし

$$\begin{aligned}&\Delta_{x_i}(a_i,b_i)F(x_1,\ldots,x_k)\\&\quad=F(x_1,\ldots,x_{i-1},b_i,x_{i+1},\ldots,x_k)-F(x_1,\ldots,x_{i-1},a_i,x_{i+1},\ldots,x_k)\end{aligned}$$

特に, $k=2$ とすると

$$\begin{aligned}&\mathrm{P}(a_1<X_1\le b_1,a_2<X_2\le b_2)\\&\qquad=F(b_1,b_2)-F(b_1,a_2)-F(a_1,b_2)+F(a_1,a_2)\end{aligned}$$

定理 2.5 の証明

定理 2.6

1次元版 (定理 2.4). 分布関数 $F(x)$ は次の性質をもつ.

(F0) $0\le F(x)\le 1$. (F1) (単調性) $x_1<x_2$ならば$F(x_1)\le F(x_2)$.

(F2) (右連続性)$F(x+0)=F(x)$. (F3) $F(-\infty)=0$, $F(\infty)=1$.

多次元版 (定理 2.6). k 次元確率変数 $\boldsymbol{X}=(X_1,\ldots,X_k)$ の分布関数 $F(x_1,\ldots,x_k)$ は次の性質をもつ.

(F0) $0\le F(x_1,\ldots,x_k)\le 1$.

(F1) 任意の実数 a_i,b_i $(a_i<b_i, i=1,\ldots,k)$ に対して

$$\Delta_{x_1}(a_1,b_1)\cdots\Delta_{x_k}(a_k,b_k)F(x_1,\ldots,x_k)\ge 0.$$

(F2) (右連続性) $F(x_1+0,\ldots,x_k+0)=F(x_1,\ldots,x_k)$.

(F3) $F(x_1,\ldots,x_{i-1},-\infty,x_{i+1},\ldots,x_k)=0$, $F(\infty,\ldots,\infty)=1$.

注意事項2

注 2.6) (F1), (F2) (F3) をみたす F に対して

$$\mathrm{P}((-\infty,x_1]\times\cdots\times(-\infty,x_k])=F(x_1,\ldots,x_k)$$

となる $(\mathbb{R}^k,\mathbb{B}_k)$ 上の確率分布 P が一意的に定まる.

一般に分布関数の定義において, すべての x_j $(j\ne i)$ について $x_j\to\infty$ とすると

$$\mathrm{P}(X_i\le x_i)=F(\infty,\ldots,\infty,x_i,\infty,\ldots,\infty)$$

が成り立つ. $F(\infty,\ldots,\infty,x_i,\infty,\ldots,\infty)$ は X_i の分布関数であるが, この場合**周辺分布関数** (**Marginal Distribution Function**) ともいう.

Example 2.3

Example 2.2 において, X の周辺分布関数は

$$F(x,y)=\mathrm{P}(X\le x, Y\le y)=\begin{cases}0=0, & x<0 \text{ or } y<0,\\ xy, & 0\le x<1,\ 0\le y<1,\\ x, & 0\le x<1,\ \ y\ge 1,\\ y, & x\ge 1,\ 0\le y<1,\\ 1 & x\ge 1,\ y\ge 1.\end{cases}$$

$$F_X(x)=\mathrm{P}(X\le x)=\mathrm{P}(X\le x, Y\le\infty)=\begin{cases}0, & x<0,\\ x, & 0\le x<1,\\ 1, & x\ge 1,\end{cases}$$

で与えられる. Y の周辺分布関数も同様に求められる.