

## 第3章 確率論の基礎

### 3.1 はじめに

前の章では、貸した資金が回収不能に陥る可能性を全く考慮せずに、利子率（収益率）を扱った。そこでは、資金が時間を通じてどのように運用されるかが最大の関心事となった（金利の期間構造！）

しかし、他人に金を貸すという行為（与信）を考えれば、常に貸し倒れの危険を避けることはできない。実際、日本の金融機関の多くは、この10年間、不良債権という名の貸し倒れの危険が異常に大きい金融資産をそのポートフォリオに抱え込む状態に苦しんできた。早い話、借用証書がタダの紙切れになる危険は、珍しくない。

危険 (risk), リスクは、不確実性に起因する経済的損失の可能性を広範に指す。ファイナンスにおいては、将来資産価格を確実に知ることができないことから生ずる危険が、主たる研究対象となる。そこで、将来資産価格を確率変数として扱うことで数学的に危険を扱うことが、必要となる。ここでは、ファイナンスにおける危険を扱うための、確率論の基礎を学ぶ。

### 3.2 確率

確率論は数学に1分野であり、現実の不確実性をとらえるものではない抽象的な演繹体系と考えることもできる。しかし、実際には確率論は他の多くの数学の分野と異なり、現実の不確実性を本来モデル化したものとして考えられている<sup>1</sup>。

確率は、集合に対して0と1の間の数を、整合的に対応づける写像として考えるのが、現代数学の立場である。そこでいう集合は、解釈的には、観測や実験・試行によってえられた結果をまとめたものである。さまざまな観測や実験・試行を行なった、すべての可能な結果  $\omega$  を集めた集合を標本空間  $\Omega$  とよぶ。

定義 1. 標本空間  $\Omega$  の部分集合をあつめたものを集合族という。 $\Omega$  の集合族  $\mathcal{F}$  が条件

1.  $\Omega \in \mathcal{F}$
2.  $A \in \mathcal{F} \implies A^c \in \mathcal{F}$
3.  $A_1, A_2, \dots, A_n, \dots \implies \bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$

を満たすとき  $\sigma$  集合族であるとよぶ。

注意 1. 標本空間の上の部分集合を集めたもののうち、 $\sigma$  集合族のようなものを考えるのは、事前にはどれが起こるかわからない可能性に、起こりやすさの数値としての確率を割り当てるとき、矛盾がおこらないようにするためと理解しよう。

例 1.  $\Omega = \{\text{晴}, \text{曇}, \text{雨}, \text{雪}\}$  とおくと、 $\mathcal{F} = \{\phi, \Omega, \{\text{晴}, \text{曇}\}, \{\text{雨}, \text{雪}\}\}$  は、 $\sigma$  集合族である。

演習 1. すぐ上の  $\Omega$  に対して、いろんな  $\sigma$  集合族が考えられる。いろいろ、作ってみよ。

<sup>1</sup> 確率論は、数学史的には「賭けの数学」として生まれ、統計学の基礎として発展したという側面をもつ。しかし、旧ソ連の数学者コルモゴロフが20世紀の前半に完全に抽象的な数学理論として再定式化するまで、多くの研究者が「確率とは何か」「不確実性とは何か」「将来がわからないということを確率で表すことに意味があるのか」などの半ば哲学的な問に囚われた。現代では、抽象的な確率論があるおかげで、数学としての確率論と現実の世界の不確実性を明確にわけて議論できるようになったともいえる。

記号	集合論の用語	確率論の用語	備考
$\Omega$	全体集合	標本空間	必ず起こる事象に対応
$\phi$	空集合	空事象	決して起こらない事象
$\omega \in \phi$	$\Omega$ の要素	状態 (state of nature)	根元事象
$A \in \mathcal{F}$	$\mathcal{F}$ の要素	事象	状態の集まったもの
$A^c$	補集合	余事象	
$\bigcup A_i$	合併集合	和事象	$\bigcup_{i=1}^n A_i$ あるいは $\bigcup_{i=1}^{\infty} A_i$
$\bigcap A_i$	共通部分	積事象	$\bigcap_{i=1}^n A_i$ あるいは $\bigcap_{i=1}^{\infty} A_i$
$A \cap B = \phi$	互いに素	$A$ と $B$ は排反事象	
$A \subset B$	部分集合	$A$ がおこるとき $B$ も起こる	

表 3.1: 集合論と確率論の用語の対応

演習 2. 標本空間  $\Omega$  としてサイコロの目の集まり, つまり  $\Omega = \{1, 2, 3, 4, 5, 6\}$  を考えるとき, いろいろな  $\sigma$  集合族が考えられる. いろいろ, 作ってみよ.

注意 2. 標本空間  $\Omega$  に対して,  $\sigma$  集合族はいくつも考えられることが, 上の演習からわかる.  $\Omega$  の部分集合族  $\mathcal{G}$  ( $\sigma$  集合族と限らない) を含む 2 つの  $\sigma$  集合族  $\mathcal{F}_1, \mathcal{F}_2$  を考えると,  $\mathcal{F}_1 \cap \mathcal{F}_2$  もまた  $\mathcal{G}$  を含む  $\sigma$  集合族となる.

演習 3. 上のことを証明せよ.

$\sigma$  集合族に関しては, 次の定理が集合のド・モルガン則から導かれる.

定理 1.

$$A_1, A_2, \dots, A_n, \dots \in \mathcal{F} \implies \bigcap_{i=1}^{\infty} A_i \in \mathcal{F}$$

演習 4. 上の定理を証明せよ.

確率論の記号の使い方は, 集合論と重なるが, 特別な言いまわしをする.  $\sigma$  集合族  $\mathcal{F}$  の要素を事象という. 特に  $\Omega$  を全事象,  $\phi$  を空事象という. 表 3.2 で確率論の言葉遣いと集合論の言葉遣いを対応させておく.

定義 2. 標本空間  $\Omega$  と  $\sigma$  集合族  $\mathcal{F}$  の対  $(\Omega, \mathcal{F})$  を可測空間という.

定義 3. 可測空間  $(\Omega, \mathcal{F})$  に対して, 写像  $P: \mathcal{F} \rightarrow [0, 1]$  が

1.  $P(\Omega) = 1$
2. (可算加法性)

$$A_i \in \mathcal{F} (i = 1, 2, \dots), A_i \cap A_j = \phi (i \neq j) \implies P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

を満たすとき,  $P$  を確率測度あるいは単純に確率であるという. また 3 つ組  $(\Omega, \mathcal{F}, P)$  を確率空間という.

ただし  $[0, 1] = \{ x \in \mathbf{R} \mid 0 \leq x \leq 1 \}$  である

演習 5.  $P(\phi) = 0$  を証明せよ .

注意 3. 確率は事前に何が起こるか分からない事象に対して, 起こりやすさを表す  $0$  と  $1$  の間の数値を矛盾なく割り当てたものであると解釈できる.  $0$  と  $1$  の間の数値を割り当てるのは, 確率論が現代的に整備される以前から, 確率が事象の場合数を全事象の場合の数で割った比率として定義されていたことを踏襲している.

演習 6. 標本空間  $\Omega$  としてサイコロの目の集まり, つまり  $\Omega = \{1, 2, 3, 4, 5, 6\}$  を考えるとき, サイコロに歪みがないとするとどのような確率空間を考えるべきだろうか .

演習 7. 高校レベルの確率論に出てくる「5人のうち, 2人が女性である確率を求めよ」という問題において, 何を標本空間とし,  $\sigma$  集合族をどのように規定すると, 都合がいいだろうか . その  $\sigma$  集合族を規定した後, 確率をどう定義することが, 問題を解くことになるのか考えてみよ .

次の定理が, 定義 3 から導かれる .

定理 2.

$$A, B \in \mathcal{F} \implies P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

[証明]  $A \cup B$  を排反な事象で表すと,  $A \cup B = (A \cap B^c) \cup (A \cap B) \cup (A^c \cap B)$  ところが,  $A$  と  $B$  も  $P(A) = P(A \cap B) + P(A \cap B^c)$ ,  $P(B) = P(A \cap B) + P(A^c \cap B)$  のように排反な事象を用いて表すことができる. これらを最初の関係式に代入して, 所望の結果を得る (証明おわり)

演習 8.  $P(A \cup B \cup C)$  に関して, 定理 2 と同様の関係を導いてみよ. (排反な事象の和として  $A \cup B \cup C$  を考えるのがヒント)

次の定理は確率の連続性とよばれる .

定理 3. 確率空間  $(\Omega, \mathcal{F}, P)$  において,

$$A_1 \subset A_2 \subset \cdots \subset A_n \subset \cdots \in \mathcal{F} \implies P\left(\bigcup_{n=1}^{\infty} A_n\right) = \lim_{n \rightarrow \infty} P(A_n)$$

[証明の概略]  $B_1 = A_1$ ,  $B_2 = A_2 \cap A_1^c$ ,  $B_i = A_i \cap (A_1 \cup \cdots \cup A_{i-1})^c$ , ( $i \geq 3$ ) とおくと,  $B_1, B_2, \cdots$  は互いに排反な事象であり,  $\bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} B_n$  であることを使う.

演習 9. 上の証明を完成させよ .

注意 4. 上の定理が連続性とよばれるのは,  $\bigcup_{n=1}^{\infty} A_n$  が増大列  $A_1, A_2, \cdots, A_n, \cdots$  の極限と考えられるからである. つまり  $\bigcup_{n=1}^{\infty} A_n$  を  $\lim_{n \rightarrow \infty} A_n$  のように記すことにすると, 定理 3 は

$$P(\lim_{n \rightarrow \infty} A_n) = \lim_{n \rightarrow \infty} P(A_n)$$

を主張しているからである .

演習 10. 事象の減少列  $A_1 \supset A_2 \supset \cdots \supset A_n \supset \cdots \in \mathcal{F}$  を考えると,  $P(\bigcap_{n=1}^{\infty} A_n) = \lim_{n \rightarrow \infty} P(A_n)$  が成立することを, 定理 3 とド・モルガン則を用いて示せ.

### 3.3 条件付確率

確率論にとって, 重要な条件付確率の概念を, ここで学ぶ.

定義 4. 確率空間  $(\Omega, \mathcal{F}, P)$  を考えるとき,  $A, B \in \mathcal{F}$  かつ  $P(B) \neq 0$  であるとき, 事象  $B$  に関する事象  $A$  の条件付確率  $P(A|B)$  を

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

と定義する.

定義 4 において,  $B \in \mathcal{F}$  を一つ決めて,  $A \in \mathcal{F}$  をいろいろ変えてみる. すると,  $\mathcal{F}$  から  $[0, 1]$  への写像として  $P(\cdot | B)$  が得られる. 次の定理が重要である.

定理 4.  $P(\cdot | B)$  は可測空間  $(\Omega, \mathcal{F})$  上の確率測度になっている.

演習 11. 定理 4 を確率測度の定義に立ちかえて, 証明せよ.

以下では大文字のアルファベットが明らかに事象であることが分かるときには,  $\mathcal{F}$  の要素であることを, いちいち断らないことにする.

条件付確率については, 以下の性質が成り立つ.

命題 1. 乗法公式  $P(A_1 \cap A_2 \cap \cdots \cap A_{n-1}) \neq 0$  のとき,

$$P(A_1 \cap A_2 \cap \cdots \cap A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \cdots P(A_n|A_1 \cap A_2 \cap \cdots \cap A_{n-1})$$

全確率公式 事象の集合  $\{A_i\}$  を  $\Omega$  の分割でそれぞれの確率はゼロでないとする. つまり  $\bigcup_{i=1}^{\infty} A_i = \Omega, A_i \cap A_j (i \neq j), P(A_i) \neq 0$ . このとき  $A \in \mathcal{F}$  に対して

$$P(A) = \sum_{i=1}^{\infty} P(A_i)P(A|A_i)$$

ベイズの定理 全確率公式と同様の条件に加えて  $P(A) \neq 0$  のとき

$$P(A_j|A) = \frac{P(A_j)P(A|A_j)}{\sum_{i=1}^{\infty} P(A_i)P(A|A_i)}$$

[証明] 乗法公式は  $n$  に関する帰納法による. 全確率公式は  $A = \bigcup_{i=1}^{\infty} (A \cap A_i)$  かつ  $A \cap A_i$  と  $A \cap A_j$  が  $i \neq j$  のとき排反であることから求められる. ベイズの定理は,  $P(A_j|A)$  を定義に戻って書き直し, 全確率公式を適用すればよい.

演習 12. 上の各命題を丁寧に証明してみよ.

### 3.4 事象の独立性

事象  $A$  と事象  $B$  が独立であるとは、ということかを学ぶ。

定義 5.  $P(A \cap B) = P(A)P(B)$  であるとき、事象  $A$  と事象  $B$  は独立であるという。

次の命題は重要である。

命題 2. 独立性と条件付確率  $P(B) \neq 0$  のとき、

$$\text{事象 } A \text{ と事象 } B \text{ は独立} \iff P(A|B) = P(A)$$

余事象の独立性 独立性は余事象に関しても成り立つ。つまり

$$\begin{aligned} \text{事象 } A \text{ と事象 } B \text{ は独立} &\iff \text{事象 } A \text{ と事象 } B^c \text{ は独立} \\ &\iff \text{事象 } A^c \text{ と事象 } B \text{ は独立} \\ &\iff \text{事象 } A^c \text{ と事象 } B^c \text{ は独立} \end{aligned}$$

演習 13. 上の命題を証明せよ。

注意 5. 3つ以上の事象の独立性を考えるとき扱いに注意する必要がある。まず、事象  $A$  と事象  $B$  が独立、事象  $B$  と事象  $C$  が独立、事象  $C$  と事象  $A$  が独立であっても、 $P(A \cap B \cap C) = P(A)P(B)P(C)$  が成り立つとは限らない（反例あり）また、 $P(A \cap B \cap C) = P(A)P(B)P(C)$  が成立していても、 $P(A \cap B) = P(A)P(B)$  が成立するとは限らない。（反例あり）

演習 14. 上の注意に関する2つの反例をそれぞれ作ってみよ（2番目は特に難しい）

3つ以上の事象の独立性は以下のように定義する。

定義 6. 事象  $A_1, A_2, \dots, A_n$  をとるとき、2つ以上のすべての組み合わせ  $A_{i_1}, A_{i_2}, \dots, A_{i_\ell}$  について、

$$P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_\ell}) = P(A_{i_1})P(A_{i_2}) \cdots P(A_{i_\ell})$$

であるとき、事象  $A_1, A_2, \dots, A_n$  は互いに独立であるという。

### 3.5 確率変数

事前にはどういう数値をとるか不確実な変数をモデル化したものである確率変数について、ここでは学ぶ。

定義 7. 可測空間  $(\Omega, \mathcal{F})$  において、写像  $X : \Omega \rightarrow \mathbf{R}$  が任意の  $x \in \mathbf{R}$  に対して、半閉区間  $(-\infty, x] = \{y \in \mathbf{R} \mid y \leq x\}$  の逆像  $X^{-1}((-\infty, x])$  が  $\mathcal{F}$  の要素であるとき、 $X$  を確率変数とよぶ。

注意 6. 定義 7 を注意深く見ると、確率変数が可測空間上で定義されていて、特定の確率測度そのものからは独立に定義されていることがわかる。実は、実数直線上の半閉区間  $(-\infty, x]$  全体からなる集合を含む最小のシグマ集合族  $\mathcal{B}$  が存在する。これをボレル集合族という。ボレル集合族は、半閉区間に加えて开区間・閉区間・片側閉片側開の区間・半开区間などあらゆる区間を含む集合族になっている。これにより、確率空間  $(\Omega, \mathcal{F}, P)$  が与えられると、区間  $B \in \mathcal{B}$  に対して  $P \circ X^{-1}(B)$  を対応させることで、写像  $P \circ X^{-1} : \mathcal{B} \rightarrow [0, 1]$  が得られて、確率測度になる。つまり、確率変数をひとつ考えることは、確率測度がきまると、実数上のあらゆる区間に対して確率を付与することができることになる。

定義 8. 可測空間  $(\Omega, \mathcal{F})$  を考えるとき、実数値関数  $f : \Omega \rightarrow \mathbf{R}$  は、任意の  $B \in \mathcal{B}$  に対して  $f^{-1}(B) \in \mathcal{F}$  であるならば、可測関数 (*measurable function*) とよばれる。

次の定理が成立する。

定理 5. 確率変数は、可測関数である。

注意 7. 可測関数は、明らかに確率変数である。

定理 6. 可測空間として  $(\mathbf{R}, \mathcal{B})$  をとる。このとき連続関数は、可測関数である。また、高々可算個の点を除いて連続な関数も可測関数である。

定理 7.  $X$  を可測空間  $(\Omega, \mathcal{F})$  上の確率変数、 $f$  を可測空間  $(\mathbf{R}, \mathcal{B})$  上の可測関数とする。 $f \circ X$  は、可測空間  $(\Omega, \mathcal{F})$  上の確率変数である。

注意 8. 確率論では、確率変数  $X$  に対してボレル集合  $B$  に対して  $X \in B$  と記すことで元々の確率空間  $(\Omega, \mathcal{F}, P)$  の  $\sigma$  集合体  $\mathcal{F}$  の要素  $\{\omega \in \Omega \mid X(\omega) \in B\}$  を表すことがある。同様に、 $a \leq X \leq b$  によって、 $[a, b] \in \mathcal{B}$  であるから  $\{\omega \in \Omega \mid X(\omega) \in [a, b]\}$  をあらわす。これにより、

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

あるいは

$$P(\{\omega \in \Omega \mid X(\omega) \in [a, b]\})$$

は、 $P(X \in B)$  あるいは  $P(a \leq X \leq b)$  のように略記できる。略記すると、「確率変数  $X$  が  $a$  と  $b$  の間の値をとる事象の確率」という感じが出るので都合がよい。

## 3.6 分布関数

統計学では、観測されるデータを確率変数の実現値として解釈する。大雑把に言えば、神様が振ったサイコロの目の結果を、われわれが観測していると考えerわけである。神様のサイコロは、確率空間  $(\Omega, \mathcal{F}, P)$  で規定されるが、分析の実際上は、可測空間  $(\Omega, \mathcal{F})$  上の確率変数  $X$  を決めて、確率空間として  $(\mathbf{R}, \mathcal{B}, P \circ X^{-1})$  をとるほうが、便利なが多い（抽象的な、不確実性よりは、数の不確実性に話題を限定したほうが便利ということ。）

可測空間  $(\Omega, \mathcal{F})$  上の確率変数  $X$  を考えて  $X^{-1}((-\infty, x]) = \{\omega \in \Omega \mid X(\omega) \leq x\}$  は、 $X$  が  $x$  以下である事象がどんなものを表す。

定義 9.  $X$  を確率空間  $(\Omega, \mathcal{F}, P)$  上の確率変数とする, 任意の  $x \in \mathbf{R}$  に対して  $F(x) = P \circ X^{-1}((-\infty, x])$  で定義される関数  $F: \mathbf{R} \rightarrow \mathbf{R}$  を分布関数 (*distribution function*) という.

注意 9. 分布関数は, データを統計処理したときの累積頻度表を理論化したものであるともいえる.

次の命題は重要である.

命題 3.  $a < b$  である任意の実数  $a, b$  に対して,

$$P(\{\omega \in \Omega \mid a < X(\omega) \leq b\}) = F(b) - F(a)$$

命題 4. 分布関数  $F$  について以下のことが成立する.

1.  $F$  は単調非減少関数である.
2. 任意の  $x \in \mathbf{R}$  に対して  $F(x) = \lim_{\varepsilon \rightarrow 0} F(x + \varepsilon)$
3.  $\lim_{x \rightarrow -\infty} F(x) = 0$
4.  $\lim_{x \rightarrow \infty} F(x) = 1$

注意 10.  $\lim_{x \rightarrow -\infty} F(x)$  が存在するとき  $F(-\infty)$ , 同様に  $\lim_{x \rightarrow \infty} F(x)$  が存在するとき  $F(\infty)$  と記す. さらに,  $\lim_{\varepsilon \rightarrow 0} F(x + \varepsilon)$  が存在するとき  $F(x+)$  とかくことがある.

### 3.7 離散型確率変数

ファイナンス数学 I で登場する確率変数の多くは離散型確率変数である. サイコロの目  $X$  は典型的な離散型確率変数である.  $X$  のとり得る値の可能性は 1, 2, 3, 4, 5, 6 の 6 個であり, それぞれの目が実現する確率は, 歪みがないサイコロであれば

$$Pr(X = x) = \frac{1}{6}, \quad (x = 1, 2, 3, 4, 5, 6)$$

つまり

とり得る値	確率
1	$\frac{1}{6}$
2	$\frac{1}{6}$
3	$\frac{1}{6}$
4	$\frac{1}{6}$
5	$\frac{1}{6}$
6	$\frac{1}{6}$

である.

定義 10. 確率空間  $(\Omega, \mathcal{F}, P)$  上の確率変数  $X$  に対して, 高々可算な集合  $E \in \mathcal{B}$  が存在して,  $P \circ X^{-1}(E) = 1$  となるとき,  $X$  を離散型確率変数とよぶ. とくに  $E$  が有限集合であるとき, 有限離散型確率変数という.



注意 11. 高々可算な集合  $E \subset \mathbf{R}$  は,  $E \in \mathcal{B}$  である. なぜならば, 1 個の数からなる集合  $\{x\}$  は

$$\{x\} = \bigcap_{n=1}^{\infty} \left[ x, x + \frac{1}{n} \right] \in \mathcal{B}$$

であり, そうした 1 個の要素からなるボレル集合を可算個集めたものもボレル集合になるからである.

上の定義 10 において  $P \circ X^{-1}(E) = 1$  は  $P(\{\omega \in \Omega \mid X(\omega) \in E\}) = 1$  であるから,  $E = \{x_1, x_2, \dots, x_n, \dots\}$  とすると,

$$\sum_{x \in E} P(X = x) = 1$$

を意味する. つまり, 離散型確率変数  $X$  で不確実な世界を眺めると, 高々可算個の事象で世界の可能性は分類されるということである. (離散型確率変数は, 高々可算個の異なる値しかとらない確率変数なのだが, 見方を変えるところということになる.)

ここで  $p_i = P(X = x_i)$  を頻度関数, 確率関数とよぶことがある.

例 2. 1 回の試行で結果  $A$  が起こる確率が  $p$  であるとき,  $n$  回試行を繰り返したとき結果  $A$  が起こる回数を確率変数  $X$  とおく.  $X$  は 0 から最大  $n$  まで  $n+1$  個の整数値をとる. 確率関数  $p_i = P(X = i)$ , ( $i = 0, 1, \dots, n$ ) は

$$p_i = {}_n C_i p^i (1-p)^{n-i}, (i = 0, 1, \dots, n)$$

である. このとき  $X$  は 2 項分布  $B(n, p)$  に従うという.

例 3. 取りうる値が 0 以上の整数値で, その確率関数が

$$p_i = e^{-\lambda} \frac{\lambda^i}{i!}, (i = 0, 1, \dots, n, \lambda > 0)$$

となる場合, 確率変数  $X$  はポアソン分布  $Po(\lambda)$  に従うとよばれる.  $Po(\lambda)$  は  $B(n, \lambda/n)$  を  $n \rightarrow \infty$  とした極限に等しい.

### 3.8 連続型確率変数

分布関数  $F$  が連続関数であるとき, 高々可算個の値をとるような確率変数でなくなる. つまり可算個の事象の確率の数え上げで, ことが済まなくなる. しかし, 分布関数の連続性に代わって, 追加的な条件を考えると, がぜん確率変数としてあつかいやすいものになる.

定義 11.  $F$  を確率空間  $(\Omega, \mathcal{F}, P)$  上の確率変数  $X$  の分布関数とする. 可測関数  $f$  が存在して, 任意の  $x \in \mathbf{R}$  に対して

$$F(x) = \int_{-\infty}^x f(\xi) d\xi$$

となるとき,  $F$  を絶対連続な分布関数,  $X$  を連続型確率変数,  $f$  を  $X$  の確率密度関数 (*probability density function*) あるいは簡単に密度関数とよぶ.

注意 12. 密度関数については,

$$f(x) \geq 0, (x \in \mathbf{R}), \int_{-\infty}^{\infty} f(\xi) d\xi = 1$$

が成立する. また  $x \in \mathbf{R}$  において  $f$  が連続ならば,

$$\frac{dF(x)}{dx} = f(x)$$

$$P(X = x) = F(x+) - F(x-) = 0$$

$$P(a < X \leq b) = \int_a^b f(\xi) d\xi$$

演習 15. 上のことを証明せよ.

注意 13. 密度関数は, 統計学に登場するヒストグラムをモデル化したものであるといってもよいかもしれない.

例 4. 密度関数が

$$f(x) = \begin{cases} \frac{1}{b-a} & (x \in (a, b)) \\ 0 & \text{その他} \end{cases}$$

であるとき, 確率変数は一様分布  $U(a, b)$  に従うという.

演習 16. 一様分布関数を求めよ.

### 3.9 多次元分布関数

第 3.4 では「神様のサイコロの性質を示す」確率空間  $(\Omega, \mathcal{F}, P)$  上の事象間の独立性の概念を学んだ. 現実には「神様がサイコロを振った結果」われわれ人間が観察することになる変量や変量間の性質のほうが, 確率空間  $(\Omega, \mathcal{F}, P)$  そのものの性質よりも関心の対象となることの方が多い.

すでに学んだように「神様がサイコロを振った結果」われわれ人間が観察することになった変量は, 事前に値が分からないという確率変数(実体は可測関数)というもので, 表すことができる. さらに, その確率変数の値の実現の仕方は, 分布関数や確率密度関数という解析学の分析対象を通じて表現されることを学んだ.

ここでは, 複数の確率変数の間の関係を扱う枠組みを示す.

#### 3.9.1 2次元分布関数

$X$  と  $Y$  を確率空間  $(\Omega, \mathcal{F}, P)$  上の確率変数とする. 今  $I = \{(x, y) \in \mathbf{R}^2 \mid a_1 \leq x \leq b_1, a_2 \leq y \leq b_2\}$  という 2次元平面上の矩形領域を考える. ただし,  $a_1, a_2 \in \mathbf{R} \cup \{-\infty\}$ ,  $b_1, b_2 \in \mathbf{R} \cup \{\infty\}$  である. このとき集合

$$\left\{ \omega \in \Omega \mid (X(\omega), Y(\omega)) \in I \right\}$$

はシグマ集合  $\mathcal{F}$  の要素である . なぜならこの集合は

$$\left\{ \omega \in \Omega \mid a_1 \leq X(\omega) \leq b_1 \right\} \cap \left\{ \omega \in \Omega \mid a_2 \leq Y(\omega) \leq b_2 \right\}$$

だからである .

定義 12. 確率空間  $(\Omega, \mathcal{F}, P)$  上の 2 つの確率変数の対  $(X, Y)$  を 2 次元確率ベクトルとよぶ .

定義 13.  $(X, Y)$  を確率空間  $(\Omega, \mathcal{F}, P)$  上の 2 次元確率ベクトルとする . このとき任意の  $(x, y) \in \mathbf{R}^2$  に対して

$$F_{XY}(x, y) = P(\left\{ \omega \in \Omega \mid X(\omega) \leq x, Y(\omega) \leq y \right\})$$

を 2 次元分布関数あるいは ,  $X$  と  $Y$  の同時分布関数とよぶ .

ここで 2 次元の同時分布関数の性質を命題の形にまとめておく .

命題 5.  $F_{XY}$  を  $X$  と  $Y$  の同時分布関数とする . このとき以下の性質が成立する .

1.  $x_1 \leq x_2, y_1 \leq y_2 \implies F_{XY}(x_1, y_1) \leq F_{XY}(x_2, y_2)$
2.  $X$  と  $Y$  の分布関数をそれぞれ  $F_X, F_Y$  と記すことにすると ,

$$(\forall x)[x \in \mathbf{R} \implies F_{XY}(x, \infty) = F_X(x)]$$

$$(\forall y)[y \in \mathbf{R} \implies F_{XY}(\infty, y) = F_Y(y)]$$

(  $F_{XY}(x, \infty), F_{XY}(\infty, y)$  をそれぞれ  $X$  と  $Y$  の周辺分布関数とよぶ . )

3.  $F_{XY}(-\infty, y) = 0, F_{XY}(x, -\infty) = 0$
4.  $F_{XY}(\infty, \infty) = 1$
5.  $P(a_1 < X \leq b_1, a_2 < Y \leq b_2) = F_{XY}(b_1, b_2) - F_{XY}(b_1, a_2) - F_{XY}(a_1, b_2) + F_{XY}(a_1, a_2)$

演習 17. 各自 , 上の性質を確認せよ .

### 3.9.2 $n$ 次元分布関数

今度は , 一般の  $n$  次元分布関数を考える .

定義 14. 確率空間  $(\Omega, \mathcal{F}, P)$  上の  $n$  個の確率変数の組  $(X_1, X_2, \dots, X_n)$  を  $n$  次元確率ベクトルとよぶ .

定義 15.  $(X_1, X_2, \dots, X_n)$  を確率空間  $(\Omega, \mathcal{F}, P)$  上の  $n$  次元確率ベクトルとする . このとき任意の  $(x_1, x_2, \dots, x_n) \in \mathbf{R}^n$  に対して

$$F_{X_1 X_2 \dots X_n}(x_1, x_2, \dots, x_n) = P(\left\{ \omega \in \Omega \mid X_1(\omega) \leq x_1, X_2(\omega) \leq x_2, \dots, X_n(\omega) \leq x_n \right\})$$

を  $n$  次元分布関数あるいは ,  $(X_1, X_2, \dots, X_n)$  の同時分布関数とよぶ .

注意 14.  $n$  次元同時分布関数に関しても命題 5 と同様の性質が成立する . また周辺分布に関しても同様に考えることができる .

### 3.10 確率変数間の独立性

この段階にいたって、複数の確率変数間の独立性が定義できる。

**定義 16.**  $X_1, X_2, \dots, X_n$  を確率空間  $(\Omega, \mathcal{F}, P)$  上の  $n$  個の確率変数とする。  $n$  個の任意の区間  $I_1, I_2, \dots, I_n$  に対して、  $n$  個の事象

$$\{\omega \in \Omega \mid X_1(\omega) \in I_1\}, \{\omega \in \Omega \mid X_2(\omega) \in I_2\}, \dots, \{\omega \in \Omega \mid X_n(\omega) \in I_n\}$$

が互いに独立であるとき、確率変数  $X_1, X_2, \dots, X_n$  は互いに独立であるという。

確率変数間の独立性は、分布関数を用いて表現することができる。

**定理 8.** 確率変数  $X_1, X_2, \dots, X_n$  が互いに独立であることの必要かつ十分条件は

$$F_{X_1 X_2 \dots X_n}(x_1, x_2, \dots, x_n) = F_{X_1}(x_1) F_{X_2}(x_2) \dots F_{X_n}(x_n), \quad (x_i \in \mathbf{R}, i = 1, 2, \dots, n)$$

[証明] 必要性は簡単である。十分性は  $n \geq 3$  の場合についてはきわめて煩雑である。

**演習 18.**  $n = 2$  の場合の定理 8 を証明してみよ。

#### 3.10.1 $n$ 次元離散型確率ベクトルの独立性

まず  $n$  次元離散型確率ベクトルを定義しておく。

**定義 17.**  $\mathbf{R}^n$  の可算集合  $E$  が存在して

$$P(\{\omega \in \Omega \mid (X_1(\omega) X_2(\omega) \dots X_n(\omega)) \in E\}) = 1$$

であるとき、  $(X_1(\omega) X_2(\omega) \dots X_n(\omega))$  を  $n$  次元離散型確率ベクトルという。

**注意 15.** 確率変数  $X$  について、  $\{\omega \in \Omega \mid X(\omega) \in B\}$  を  $X \in B$  のように略記したように、  $n$  次元確率ベクトルについても、

$$\{\omega \in \Omega \mid (X_1(\omega) X_2(\omega) \dots X_n(\omega)) \in E\}$$

を  $(X_1 X_2 \dots X_n) \in E$  のように略記すると便利である。

**注意 16.** もし  $(X_1 X_2 \dots X_n)$  が離散型  $n$  次元確率ベクトルであるならば、各  $X_i, (i = 1, \dots, n)$  に対して、  $P(X_i \in E_i) = 1$  となる  $\mathbf{R}$  の可算集合  $E_i$  が存在する

**定理 9.**  $n$  個の確率変数  $X_1, X_2, \dots, X_n$  が互いに独立であるための必要十分条件は、注意 16 にある各  $E_i, (i = 1, \dots, n)$  の任意の要素  $x_i \in E_i$  に対して

$$P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) = \prod_{i=1}^n P(X_i = x_i) \quad (3.1)$$

が成立することである。

[証明] (必要性)  $X_1, X_2, \dots, X_n$  が互いに独立であるとする. 適当な  $\Delta > 0$  に対して区間  $I_i = (x_i - \Delta, x_i + \Delta)$  と定義すると確率変数の独立性の定義 17 より

$$P(X_1 \in I_1, X_2 \in I_2, \dots, X_n \in I_n) = \prod_{i=1}^n P(X_i \in I_i)$$

となる. ここで, 各区間  $I_i$  はボレル集合であることに注意.  $\Delta \rightarrow 0$  としたとき,  $I_i \rightarrow \{x_i\}$  であることと確率の連続性を意味する定理 3 より

$$P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) = \prod_{i=1}^n P(X_i = x_i)$$

が成立する.

(十分性) 任意の  $x_i \in E_i, (i = 1, \dots, n)$  に対して (3.1) が成立したとする. ボレル集合  $B_i = \{y \in E_i \mid y \leq x_i\}$  と定義すると

$$\begin{aligned} F_{X_1 X_2 \dots X_n}(x_1, x_2, \dots, x_n) &= \sum_{y_1 \in B_1} \sum_{y_2 \in B_2} \dots \sum_{y_n \in B_n} P(X_1 = y_1, X_2 = y_2, \dots, X_n = y_n) \\ &= \sum_{y_1 \in B_1} P(X_1 = y_1) \sum_{y_2 \in B_2} P(X_2 = y_2) \dots \sum_{y_n \in B_n} P(X_n = y_n) \\ &= F_{X_1}(x_1) F_{X_2}(x_2) \dots F_{X_n}(x_n) \end{aligned}$$

であるから, 定理 8 より,  $X_1, X_2, \dots, X_n$  が互いに独立であることがわかった.

### 3.10.2 $n$ 次元離散型確率ベクトルの条件付確率

$n$ 次元離散型確率ベクトルが与えられているとき, ベクトルの各要素の確率変数のとる値を条件とする, 条件付確率を考えることができる.

たとえば,

$$P(X_i = x_i \mid X_j = x_j) = \frac{P(X_i = x_i, X_j = x_j)}{P(X_j = x_j)}$$

を考えることで,  $X_j = x_j$  であるとき  $X_i$  が  $x_i$  となる条件付確率と定義できる.

注意 17. 定理 8 より, 2つの確率変数  $X_i, X_j, (i \neq j)$  が独立であるならば,  $P(X_i = x_i \mid X_j = x_j) = P(X_i = x_i)$  がわかる.

### 3.10.3 $n$ 次元連続型確率ベクトルの独立性

今度は,  $n$ 次元確率ベクトルが連続型である場合を考えてみる.

定義 18. 確率ベクトル  $(X_1(\omega), X_2(\omega), \dots, X_n(\omega))$  の分布関数を  $F_{X_1 X_2 \dots X_n}(x_1, x_2, \dots, x_n)$  とする. このとき,  $\mathbf{R}^n$  上の非負値可測関数  $f_{X_1 X_2 \dots X_n} : \mathbf{R}^n \rightarrow \mathbf{R}_+$  が存在して, 任意の  $(x_1, x_2, \dots, x_n) \in \mathbf{R}^n$  に対して

$$F_{X_1 X_2 \dots X_n}(x_1, x_2, \dots, x_n) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \dots \int_{-\infty}^{x_n} f_{X_1 X_2 \dots X_n}(\xi_1, \xi_2, \dots, \xi_n) d\xi_1 d\xi_2 \dots d\xi_n$$

が成立するとき，分布関数は絶対連続であるといい．確率変数ベクトルは連続であるという．また  $f_{X_1 X_2 \dots X_n}$  を確率ベクトル  $(X_1 X_2 \dots X_n)$  の密度関数とよばれる（あるいは  $X_1, X_2, \dots, X_n$  の同時密度関数とよばれる．）

連続型確率ベクトルに関しては，連続型確率変数と同様の性質がなりたつ．たとえば， $(x_1 x_2 \dots x_n)$  において密度関数  $f_{X_1 X_2 \dots X_n}$  が連続であるとき，

$$f_{X_1 X_2 \dots X_n}(x_1 x_2 \dots x_n) = \frac{\partial^n}{\partial x_1 \partial x_2 \dots \partial x_n} F_{X_1 X_2 \dots X_n}$$

が成立する．

また周辺密度関数も考えることができる．例えば  $n = 4$  のとき連続型確率ベクトル  $(X_1 X_2 X_3 X_4)$  に対する  $(X_1 X_4)$  の周辺分布関数は

$$F_{X_1 X_4}(x_1, x_4) = F_{X_1 X_2 X_3 X_4}(x_1, \infty, \infty, x_4)$$

であるが，その密度関数は

$$f_{X_1 X_4}(x_1, x_4) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X_1 X_2 X_3 X_4}(x_2, x_3) dx_2 dx_3$$

で与えられる．

連続型確率ベクトルに関しても定理9と同様の以下の定理が成立する．

定理10.  $n$ 次元連続型確率ベクトル  $(x_1 x_2 \dots x_n)$  の密度関数を  $f_{X_1 X_2 \dots X_n}$ ，各  $X_i$  の周辺密度関数を  $f_{X_i}$  とする．このとき  $X_1, X_2, \dots, X_n$  が互いに独立であるための必要十分条件は，任意の  $(x_1 x_2 \dots x_n) \in \mathbf{R}^n$  に対して

$$f_{X_1 X_2 \dots X_n}(x_1, x_2, \dots, x_n) = \prod_{i=1}^n f_{X_i}(x_i)$$

がなりたつことである．

演習19. 定理10を証明せよ．

この節をおわるにあたって，条件付密度関数を定義しておこう．

定義19.  $f_{XY}$  を  $(XY)$  の密度関数， $f_X, f_Y$  をそれぞれ  $X, Y$  の周辺密度関数とする． $f_Y(y) > 0$  であるとき

$$f_{X|Y}(x|y) = \frac{f_{XY}(x, y)}{f_Y(y)}$$

で定義される  $f_{X|Y}$  を  $Y = y$  を条件とする  $X$  の条件付密度関数という．また，

$$F_{X|Y}(x|y) = \int_{-\infty}^x f_{X|Y}(\xi|y) d\xi$$

で定義される  $F_{X|Y}$  を  $Y = y$  を条件とする  $X$  の条件付分布関数という．

### 3.11 期待値

この節では，確率空間  $(\Omega, \mathcal{F}, P)$  上の確率変数  $X$ ，その分布関数  $F$  を考える．

高度な話をするならば，確率論は，可測関数  $\phi: \mathbf{R} \rightarrow \mathbf{R}$  をとるとき， $F$  に対する広義の積分（ルベグ積分）を考えることで期待値あるいは分散他の高次モーメントに関する一般論が，統一的に展開される．

しかし，ここではルベグ積分の一般論を展開する余裕がないので，授業でよく用いられる，離散型確率変数と，密度関数をもつ絶対連続な分布をもつ連続型確率変数を分けて扱う．

定義 20. 離散型確率変数  $X$  を考える．このとき  $P(X \in E) = 1$  なる高々可算な集合を  $E = \{a_1, a_2, \dots\}$  であらわすとすると，確率関数は  $f(a_i) = P(X = a_i)$  である． $X$  の期待値  $E(X)$  を

$$E(X) = \sum_{i=1}^{\infty} a_i f(a_i) \quad (3.2)$$

で定義する．

注意 18. ルベグ積分の立場からみると，(3.2) の右辺は

$$\sum_{i=1}^{\infty} a_i f(a_i) = \int_{\mathbf{R}} x dF(x) \quad (3.3)$$

と書かれる．

定義 21. 連続型確率変数  $X$  を考える．分布関数を  $F$ ，密度関数を  $f$  と記すことにすると， $X$  の期待値  $E(X)$  は

$$E(X) = \int_{\mathbf{R}} x f(x) dx \quad (3.4)$$

で定義される．

注意 19. ルベグ積分の立場からみると，(3.4) の右辺は，離散型の場合同様

$$\int_{\mathbf{R}} x f(x) dx = \int_{\mathbf{R}} x dF(x) \quad (3.5)$$

と書かれる．

注意 20. ルベグ積分の立場からみると，確率空間  $(\Omega, \mathcal{F}, P)$  上の確率変数  $X$ ，その分布関数  $F$  を考え，さらに可測関数  $\phi: \mathbf{R} \rightarrow \mathbf{R}$  を考えるとき， $\phi(X)$  の期待値  $E(\phi(X))$  を

$$E(\phi(X)) = \int_{\mathbf{R}} \phi(x) dF(x)$$

と定義する．

証明を省くが，期待値に関しては離散型・連続型をとわず以下の定理が成立する．

定理 11. 1.  $\phi(x) = c$  (定数) のとき  $E(\phi(X)) = c$

2.  $E(|\phi(X)|) < \infty, E(|\psi(Y)|) < \infty$  のとき, 任意の実数  $a, b$  に対して,  $E(a\phi(X) + b\psi(X)) = aE(\phi(X)) + bE(\psi(X))$
3.  $E(|\phi(X)|) < \infty, E(|\psi(Y)|) < \infty$  でかつ,  $\phi(x) \leq \psi(x)$  としよう. このとき  $E(\phi(X)) \leq E(\psi(X))$
4.  $|E(\phi(X))| \leq E(|\phi(X)|)$

演習 20. 証明を各人考えよ.

確率ベクトル  $X = (X_1 X_2 \cdots X_n)$  と可測関数  $\phi: \mathbf{R}^n \rightarrow \mathbf{R}$  に対して  $\phi(X)$  の期待値は,  $Y = \phi(X)$  の期待値として定義される.  $X$  が離散型のときは, 確率関数  $f$  を使って

$$E(\phi(X)) = \sum_{i_1} \cdots \sum_{i_n} \phi(x_{i_1}, \cdots, x_{i_n}) f(a_{i_1}, \cdots, a_{i_n})$$

$X$  が連続型のときは, 密度関数  $f$  を使って

$$E(\phi(X)) = \int \cdots \int_{\mathbf{R}^n} \phi(x_{i_1}, \cdots, x_{i_n}) f(x_{i_1}, \cdots, x_{i_n}) dx_1 \cdots dx_n$$

で表される.

次の定理が成立する.

定理 12.  $X$  と  $Y$  が互いに独立で,  $E(|\phi_1(X)|) < \infty, E(|\phi_2(Y)|) < \infty$  が成立しているとき

$$E(\phi_1(X)\phi_2(Y)) = E(\phi_1(X))E(\phi_2(Y))$$

演習 21. 定理 12 を証明せよ.

### 3.12 分散・共分散・モーメント

前の節では, 期待値を可測関数  $\phi: \mathbf{R} \rightarrow \mathbf{R}$  を使った形で定義しておいた. これには理由がある. 実は,

$$E(\phi(X)) = \int_{\mathbf{R}} \phi(x) dF(x)$$

は, 分布関数に依存して計算される. そこで  $\phi$  をいろいろ変更することで  $F$  の性質がわかるのではないかと期待される (実際それは正しい) この目的で使われる可測関数として

$$\phi(x) = x^m, (m = 1, 2, 3, \cdots) \tag{3.6}$$

あるいは  $\mu = E(X)$  として

$$\phi(x) = (x - \mu)^m, (m = 1, 2, 3, \cdots) \tag{3.7}$$



定義 22. 可測関数として (3.6) の  $\phi$  を用いた時の  $\phi(X)$  の期待値, つまり  $E(X^m)$  を  $X$  の (原点まわりの)  $m$  次のモーメントといい,  $\mu'_m$  と書く. 可測関数として (3.7) の  $\phi$  を用いた時の  $\phi(X)$  の期待値, つまり  $E((X - \mu)^m)$  を  $X$  の (期待値まわりの)  $m$  次のモーメントといい,  $\mu_m$  とかく. 特に  $m = 2$  のときの期待値回りのモーメントを分散とよび  $V(X)$  あるいは  $Var(X)$  とかいたりする.

注意 21. 分散  $V(X)$  は定義より負の値をとることはない.

分散  $V(X)$  は  $X$  のバラツキを期待値  $\mu$  からのずれ (の 2 乗) を平均することで測っていると考えられる.

なお, 平均を表すギリシャ文字として  $\mu$  がよく使われるのに対して分散を表すギリシャ文字として  $\sigma^2$  がよく使われる. 分散の平方根を標準偏差といい,  $\sigma$  であらわす.

なお,  $m = 3, 4$  に関する期待値回りのモーメントも重要でそれらを

$$\frac{E((X - \mu)^3)}{\sigma^3}, \frac{E((X - \mu)^4)}{\sigma^4}$$

のように, 標準偏差  $\sigma$  を用いて基準化したものを歪度, 尖度とよぶ.

演習 22. 離散型, 連続型のそれぞれに関して, 分散  $V(X)$  の計算方法を示せ. さらに, 負の値をとることがないことを示せ.

ここで分散の性質をまとめておく.

命題 6.  $E(X^2) < \infty$  とする.

1.  $V(X) = E(X^2) - [E(X)]^2$
2.  $V(aX + b) = a^2V(X)$ , ( $a, b \in \mathbf{R}$ )
3.  $V(X) = 0$  ならば  $P(X = \mu) = 1$

演習 23. 上の命題の最初の 2 つを証明せよ.

次の定理はチェビシェフの不等式として知られている. 確率変数の分布の形によらずに成り立つ関係式であるために, 統計学において, データが事前にどのような分布をするか分からないときに, 統計的な推測をするときに役立つ.

定理 13. 平均  $\mu = E(X)$ , 分散  $\sigma^2 = V(X)$  であるとき, 任意の定数  $k \geq 0$  に対して以下の不等式が成立する.

$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2} \tag{3.8}$$

[証明]  $B = \{x \mid |x - \mu| \geq k\sigma\}$  とおく .

$$\begin{aligned}
 \sigma^2 &= \int_{\mathbf{R}} (x - \mu)^2 dF(x) \\
 &= \int_B (x - \mu)^2 dF(x) + \int_{B^c} (x - \mu)^2 dF(x) \\
 &\geq \int_B (x - \mu)^2 dF(x) \\
 &\geq \int_B k^2 \sigma^2 dF(x) \\
 &= k^2 \sigma^2 \int_B dF(x) \\
 &= k^2 \sigma^2 P(X \in B) \\
 &= k^2 \sigma^2 P(|X - \mu| \geq k\sigma)
 \end{aligned}$$

である . 途中で集合  $B$  では  $|x - \mu| \geq k\sigma$  が成立していることを使った (証明おわり)

2つの確率変数  $X, Y$  に対して共分散を定義しておく .

**定義 23.**  $\mu_X = E(X), \mu_Y = E(Y)$  とする .  $E((X - \mu_X)(Y - \mu_Y))$  を  $X$  と  $Y$  の共分散とよび ,  $Cov(X, Y)$  と記す . また  $V(X), V(Y)$  がともにゼロでないとき

$$\rho(X, Y) = \frac{Cov(X, Y)}{\sqrt{V(X)V(Y)}}$$

を相関係数という .

**命題 7.**  $E(|X|) < \infty, E(|Y|) < \infty, E(|XY|) < \infty$  のとき

1.  $Cov(X, Y) = E(XY) - E(X)E(Y)$
2.  $Cov(ax + b, cY + d) = acCov(X, Y)$
3.  $X, Y$  が互いに独立ならば  $Cov(X, Y) = 0$  (当然相関係数  $\rho(X, Y) = 0$ )

**演習 24.**  $V(X), V(Y)$  が存在するとして , 定数  $a, b$  に対して

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

を示せ .

**注意 22.**  $X, Y$  が互いに独立ならば , 共分散と相関係数がゼロであるが , 共分散あるいは相関係数がゼロであるからといって ,  $X, Y$  が互いに独立であるとは限らない (よく間違える人がいる !)

**演習 25.** とともに最低 3 個の値をとる 2 つの確率変数  $X, Y$  を考え , 共分散がゼロ (当然相関係数もゼロ) であるが独立ではない例を作れ .

この節の最後に , 各次のモーメントと分布関数を結びつけるものとして積率母関数を紹介しておく .

定義 24. 確率変数  $X$  に対して実数  $t$  の関数を

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

と定義し、これを積率母関数という。

注意 23. 積率母関数  $M_X(t)$  は定義式の右辺が定義される範囲での、関数と考えること。ただし、具体的な例として表れる、正規分布、2項分布、ガンマ分布などよく知られたの分布を考えるとき、大抵  $M_X(t)$  は全域で定義される。

以下、証明抜きで積率母関数の性質を紹介する。

命題 8. 1.  $M_{aX+b}(t) = e^{bt} M_X(at)$

2.  $X, Y$  が互いに独立であるならば、 $M_{X+Y}(t) = M_X(t)M_Y(t)$  (逆は成立しない)

3. 原点まわりの  $m$  次モーメント  $\mu'_m = E(X^m)$  を使うと

$$M_X(t) = 1 + \frac{\mu'_1}{1!}t + \frac{\mu'_2}{2!}t^2 + \cdots + \frac{\mu'_n}{n!}t^n + \cdots$$

が成り立つ。

最後の性質は重要である。 $t$  の関数としての積率母関数が分かっているならば任意の原点まわりのモーメントは  $\mu'_k = M_X^{(k)}(0)$  によって求まる。

さらに、ある条件の下で、積率母関数の同等性と分布の同等性が成立する。よって、2つの確率変数が同じ分布であることを示すには、積率母関数が同じであることを示せばよい。

注意 24. 積率母関数は、実用的に役に立つが、数学的には特性関数とよばれる

$$\varphi(t) = E(e^{itX})$$

の方が、汎用性が高い。しかし複素関数に関する知識が必要なので、ここでは扱わない。

### 3.13 条件付期待値

以前、3.3節で条件付確率を定義した。定理 4 が示すように、確率空間  $(\Omega, \mathcal{F}, P)$  において、条件付確率自体も条件となる可測集合  $S \in \mathcal{F}$  を固定すると、確率測度  $P(\cdot|S)$  とみなせる。当然、確率変数  $X$  が定まると、分布関数を  $P(\cdot|S)$  によって

$$F(x|S) = P(X^{-1}((-\infty, x]|S))$$

として、条件付分布関数を考えることができる。

定義 25. 確率空間  $(\Omega, \mathcal{F}, P)$  と確率変数  $X$  を考える。 $S \in \mathcal{F}$  に対して条件付分布  $F(x|S)$  を使って計算される

$$\int_{-\infty}^{\infty} x dF(x|B)$$

が値をもつとき、この値を  $S$  を条件とした  $X$  の条件付期待値とよぶ。

条件付期待値は非常に重要である．後でふれるように，回帰分析をはじめ，多くの推定にかかわることが条件付期待値をもとめることに帰着される．

条件  $S \in \mathcal{F}$  の与え方として，典型的なのは確率変数  $Y$  に対してボレル集合  $B \in \mathcal{B}$  の逆像

$$Y^{-1}(B) = \{\omega \in \Omega \mid Y(\omega) \in B\} \in \mathcal{F}$$

を考えることである．実際，ボレル集合  $B$  としてよくつかわるのは

$[a, b]$  区間,  $\{y\}$  1 点集合

などである．以下では，確率変数  $Y$  とボレル集合から決まる事象  $Y^{-1}(B)$  を  $Y \in B$  のように記す．このように書いたほうが，事前に値の分からない確率変数  $Y$  がある「範囲」の値をとる事象のように読みやすいからである．同様に条件付分布も  $P(X \leq x \mid Y \in B)$  を  $F_{X|Y \in B}(x)$  のように書く．

### 3.13.1 条件付期待値：離散型

さて  $X, Y$  をともに，離散型確率変数とする． $B$  をボレル集合とするととき，条件付確率関数と条件付期待値がどのように計算されるかをみる．さて定義より  $P(X \in E_X) = 1, P(Y \in E_Y) = 1$  となる高々可算な集合  $E_X = \{x_1, x_2, \dots\}, E_Y = \{y_1, y_2, \dots\}$  が存在する．このとき

$$f(x, y) = P(X = x, Y = y), \quad x \in E_X, y \in E_Y$$

を確率関数といった．このとき

$$\begin{aligned} F_{X|Y \in B}(x) &= \frac{P(X \leq x, Y \in B)}{P(Y \in B)} \\ &= \sum_{x_i \leq x} \frac{\sum_{y \in B \cap E_Y} f(x_i, y)}{P(Y \in B)} \end{aligned}$$

である．これより， $Y \in B$  を条件とする条件付確率関数は

$$f_{X|Y \in B}(x) = \frac{\sum_{y \in B \cap E_Y} f(x, y)}{P(Y \in B)}, \quad x \in E_X$$

がわかる．条件付期待値はこの  $f_{X|Y \in B}(x)$  を用いて

$$E[X|Y \in B] = \sum_{x \in E_X} x f_{X|Y \in B}(x)$$

で計算される．

演習 26. 2個の歪みのないサイコロをふる試行を考える．2つのサイコロの目の和を  $X$  で表わす．また第 1 のサイコロの目の値を  $Y$  であらわす． $B = \{2, 4, 6\}$  とおくととき

1. 高々可算な集合  $E_X, E_Y$  はどんな集合か
2. 確率関数  $f(x, y)$  はどんな関数か
3. 条件付分布関数  $F_{X|Y \in B}(x)$  と条件付確率関数  $f_{X|Y \in B}(x)$  を求めよ

4. 条件付期待値  $E[X|Y \in B]$  を求めよ

注意 25.  $B = \{b\}$  をとるとき, 条件付確率関数は

$$f_{X|Y \in B}(x) = \frac{f(x, y)}{f_Y(b)}$$

で与えられる. ここで  $f_Y(b)$  は周辺確率  $f_Y(b) = P(Y = b)$  である.

## 3.13.2 条件付期待値: 連続型

次に連続型の確率変数のうち密度関数をもつ場合の条件付期待値をあつかう. さて  $X, Y$  をともに, 連続型確率変数とする.  $B$  をボレル集合とすると, 条件付密度関数と条件付期待値が同時密度関数  $f(x, y)$  からどのように計算されるかをみる.

$$\begin{aligned} F_{X|Y \in B}(x) &= \frac{P(X \leq x, Y \in B)}{P(Y \in B)} \\ &= \int_{-\infty}^x \frac{\int_B f(u, v) dv}{P(Y \in B)} du \end{aligned}$$

である. これより,  $Y \in B$  を条件とする条件付密度関数は

$$f_{X|Y \in B}(x) = \frac{\int_B f(x, y) dy}{P(Y \in B)}$$

がわかる. 条件付期待値はこの  $f_{X|Y \in B}(x)$  を用いて

$$E[X|Y \in B] = \int_{-\infty}^{\infty} x f_{X|Y \in B}(x) dx$$

で計算される.

注意 26. ボレル集合として  $B = \{b\}$  のような 1 点集合を考えると, 実は条件付確率密度を上のように考えるのはまずい. 連続型の場合,  $P(Y \in B) = P(\{b\}) = 0$  となってしまうからである. 実は条件付期待値を測度論の範囲で議論する場合, 問題は生じてこないが, この講義のレベルでは, 1 点集合  $B = \{b\}$  に収束するボレル集合の減少列  $\{B_\varepsilon\}$  を使って, この節での定義の極限として

$$f_{X|Y=b} = \frac{f(x, y)}{f_Y(b)}$$

とするのが実用的である. ここで  $f_Y(y)$  は  $Y$  の周辺密度関数

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y) dx$$

である.

### 3.14 条件付期待値と最小2乗法

$X, Y$  を2つの確率変数とする.  $X$  と  $Y$  が独立でないとする.  $X$  の値が観察されるとき  $Y$  がどのような値が観察されるかを予想したいということはよくある ( $X$  を10年ものの国債価格,  $Y$  をTOPIX みたいな株価指数を考えるなど)

$X$  の値  $x$  と  $Y$  の値  $y$  に  $y = g(x)$  のような関数関係があったとして

$$\ell = E((Y - g(X))^2) \quad (3.9)$$

を最小にする  $g$  を求めるのが最小2乗法の原理である.

$X = x$  を所与としたときの  $Y$  の条件付期待値を  $E(Y|X = x)$  であらわす. 次の定理が成立する.

定理 14.  $E(Y^2) < \infty, E(g(X)^2) < \infty$  が成立する範囲で,  $\ell$  を最小にする  $g$  は  $E(Y|X = x)$  で与えられる.

(証明は省く)

実は  $E(Y|X = x)$  を  $Y$  の  $X$  への回帰とよぶ. データ解析の立場からは, すべての関数のうちから  $g$  を選ぶことよりも,

$$g(x) = \beta_0 + \beta_1 x$$

のような線形関数の範囲で  $g$  を探ることが多い.

演習 27. 線形関数の範囲で  $g$  を探す場合, (3.9) に上の線形関数を代入して計算した  $\ell$  に関して,

$$\frac{\partial \ell}{\partial \beta_0}, \frac{\partial \ell}{\partial \beta_1}$$

を求めよ. 求める  $g$  はこの2つがともに0になった場合である (最小化の必要条件)

$\beta_0, \beta_1$  に関する方程式が得られることになる.  $\beta_0, \beta_1$  をもとめよ.

(続く)