第 1 章「確 率 変 数 と 確 率 分 布」

1．確率変数（Random Variable）

$X = X(\omega)\ :\ \Omega \to R^1$ への可測写像

・離散的 (discrete) な場合と連続的（continuous）な場合がある．

2．分布関数（Distribution Function）

確率変数 X が離散，連続いずれの場合でも，任意の実数 x に対して

$$F(x) = P(X \leq x) = P\left(\omega \in \Omega :\ X(\omega) \leq x\right)$$

が定義される．これを確率変数 X の分布関数という．

・分布関数の性質

(D1) 任意の x に対して $0 \leq F(x) \leq 1$ であり，かつ，

$$F(-\infty) = 0, \quad F(+\infty) = 1$$

(D2) 単調非減少，すなわち，$x < y$ ならば $F(x) \leq F(y)$
(D3) 右連続，すなわち，任意の x に対して $F(x) = F(x+)$ である．しかし，必ずしも左連続ではない．

3．確率関数と密度関数 (Probability and Density Functions)

・**確率関数**（離散的な確率変数に対して定義）

$$P(X = x) = P(X \leq x) - P(X < x)$$

・確率関数 $P(X = x)$ の性質

任意の x に対して $P(X = x) \geq 0$，かつ $\sum_x P(X = x) = 1$

・**密度関数**（連続的な確率変数に対して定義）

$$f(x) = \frac{d\,F(x)}{d\,x} = F'(x)$$

（注）連続的な場合には，

$$P(X = x) = 0 \quad (\text{任意の実数}\,x\,\text{に対して})$$

である．したがって，確率関数は無意味となる．

・密度関数の意味合い

$$f(x) = \lim_{h \to 0} \frac{F(x+h) - F(x)}{h} = \lim_{h \to 0} \frac{P(x < X < x+h)}{h}$$

より，h が微小のとき，$P(x < X < x+h) \cong h \times f(x)$ (底辺 $\times$ 高さ) である．

・密度関数による確率計算

$$\begin{aligned} P(a < X < b) &= P(a \leq X \leq b) = \int_a^b f(x)\,dx = F(b) - F(a) \\ P(X < c) &= P(X \leq c) = \int_{-\infty}^c f(x)\,dx = F(c) \end{aligned}$$

・密度関数 $f(x)$ の性質

任意の x に対して $f(x) \geq 0$，かつ $\int_{-\infty}^{\infty} f(x)\,dx = 1$

４．分布の特性値

・平均（Mean）（あるいは確率変数の**期待値（Expectation）**）：平均は分布の重心である．

$$E(X) = \begin{cases} \sum_x x\,P(X = x) & \text{（離散的な場合）} \\ \\ \int_{-\infty}^{\infty} x\,f(x)\,dx & \text{（連続的な場合）} \end{cases}$$

（注）とりうる値が加算無限，あるいは連続的な確率変数の期待値は必ずしも存在しない．

例１： 離散的な場合

$$P(X = n) = \frac{6}{n^2\pi^2} \quad (n = 1, 2, \cdots), \qquad \sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}$$

例２： 連続的な場合

$$f(x) = \frac{1}{\pi}\,\frac{1}{x^2+1}, \quad \int_{-\infty}^{\infty} \frac{1}{\pi}\,\frac{1}{x^2+1} = \left[\frac{1}{\pi}\tan^{-1} x\right]_{-\infty}^{\infty} = 1$$

・分散（Variance）

$$V(X) = \begin{cases} \sum_x (x - E(X))^2 \, P(X = x) = \sum_x x^2 \, P(X = x) - E^2(X) & \text{（離散的な場合）} \\ \\ \int_{-\infty}^{\infty} (x - E(X))^2 \, f(x) \, dx = \int_{-\infty}^{\infty} x^2 \, f(x) \, dx - E^2(X) & \text{（連続的な場合）} \end{cases}$$

・標準偏差（Standard Deviation）

$$SD(X) = \sqrt{V(X)}$$

を確率変数 X の標準偏差という.

・標準化（Normalization）

$$Z = \frac{X - E(X)}{SD(X)}$$

は期待値 0, 分散 1 の確率変数となる．この変換を確率変数 X の**標準化**という.

・k 次のモーメント

$$E(X^k) = \begin{cases} \sum_x x^k \, P(X = x) & \text{（離散的な場合）} \\ \\ \int_{-\infty}^{\infty} x^k \, f(x) \, dx & \text{（連続的な場合）} \end{cases}$$

定理　k 次の絶対モーメント $E(|X|^k)$ が存在すれば，それより低い次数の絶対モーメントが（したがってモーメントも）存在する.

[証明ーその１] $0 < j < k$ のとき，任意の x に対して

$$|x|^j \leq 1 + |x|^k$$

が成り立つ．なぜなら，$|x| \leq 1$ ならば $|x|^j \leq 1 \leq 1 + |x|^k$ であり，$|x| > 1$ ならば $|x|^j \leq |x|^k$ となるからである．したがって,

$$E\left(|X|^j\right) \leq E\left(1 + |X|^k\right) = 1 + E\left(|X|^k\right) < \infty \qquad \text{（等号成立の理由は５節参照）}$$

を得る.

[証明ーその２] $0 < j < k$ のとき，イェンセン（Jensen）の不等式（下を参照）を使って

$$E\left(|X|^k\right) = E\left((|X|^j)^{k/j}\right) \geq \left(E(|X|^j)\right)^{k/j}$$

を得る．すなわち,

$$\left(E\left(|X|^k\right)\right)^{1/k} \geq \left(\left(E(|X|^j)\right)\right)^{1/j} \qquad (j < k)$$

となる.

・イェンセンの不等式：$h(x)$ が凸（convex）関数ならば,

$$E(h(X)) \geq h(E(X))$$

が成り立つ.

[証明] 凸関数の性質から，$h(X)$ を下から支える直線

$$h(X) \geq h(\mu) + b(X - \mu)$$

が存在する．ただし，$\mu = E(X)$ である．ここで両辺の期待値をとれば結論を得る.

(注) 上の定理の証明では，$|X|^\alpha$ $(\alpha > 1)$ が凸関数となることを使った.
(問１) k 次のモーメントが存在すれば，$E(X - E(X))^k$ が存在することを示せ.

５．確率変数の関数の分布

$g(X)$: 確率変数 X の可測関数

は確率変数である．このとき，

$$E\left(g(X)\right) = \begin{cases} \sum_x g(x) \times P(X = x) & \text{（離散的な場合）} \\ \\ \int_{-\infty}^{\infty} g(x)\, f(x)\, dx & \text{（連続的な場合）} \end{cases}$$

(問２) a, b を定数とするとき，$E(aX + b) = aE(X) + b$ となることを示せ.
(問３) $V(X) = E(X - E(X))^2 = E(X^2) - E^2(X)$ と表せることを示せ.
(問４) $|E(g(X))| \leq E(|g(X)|)$ となることを示せ.

６．積率母関数（Moment Generating Function）など

・確率母関数（Probability Generating Function）

離散分布に対して，次のように定義される.

$$P(t) = E(t^X) = \sum_x t^x\, P(X = x)$$

例： 二項分布 $P(X = k) = {}_n C_k p^k (1 - p)^{n-k}$ の場合

$$\begin{aligned} P(t) &= \sum_{k=0}^{n} t^k\, {}_nC_k p^k (1-p)^{n-k} = \sum_{k=0}^{n} {}_nC_k (pt)^k (1-p)^{n-k} \\ &= (pt + 1 - p)^n \end{aligned}$$

・確率母関数の効用

$$E(X(X-1)\cdots(X-k+1)) = \left.\frac{d^k\, P(t)}{d\, t^k}\right|_{t=1} = P^{(k)}(1)$$

左辺のモーメントを k 次の**階乗モーメント**という.

（問５） 確率母関数を使って，二項分布の平均と分散を求めよ．

・積率母関数（Moment Generating Function）

$$M(t) = E(e^{tX})$$

を確率変数 X の積率母関数という．

例： 二項分布 $P(X=k) = {}_n C_k p^k (1-p)^{n-k}$ の場合

$$\begin{aligned} M(t) &= \sum_{k=0}^{n} e^{tk}\,{}_nC_k p^k (1-p)^{n-k} = \sum_{k=0}^{n} {}_nC_k (pe^t)^k (1-p)^{n-k} \\ &= (pe^t + 1 - p)^n \end{aligned}$$

例： 標準正規分布 $f(x) = e^{-x^2/2}/\sqrt{2\pi}$ の場合

$$\begin{aligned} M(t) &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{tx} \exp\left\{-\frac{x^2}{2}\right\} dx \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{(x-t)^2 - t^2}{2}\right\} dx \\ &= e^{t^2/2} \end{aligned}$$

（注） 積率母関数はすべての t に対して存在するとは限らない．また，存在しない場合もある．

・積率母関数の効用

$$E(X^k) = \left.\frac{d^k M(t)}{dt^k}\right|_{t=0} = M^{(k)}(0)$$

（問６） 積率母関数を使って，二項分布と標準正規分布の平均と分散を求めよ．
（問７） 標準正規分布の積率母関数を使って，$\mathrm{N}(\mu, \sigma^2)$ の平均と分散を求めよ．

・特性関数（Characteristic Function）

$$\phi(t) = E(e^{itX}) = M(it) \qquad (i \text{ は虚数単位})$$

を確率変数 X の特性関数という．特性関数は常に存在する．

・特性関数の効用

1. 分布関数と特性関数は１対１に対応する（フーリエ変換の性質）.

2. **逆変換公式**により，特性関数から分布関数を求めることができる.

$$F(y) - F(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{-ity} - e^{-itx}}{-it}\, \phi(t)\, dt$$

3. 非負値確率変数の場合には,

$$F(x) = \frac{1}{\pi} \int_0^{\infty} \mathrm{Re}\left[\frac{1 - e^{-itx}}{it}\, \phi(t)\right] dt$$

4. **（連続性定理）**分布関数の列 $\{F_n\}$ が分布関数 F に F のすべての連続点 x に対して収束 (**分布収束**) することと，対応する特性関数列 $\{\phi_n\}$ が ϕ に収束することは同値である．すなわち,

$$\lim_{n\to\infty} F_n(x) = F(x) \qquad (F \text{ のすべての連続点 } x \text{ に対して})$$

$$\Longleftrightarrow \lim_{n\to\infty} \phi_n(t) = \phi(t) \qquad (\text{任意の } t \text{ に対して})$$

注 連続性定理は中心極限定理の証明に使われる.

７．確率不等式（Probability Inequality）

・マルコフの不等式

非負値のみをとる確率変数 X（$\Leftrightarrow P(X \geq 0) = 1$）が有限の期待値をもつとき，任意の正数 a に対して

$$P(X \geq a) \leq \frac{1}{a}\, E(X)$$

（証明）非負の x に対して，次の関数を定義する.

$$g(x) = \begin{cases} 0 & (x < a) \\ a & (x \geq a) \end{cases}$$

このとき，$\mu = E(X) \geq E(g(X)) = a\, P(X \geq a)$ が成り立つ．ゆえに，不等式を得る.

・チェビシェフの不等式

確率変数 X が期待値 μ, 分散 σ^2 をもつとき，任意の正数 ε, k に対して

$$P\left(|X - \mu| \geq \varepsilon\right) \leq \frac{\sigma^2}{\varepsilon^2} \Longleftrightarrow P\left(|X - \mu| \geq k\sigma\right) \leq \frac{1}{k^2}$$

(注) 確率収束を証明する場合に使われる.

・片側チェビシェフの不等式

$$P\,(X-\mu \geq k\sigma) \leq \frac{1}{1+k^2}$$

(証明) 定数 c に対して，次の関数を定義する.

$$g(x)=(x-c)^2, \qquad h(x)=\begin{cases}(\mu+k\sigma-c)^2 & (x\geq \mu+k\sigma)\\ 0 & (x<\mu+k\sigma)\end{cases}$$

このとき，$E(g(X))\geq E(h(X))$ が成り立つ．あとは，適当に c を決めればよい.

・コーシー＝シュワルツの不等式

$$(E(XY))^2 \leq E^2(X)\,E^2(Y)$$

(注) 左辺の期待値は多変数の確率変数に関するもの．この不等式は相関係数の絶対値が常に 1 以下となることを示すのに使われる.

・カントロビッチの不等式

確率変数 X が $0<m\leq X\leq M$ の範囲で分布するとき，

$$E(X)\,E\left(\frac{1}{X}\right) \leq \frac{(m+M)^2}{4mM}$$

(証明) $0<m\leq x\leq M$ に対して，

$$0\leq (M-x)(x-m)=(M+m-x)x-Mm \qquad\Longleftrightarrow\qquad \frac{1}{x}\leq \frac{M+m-x}{Mm}$$

したがって，

$$E(X)\;E\left(\frac{1}{X}\right) \leq \frac{E(X)(M+m-E(X))}{Mm}$$

ここで，

$$E(X)(M+m-E(X)) = -(E(X)-(M+m)/2)^2+(M+m)^2/4 \leq (M+m)^2/4$$

より，不等式を得る.