

確率空間における大数の強法則

石田陸哉

2020年1月31日

目次

1 はじめに

自分は来年度以降, データサイエンス系の大学院に進学することから, ゼミの中で来年度以降の準備として統計学を勉強していた. このレポートはその際に使った「入門 確率解析とルベーク積分」を元にした勉強ノートである. 基本的には主に「入門 確率解析とルベーク積分」を参考にしたが, 2,3 章は「講義: 確率・統計」, 4 章は名古屋大学 中島誠先生の「確率論講義ノート」も参考にした.

2 確率の古典的定義

そもそも確率には古典的定義と公理的定義とがあり、公理的定義を考える際に確率空間の考え方が必要になる。では、なぜ公理的定義が必要なのかを含めて確率の定義について記す。

「サイコロ1個を投げて1の目が出る確率」などを考えるときに用いるのがラプラスによる確率の古典的定義であり、どの単一事象も起こる確率が同じである（同様に確からしい）ということを前提にしている。

定義 2.1 (確率の古典的定義)

ある試行 T に対する標本空間 Ω に含まれるどの根元事象も同様の確からしさと起こるとする。このとき、任意の事象 A に対して A が起こる確率を $P(A)$ と書き、次で定義する。

$$P(A) = \frac{n(A)}{n(\Omega)}$$

上の定義で用いた言葉、記号は以下のように定義する。

1. 試行: 同じ条件下で繰り返し行うことのできる実験や観測
2. 根元事象: 試行により起こりうる事柄の最小単位
3. 標本空間: 根元事象全ての集合
4. 事象: 標本空間の任意の部分集合
5. $n(A)$: 事象 A に含まれる根元事象の個数

例えば、サイコロを1個投げる場合を考えると、

1. 試行: サイコロを1個投げる
2. 根元事象: $\{1(\text{の目が出る})\}, \{2\}, \dots, \{6\}$
3. 標本空間: $\Omega = \{1(\text{の目が出る}), 2, \dots, 6\}$
4. 事象: 偶数が出る事象 $A = \{2, 4, 6\}$, 4以下が出る事象 $B = \{1, 2, 3, 4\}, \dots$ など
5. $n(A) = 3, n(B) = 4, n(\Omega) = 6$

5より、偶数の目が出る確率は $P(A) = \frac{n(A)}{n(\Omega)} = \frac{1}{2}$ 。同様に、4以上の目が出る確率は $P(B) = \frac{2}{3}$ となる。以上が確率の古典的定義とその例であり、直感的なものとなっている。

3 確率の公理的定義

しかし、定義 (2.1) で確率を考えると、不都合が生じることがある。

例 3.1 (確率の古典的定義で生じる不都合な例)

実数上の区間 $[0,1]$ からでたらめに 1 点 X を選ぶとする。どの点も同様の確からしさで選ばれるものとする。このとき、ある 1 点 c を選ぶ場合の数は 1 通り、 $[0,1]$ から 1 点を選ぶ場合の数は無限通りある。したがってこの場合の古典的定義による確率は $\frac{1}{\infty} = 0$ である。すると、選んだ 1 点が $[\frac{1}{2}, 1]$ に含まれる確率も 0 になる。なぜなら 1 点 c を選ぶ確率は 0 であり、0 をいくつ足しても 0 であるからだ。直感的には $[0,1]$ から 1 点選んだときに $[\frac{1}{2}, 1]$ に含まれる確率は $\frac{1}{2}$ になりそうだがそれとは異なってしまふ。

この例を含めた古典的定義による不都合に決着をつけたのがコルモゴロフによる確率の公理的定義である。公理とは証明するものではなく前提として仮定されるものである。最小限の基本的性質を公理とし、その公理を満たすものは全て確率であるとした。

3.1 確率空間

古典的定義での言葉と対応させると、標本空間を Ω 、事象全体の集合を σ 代数 \mathcal{B} 、事象の確率 (測度) を P としたときに、その 3 つの組 (Ω, \mathcal{B}, P) のことを確率空間という。 Ω は確率を考えたい集合として、その部分集合族である σ 代数 \mathcal{B} と、 \mathcal{B} の各元に確率を与える確率測度 P について定義する。

3.1.1 代数 \mathcal{B}

σ 代数 \mathcal{B} は集合族とよばれる Ω の部分集合の集合であり、確率測度を考えることができる集合全体というイメージを厳格に定義する。

定義 3.1 (σ 代数 \mathcal{B})

集合 Ω の部分集合の族 \mathcal{B} が σ 代数 \mathcal{B} であるとは,

- (1) $\emptyset \in \mathcal{B}$
- (2) $A \in \mathcal{B}$ ならば, $A^c \in \mathcal{B}$
- (3) $A_1, A_2, \dots \in \mathcal{B}$ ならば, $\bigcup_{i=1}^{\infty} A_i \in \mathcal{B}$

を満たすことである.

このとき, (2), (3) よりド・モルガンの法則から $\bigcup_{i=1}^{\infty} A_i \in \mathcal{B}$ ならば $\bigcap_{i=1}^{\infty} A_i \in \mathcal{B}$ であることがいえる. このことから測れる集合全体は補集合をとること, 和集合をとること, 共通部分をとることについて可算回の演算をしても閉じていることがこの定義 3.1 によって保証されていることがわかる.

この性質を用いると以下の公式が導かれる.

公式 3.1 上極限と下極限の可測性

$A_1, A_2, \dots \in \mathcal{B}$ ならば $\limsup_{n \rightarrow \infty} A_n \in \mathcal{B}$ および $\liminf_{n \rightarrow \infty} A_n \in \mathcal{B}$ である.

ここでの可測性は確率を測ることができるという意味で, \mathcal{B} に含まれる集合には確率 P を与えるという前提のもとであることに注意する.

証明

$$\limsup_{n \rightarrow \infty} A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k$$

であるので, $A_1, A_2, \dots \in \mathcal{B}$ より, $\bigcup_{k=n}^{\infty} A_k \in \mathcal{B}$ であり, そのことから, $\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k \in \mathcal{B}$ が示された.

$$\liminf_{n \rightarrow \infty} A_n = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k$$

についても同様の証明ができる.

3.1.2 確率測度

(Ω, \mathcal{B}, P) を確率空間と呼ぶが, 集合 Ω とその上の σ 代数をペアにした (Ω, \mathcal{B}) を可測空間という. この \mathcal{B} の元に確率を与える P について定義する. この際 P は Ω 上の関数ではなく, σ 代数 \mathcal{B} から実数 \mathbb{R} (または区間 $[0, 1]$) への関数であり, 定義域と

値域がしっかり定まっている.

P が満たすべき条件は以下である.

定義 3.2 (確率測度)

P が可測空間 (Ω, \mathcal{B}) の上の確率測度であるとは,

- (1) 任意の $A \in \mathcal{B}$ について, $P(A) \geq 0$ をみたす. とくに $P(\Omega) = 1$ である.
- (2) $A_1, A_2, \dots \in \mathcal{B}$, かつ互いに素, すなわち $A_i \cap A_j = \emptyset (i \neq j)$ ならば,

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

をみたすことである.

また, $A \cap B \neq \emptyset$ であるときも,

$$P(A \cup B) = P(A \cap B^c) + P(B \cap A^c) + P(A \cap B)$$

のように分解することも可能で, かつ, $P(A) + P(B)$ も

$$P(A) + P(B) = P(A \cap B^c) + P(A \cap B) + P(A \cap B^c) + P(A \cap B)$$

のように分解できることから, 確率測度の劣加法性

$$P(A \cup B) \leq P(A) + P(B)$$

がいえる.

以上の定義からいくつか確率の基本性質がいえる.

定理 3.1 (確率の基本的性質)

$A, B \in \mathcal{B}$ としたとき,

- (1) $P(A^c) = 1 - P(A)$.
- (2) $P(\emptyset) = 0$.
- (3) $A \subset B$ ならば, $P(A) \leq P(B)$.
- (4) $P(A) \leq 1$

証明

(1) $\Omega = A \cup A^c$ で, A と A^c は互いに素だから, 定義 3.2 より,

$$\begin{aligned} 1 &= P(\Omega), \\ &= P(A \cup A^c), \\ &= P(A) + P(A^c), \\ P(A^c) &= 1 - P(A). \end{aligned}$$

(2) $\Omega^c = \emptyset$ だから, 定義 4.2(1) $P(\Omega) = 1$, 定理 3.1(1) より

$$\begin{aligned} P(\emptyset) &= 1 - P(\Omega), \\ P(\emptyset) &= 0. \end{aligned}$$

(3) $B = A \cup (A^c \cap B)$ であるから, 定理 3.1(1)(2) より,

$$P(B) = P(A) + P(A^c \cap B) \geq P(A).$$

(4) $A \subset \Omega$ であるから, 定義 4.2(1) $P(\Omega) = 1$, 定理 3.1(3) より

$$P(A) \leq P(\Omega) = 1.$$

3.1.3 確率変数

測度論では可測関数と呼ばれるものであり, 公式 3.1 で確認した可測性という言葉と関係してくる. 確率測度 P は σ 代数 \mathcal{B} から実数 \mathbb{R} への写像であったが, 確率変数 X は Ω から S への写像であり, S を見本空間と呼ぶ.

例えば確率変数 X^1 が日本人全員 $\Omega^1 = \{\omega_1, \omega_2, \dots, \omega_N\}$ (日本人の総数を N 人とする) の身長を測るものだとしたら, それぞれの身長の集合 $S_l = \{l_1, l_2, \dots, l_N\}$ (全員身長が異なるものとする) が見本空間であり, $X^1(\omega_i) = l_i \in S_l$ で表せる.

見本空間と一言で言っても, 硬貨投げであれば $\{0, 1\}$, サイコロ投げであれば $\{0, 1, 2, 3, 4, 5, 6\}$, と状況によって値域は変わってくる. そこで値域は広く取り, $X : \Omega \rightarrow \mathbb{R}$ と確率変数 X は Ω から実数全体 \mathbb{R} への写像と見なすことにする.

このとき,

定義 3.3 (確率変数)

確率空間 (Ω, \mathcal{B}, P) から \mathbb{R} への写像 X が確率変数であるとは, 任意の $a \in \mathbb{R}$ について, $(-\infty, a]$ の逆像 $X^{-1}((-\infty, a]) = \{\omega \in \Omega : X(\omega) \leq a\} \in \mathcal{B}$ をみたすことである.

この定義について, 具体的に考えてみる.

例 3.1) 硬貨投げ

$\Omega = \{T, H\}$ であるから, 部分集合族である \mathcal{B} は $\mathcal{B} = \{\emptyset, T, H, \Omega\}$. T(裏) なら 0, H(表) なら 1 を返す確率変数 X について, 逆像を考えると

$$X^{-1}((-\infty, a]) = \{\omega \in \Omega : X(\omega) \leq a\} = \begin{cases} \emptyset & a < 0 \\ T & 0 \leq a < 1 \\ \Omega & a \geq 1 \end{cases}$$

例 3.2) サイコロ投げ

$\Omega = \bigcup_{i=1}^6 D_i$ ($i \neq j$ ならば $D_i \cap D_j = \emptyset$), Ω の部分集合族 \mathcal{B} , $Y(\omega) = i$ ($\omega \in D_i$) となる確率変数 Y を考えると, Y は 1 から 6 までの値しか取らないので,

$$Y^{-1}((-\infty, a]) = \{\omega \in \Omega : Y(\omega) \leq a\} = \begin{cases} \emptyset & a < 1 \\ D_1 & 1 \leq a < 2 \\ D_1 \cup D_2 & 2 \leq a < 3 \\ D_1 \cup D_2 \cup D_3 & 3 \leq a < 4 \\ D_1 \cup D_2 \cup D_3 \cup D_4 & 4 \leq a < 5 \\ D_1 \cup D_2 \cup D_3 \cup D_4 \cup D_5 & 5 \leq a < 6 \\ \Omega & a \geq 6 \end{cases}$$

となり, どの集合もそれぞれの場合の σ 代数 \mathcal{B} に属していることがわかる.

ここで, \mathcal{B} は σ 代数であるから, 補集合について閉じているはず. つまり, 任意の $a \in \mathbb{R}$ について, X が可測関数ならば,

$$\{\omega \in \Omega : X(\omega) \leq a\}^c = \{\omega \in \Omega : X(\omega) > a\} \in \mathcal{B}.$$

であるから, 確率変数の定義は

任意の $a \in \mathbb{R}$ について, (a, ∞) の逆像 $X^{-1}((a, \infty)) \in \mathcal{B}$ をみたすことである.

と同値であることを確認しておきたい.

次に $(-\infty, a)$ について考える.

$$(-\infty, a) = \bigcup_{n=1}^{\infty} (-\infty, a - \frac{1}{n}]$$

であることに注目すると,

$$\begin{aligned} X^{-1}((-\infty, a)) &= \{\omega \in \Omega : X(\omega) < a\} \\ &= \bigcup_{n=1}^{\infty} \{\omega \in \Omega : X(\omega) \leq a - \frac{1}{n}\} \\ &= \bigcup_{n=1}^{\infty} X^{-1}((-\infty, a - \frac{1}{n}]). \end{aligned}$$

となる. $X^{-1}((-\infty, a - \frac{1}{n}])$ は \mathcal{B} に属することは既に確認した. \mathcal{B} に属する集合の加算個の和であるから, $(-\infty, a)$ の逆像も \mathcal{B} に属することがわかる. 補集合 $[a, \infty)$ の逆像についても同じことがいえる.

以上から任意の $a \in \mathbb{R}$ について $(-\infty, a]$, (a, ∞) , $(-\infty, a)$, $[a, \infty)$ の逆像は \mathcal{B} に属することがわかり, そこからさらに次の補題 3.1 がわかる.

補題 3.1

$a < b$ について, $[a, b]$, $[a, b)$, $(a, b]$, (a, b) などの逆像も \mathcal{B} に属している.

たとえば, $(a, b]$ については,

$$\begin{aligned} X^{-1}((a, b]) &= \{\omega \in \Omega : a < X(\omega) \leq b\} \\ &= \{\omega \in \Omega : X(\omega) > a\} \cap \{\omega \in \Omega : X(\omega) \leq b\} \\ &= X^{-1}((a, \infty)) \cap X^{-1}((-\infty, b]) \in \mathcal{B} \end{aligned}$$

のように示せる. また特に $X^{-1}(\{a\}) \in \mathcal{B}$ についても,

$$\{\omega \in \Omega : X(\omega) = a\} = X^{-1}((-\infty, a]) \cap X^{-1}([a, \infty)) \in \mathcal{B}$$

によって示せる.

確率変数 $X : \Omega \rightarrow \mathbb{R}$ であり, \mathbb{R} 上の σ 代数 \mathcal{B} を $\mathcal{F}_{\mathbb{R}}$ で表すとき, 任意の $A \in \mathcal{F}_{\mathbb{R}}$ について, $X^{-1}(A) = \{\omega \in \Omega : X(\omega) \in A\} \in \mathcal{B}$ になることが示される. このことを用いて, 確率空間 (Ω, \mathcal{B}, P) から可測空間 (S, \mathcal{F}) への写像が可測であることは, 任意の $A \in \mathcal{F}$ について $X^{-1}(A) \in \mathcal{B}$ が成り立つことと定義する.

たとえば, 2 回硬貨投げの空間における σ 代数 $\mathcal{B}^{1,2}$ は, それぞれの効果の裏表に対応する T_1, H_1, T_2, H_2 を用いて, $\mathcal{B}^{1,2} = \{\emptyset, H_1 \cap H_2, H_1 \cap T_2, T_1 \cap H_2, T_1 \cap T_2, H_1, T_1, H_2, T_2, (H_1 \cap H_2) \cup (T_1 \cap T_2), (H_1 \cap T_2) \cup (T_1 \cap H_2), \Omega \setminus (H_1 \cap H_2), \Omega \setminus (T_1 \cap T_2), \Omega \setminus (H_1 \cap T_2), \Omega \setminus (T_1 \cap T_2), \Omega\}$ で表す. $\{\omega \in \Omega : X_1(\omega) = i, X_2(\omega) = j\}$ の形の

集合は $\mathcal{B}^{1,2}$ に属していて、このことを X_1, X_2 は \mathcal{B}^1 で可測であると表現する.

ここで、 X_1 だけを考えてとき、 $\mathcal{B}^1 = \{\emptyset, H_1, T_1, \Omega\}$ を考えると、 $\{\omega \in \Omega : X_1(\omega) = i\}$ の形の集合は \mathcal{B}^1 に属していて、 $\mathcal{B}^{1,2}$ は無駄に大きいことがわかる. 逆に、 \mathcal{B}^1 より小さい σ 代数 (この場合は $\{\emptyset, \Omega\}$) を考えると、 $\{\omega \in \Omega : X_1(\omega) = i\}$ の形の集合は属していないことがわかる. つまり、 \mathcal{B}^1 は X_1 を考えるのにぴったりの σ 代数であり、このとき、 \mathcal{B} を X_1 の生成する σ 代数とよび、 $\mathcal{B}^1 = \sigma(X_1)$ のように表す. 複数の確率変数については、 X_1, X_2, \dots, X_n の生成する $\mathcal{B}^{1,2,\dots,n}$ は $\sigma(X_1, X_2, \dots, X_n)$ であると表現する.

大数の法則の証明にて必要な確率変数の独立性について定義する.

定義 3.4 確率変数の独立性

確率変数 X と Y が独立であるとは、 X の生成する σ 代数 $\sigma(X)$ と Y の生成する σ 代数 $\sigma(Y)$ が独立であるということ、複数の確率変数 X_1, X_2, \dots が独立なこともそれらが生成する $\sigma(X_1), \sigma(X_2), \dots$ が独立なことと定義する.

定義 3.5 σ 代数の独立性

2つの σ 代数 $\mathcal{A}_1 \in \mathcal{B}$ と $\mathcal{A}_2 \in \mathcal{B}$ が独立であるとは、任意の $A_1 \in \mathcal{A}_1$ と $A_2 \in \mathcal{A}_2$ について、事象 A_1 と A_2 が独立であることである.

事象が独立であるとは、確率空間 (Ω, \mathcal{B}, P) において、

$$P(A \cap B) = P(A) \times P(B) \quad (A, B \in \mathcal{B})$$

をみtasことである.

定義 3.6 複数の σ 代数の場合

3つ以上の σ 代数 $\mathcal{A}_1, \dots, \mathcal{A}_n$ が独立であるとは、任意の $A_1 \in \mathcal{A}_1, \dots, A_n \in \mathcal{A}_n$ について、

$$P\left(\bigcap_{i=1}^n A_i\right) = \prod_{i=1}^n P(A_i)$$

をみtasことで、無限個の σ 代数について、それらが独立であるとはそのうちの有限個を選んでも独立なことと定義する.

一般の確率変数について独立であるときに幾つかの公式が導ける.

公式 3.2 独立性と平均, 分散

確率変数 X と Y が独立ならば, 任意の $a \in \mathbb{R}$ について,

$$(1) E(XY) = E(X) \times E(Y)$$

$$(2) V(X + Y) = V(X) + V(Y)$$

$$(3) V(aX) = a^2V(X)$$

が成り立つ.

3.1.4 ボレル・カンテリの第一定理

σ 代数の性質を用いて確率論の収束定理を述べる.

定義 3.7 (ボレル・カンテリの第一定理)

$A_1, A_2, \dots \in \mathcal{B}$ が $\sum_{n=1}^{\infty} P(A_n) < \infty$ をみたすとき,

$$P\left(\limsup_{n \rightarrow \infty} A_n\right) = 0$$

証明

上極限の定義から,

$$\limsup_{n \rightarrow \infty} A_n = \bigcap_{n=1}^{\infty} \bigcup_{k \geq n} A_k$$

である. $\sum_{m=1}^{\infty} P(A_m) < \infty$ より, $\bigcup_{k \geq n} A_k$ は n について狭義単調減少であるから,

$$P(\limsup_{n \rightarrow \infty} A_n) = P\left(\lim_{n \rightarrow \infty} \bigcup_{k \geq n} A_k\right)$$

であることと, 確率測度の劣加法性を用いて,

$$0 \leq P(\limsup_{n \rightarrow \infty} A_n) = P\left(\lim_{n \rightarrow \infty} \bigcup_{k \geq n} A_k\right) = \lim_{n \rightarrow \infty} P\left(\bigcup_{k \geq n} A_k\right) \leq \lim_{n \rightarrow \infty} \sum_{k \geq n} P(A_k)$$

$\sum_{m=1}^{\infty} P(A_m) < \infty$ より, 右辺は 0 に収束するので $P(\limsup_{n \rightarrow \infty} A_n) = 0$.

4 大数の法則

4.1 大数の法則

定義 4.1 大数の法則

データに対応する確率変数 X_1, X_2, \dots, X_n において、このデータの平均に対応する確率変数 $\frac{X_1 + X_2 + \dots + X_n}{n}$ は n を大きくした極限を考えると母平均 m に近づく (p22).

大数の法則のなかでも、確率収束について述べた弱法則と概収束について述べた強法則とがある。ここでは収束の種類について細かくは述べないが、概収束の方が条件が強く、概収束するならば、確率収束するといえるということだけ述べ、弱法則と強法則、各々の証明をしていく。

4.2 大数の弱法則

定理 4.1 大数の弱法則

X_1, X_2, \dots は独立で、 $E(X_i) = m, V(X_i) < \infty$ とする。このとき、任意の ϵ について

$$\lim_{n \rightarrow \infty} P\left(\omega \in \Omega : \left| \frac{X_1(\omega) + X_2(\omega) + \dots + X_n(\omega)}{n} - m \right| < \epsilon\right) = 0$$

が成り立つ。

4.2.1 チェビシエフの不等式

定理 4.2 (チェビシエフの不等式)

a, b を正の数とすると、

$$P(\omega \in \Omega : |X(\omega)| > b) \leq \frac{E(|X|^a)}{b^a}$$

証明

$\{\omega \in \Omega : |X(\omega)| > b\} = A$ において,

$$E(|x|^a) = \int |X|^a dP \geq \int_A |X|^a dP$$

$|X| > b$ より,

$$\begin{aligned} E(|X|^a) &\geq \int_A |X|^a dP \geq \int_A b^a dP \\ &= b^a \times P(A) = b^a \times P(\omega \in \Omega : |X(\omega)| > b) \end{aligned}$$

$$\frac{E(|X|^a)}{b^a} \geq P(\omega \in \Omega : |X(\omega)| > b).$$

よって示された.

このチェビシエフの不等式を変形すると,

4.1

$$P(\omega \in \Omega : |X(\omega) - E(X)| > \epsilon) \leq \frac{V(X)}{\epsilon^2}$$

を導くことができる.

証明 $Y(\omega) = X(\omega) - E(X(\omega))$ とおくと, 期待値と分散の公式から,

$$V(X) = E[(X - E(X))^2] = E(Y^2)$$

チェビシエフの不等式 (補題) より,

$$P(\omega \in \Omega : |X(\omega) - E(X(\omega))| > \epsilon) = P(\omega \in \Omega : |Y(\omega)| > \epsilon) < \frac{E(Y^2)}{\epsilon^2} = \frac{V(X)}{\epsilon^2}$$

よって導かれた.

これを用いて大数の弱法則を証明する.

4.2.2 弱法則の証明

X_1, X_2, \dots, X_n を独立で同分布, $E(X_i) = m, V(X_i) = v < \infty$ の確率変数とすると,

$$S_n = \frac{X_1 + X_2 + \dots + X_n}{n}$$

の平均は

$$\begin{aligned} E(S_n) &= E\left(\frac{X_1 + X_2 + \dots + X_n}{n}\right) \\ &= \frac{1}{n} (E(X_1) + \dots + E(X_n)) = \frac{1}{n} nm = m \end{aligned}$$

分散は, X_i は独立だから公式 (4.1), より

$$\begin{aligned} V(A_n) &= V\left(\frac{X_1 + X_2 + \dots + X_n}{n}\right) \\ &= \frac{1}{n^2} V(X_1 + X_2 + \dots + X_n) = \frac{1}{n^2} nv = \frac{v}{n} \end{aligned}$$

S_n について式 (4.2) を用いると,

$$P\left(\omega \in \Omega : \left|\frac{X_1 + X_2 + \dots + X_n}{n} - m\right| > \epsilon\right) < \frac{v}{n\epsilon^2}$$

を得る. $n \rightarrow \infty$ を考えたとき, (右辺) $= \frac{v}{n\epsilon^2} \rightarrow 0$ がいえるので,

$$\lim_{n \rightarrow \infty} P\left(\omega \in \Omega : \left|\frac{X_1 + X_2 + \dots + X_n}{n} - m\right| > \epsilon\right) = 0$$

よって, 大数の弱法則がいえた.

4.3 大数の強法則

定理 4.3 (大数の強法則)

X_1, X_2, \dots は独立かつ同分布で, $E(X_i) = m, V(X_i) < \infty$ とする. このとき,

$$P\left(\omega \in \Omega : \lim_{n \rightarrow \infty} \frac{X_1(\omega) + X_2(\omega) + \dots + X_n(\omega)}{n} = m\right) = 1$$

が成り立つ.

この定理 (4.4) を証明するために幾つかの準備をする.

X_1, X_2, \dots は独立で同分布とする. 平均を m , 分散を v で表すとき,

$$S_n = X_1 + X_2 + \dots + X_n$$

を考える.

4.3.1 コルモゴロフの不等式

補題 4.1 (コルモゴロフの不等式)

$m = 0$ のとき, 任意の $a > 0$ について,

$$P(\omega \in \Omega : \max_{1 \leq k \leq n} |S_k| \geq a) \leq \frac{E[S_n^2]}{a^2}$$

証明

$v = \infty$ のとき, $E[S_n^2] = nv = \infty$ となり, 不等式を満たす.

$v < \infty$ のとき,

$$A = \{\omega \in \Omega : \max_{1 \leq k \leq n} |S_k(\omega)| \geq a\}$$

$$A_k = \{\omega \in \Omega : \max_{1 \leq l \leq k-1} |S_l(\omega)| < a, |S_k(\omega)| \geq a\}$$

とおくと, A_i は互いに素で $\bigcup_{k=1}^n A_k = A$ を満たすので,

$$\begin{aligned} E(S_n^2) &= \int S_n^2 dP \geq \int_A S_n^2 dP = \sum_{k=1}^n \int_{A_k} S_n^2 dP \\ &= \sum_{k=1}^n \int_{A_k} (S_k + (X_{k+1} + \dots + X_n))^2 dP \\ &= \sum_{k=1}^n \int_{A_k} S_k^2 dP + 2 \sum_{k=1}^n \int_{A_k} S_k (X_{k+1} + \dots + X_n) dP + \sum_{k=1}^n \int_{A_k} (X_{k+1} + \dots + X_n)^2 dP \end{aligned}$$

ここで, A_k と S_k は X_1, \dots, X_k で定まり, X_1, \dots, X_k は, $k+1 \leq j \leq n$ を満たす X_j とは独立であるから, 公式 () より,

$$\int_{A_k} S_k X_j dP = E[(S_k 1_{A_k}) X_j] = E[S_k 1_{A_k}] \times E[X_j] = 0$$

$E[X] = m = 0$ より, (第二項) = 0.

A_k 上で $|S_k(\omega)| \geq a$ なので $S_k(\omega)^2 \geq a^2$. よって,

$$\begin{aligned} E(S_n^2) &= \sum_{k=1}^n \int_{A_k} S_k^2 dP + \sum_{k=1}^n \int_{A_k} (X_{k+1} + \dots + X_n)^2 dP \\ &\geq \sum_{k=1}^n \int_{A_k} a^2 dP = a^2 P(A) \end{aligned}$$

$$P(A) \leq \frac{E(S_n^2)}{a^2}.$$

$A = \{\omega \in \Omega : \max_{1 \leq k \leq n} |S_k(\omega)| \geq a\}$ より, 示された.

4.3.2 クロネッカーの補題

補題 4.2 (クロネッカーの補題)

$\sum_{n=1}^{\infty} \frac{x_n}{n}$ が存在するならば,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n x_i = 0$$

証明

$y_n = \sum_{i=1}^n \frac{x_i}{i}$ とおく. このとき, x_i についての式にすると,

$$y_n - y_{n-1} = \frac{x_n}{n}$$

$$i(y_n - y_{n-1}) = x_i.$$

x_i の和について,

$$\begin{aligned} \sum_{i=1}^n x_i &= \sum_{i=1}^n i(y_n - y_{n-1}) = \sum_{i=1}^n \sum_{j=1}^i (y_n - y_{n-1}) \\ &= \sum_{j=1}^n \sum_{i=j}^n (y_n - y_{n-1}) = \sum_{j=1}^n \{(y_j - y_{j-1}) + (y_{j+1} - y_j) + \dots + (y_n - y_{n-1})\} \\ &= \sum_{j=1}^n (y_n - y_{j-1}) \end{aligned}$$

最後の式をある数 p で区切って, $\sum_{j=1}^p (y_n - y_{j-1})$ と $\sum_{j=p+1}^n (y_n - y_{j-1})$ の2つの和に分ける. $\lim_{n \rightarrow \infty} y_n = \sum_{i=1}^{\infty} \frac{x_i}{i}$ より, 収束することから, $\{y_n\}$ はコーシー列になっている. そこで, p は

$$\max_{p+1 \leq j \leq n} |y_n - y_{j-1}| < \epsilon$$

を満たすようにとる. このとき, 先ほど分けた2つの和は,

$$\begin{aligned} \left| \sum_{j=p+1}^n (y_n - y_{j-1}) \right| &\leq \max_{p+1 \leq j \leq n} |y_n - y_{j-1}| \times n \leq \epsilon \times n \\ \left| \sum_{j=1}^p (y_n - y_{j-1}) \right| &\leq \sum_{j=p+1}^n (|y_n| + |y_{j-1}|) \leq 2Kp \end{aligned}$$

を満たす. ここでは $|y_n| \leq K$ をなるように K を選んでいる.

ここで, $n > \frac{2Kp}{\epsilon}$ となる n を選ぶと, ($\epsilon > \frac{2Kp}{n}$)

$$\left| \frac{1}{n} \sum_{i=1}^n x_i \right| \leq \frac{1}{n} \left\{ \left| \sum_{j=1}^p (y_n - y_{j-1}) \right| + \left| \sum_{j=p+1}^n (y_n - y_{j-1}) \right| \right\} \leq \frac{2Kp}{n} + \epsilon < 2\epsilon$$

を得る. これによりこの補題の証明を終わる.

4.3.3 強法則の証明

証明 期待値が $E[X_i] = 0$, 分散が $V(X_i) = v (< \infty)$ である X_n を考える. $T_n = \sum_{k=1}^n \frac{X_k}{k}$ とおいて, $n > j$ のとき,

$$\begin{aligned} E[(T_n - T_j)^2] &= E \left[\left(\sum_{k=1}^n \frac{X_k}{k} - \sum_{k=1}^j \frac{X_k}{k} \right)^2 \right] \\ &= E \left[\left(\frac{X_{j+1}}{j+1} + \frac{X_{j+2}}{j+2} + \cdots + \frac{X_n}{n} \right)^2 \right] \\ &= E \left[\sum_{k=j+1}^n \left(\frac{X_k}{k} \right)^2 + 2 \sum_{l=j+1}^n \sum_{m=j+1(m \neq j)}^n \frac{X_l}{l} \times \frac{X_m}{m} \right] \end{aligned}$$

X_n の独立性を用いると,

$$\begin{aligned} &= \sum_{k=j+1}^n E \left[\frac{X_k^2}{k^2} \right] + 2 \sum_{l=j+1}^n \sum_{m=j+1(m \neq j)}^n \left(E \left[\frac{X_l}{l} \right] \times E \left[\frac{X_m}{m} \right] \right) \\ &= \sum_{k=j+1}^n \frac{1}{k^2} E[X_k^2] + 2 \sum_{i=j+1}^n \sum_{m=j+1(m \neq j)}^n \frac{1}{lm} (E[X_i] \times E[X_m]) \end{aligned}$$

$E[X_n] = 0$, $E[X_n^2] = v(X_n) = v$ より,

$$= \sum_{k=j+1}^n \frac{v}{k^2}$$

ここで, $\sum_{k=1}^{\infty} \frac{1}{k^2} < \infty$ であるから, 狭義単調増加な k_n を

$$\sum_{k=k_n+1}^{\infty} \frac{v}{k^2} < 2^{-3n}$$

をみたすようにとる.

$$A_n = \{\omega \in \Omega : \sup_{k \geq k_n} |T_k - T_{k_n}| > 2^{-n}\}$$

とおき, コルモゴロフの不等式 () を $T_n - T_{k_n}$ に適用すると,

$$\begin{aligned} P(A_n) &= P(\omega \in \Omega : \sup_{k \geq k_n} |T_k - T_{k_n}| > 2^{-n}) \\ &\leq \frac{E[(T_n - T_{k_n})^2]}{2^{-2n}} \\ &= 2^{2n} \sum_{k=k_n+1}^{\infty} \frac{v}{k^2} \end{aligned}$$

$\sum_{k=k_n+1}^{\infty} \frac{v}{k^2} < 2^{-3n}$ より,

$$P(A_n) < 2^{-n}$$

これより, $\sum_{n=1}^{\infty} P(A_n) < \infty$ をみたすので, ボレル = カンテリの第一定理 () より $P(\limsup_{n \rightarrow \infty} A_n) = 0$. ド・モルガンの法則から,

$$P((\limsup_{n \rightarrow \infty} A_n)^c) = P(\liminf_{n \rightarrow \infty} A_n^c) = 1 - P(\limsup_{n \rightarrow \infty} A_n) = 1$$

このとき, 補集合 A_n^c は $A_n^c = \{\omega \in \Omega : \sup_{k \geq k_n} |T_k - T_{k_n}| \leq 2^{-n}\}$ である.

$$\text{ここで, } \Omega' := \liminf_{n \rightarrow \infty} A_n^c$$

$$\Omega'' := \left\{ \omega \in \Omega : \lim_{n \rightarrow \infty} \sum_{i=1}^n \frac{1}{n} X_i = 0 \right\}$$

としたとき, $\Omega'' \supset \Omega'$ を示すことができれば, 確率の公式 () から $1 \geq P(\Omega'') \geq P(\Omega') = 1$ より $P(\Omega'') = 1$ がわかる. よって, ここから $\omega \in \Omega''$ が必ず, $\lim_{n \rightarrow \infty} \sum_{i=1}^n \frac{1}{n} X_i = 0$ を満たすことを示していく.

$$P(\liminf_{n \rightarrow \infty} A_n^c) = 1$$

より, $\bigcup_{n=1}^{\infty} \bigcap_{k \geq n} A_k^c$ の確率が 1 であることからほとんど全ての $\omega \in \Omega$ について, ある十分に大きな $N = N(\omega)$ が存在して, $k \geq N$ について $\omega \in A_k^c$. つまり,

$$|T_k - T_{k_n}| \leq 2^{-n}$$

を満たす. さらに, $k, l \geq N$ について, 三角不等式を用いると,

$$\begin{aligned} |T_k(\omega) - T_l(\omega)| &= |T_k(\omega) - T_{k_n}(\omega) + T_{k_n}(\omega) - T_l(\omega)| \\ &\leq |T_k(\omega) - T_{k_n}(\omega)| + |T_{k_n}(\omega) - T_l(\omega)| \\ &\leq \frac{1}{2^n} + \frac{1}{2^n} = \frac{1}{2^{n-1}} \end{aligned}$$

$n \rightarrow \infty$ のとき,

$$|T_k(\omega) - T_l(\omega)| \leq \frac{1}{2^{n-1}} \rightarrow 0$$

以上から, $T_k(\omega) = \sum_{i=1}^k \frac{X_i}{i}$ はコーシー列であり, 収束する. $\sum_{i=1}^k \frac{X_i}{i}$ が存在するので, クロネッカーの補題より $\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n X_i = 0$.

$\Omega'' \in \Omega'$ より,

$$P(\Omega'') = P(\omega \in \Omega : \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n X_i = 0) = 1$$

今回, 期待値 $E[X_i] = m = 0$ なので大数の強法則を満たしている. $E[X_i] = m \neq 0$ の場合でも, $Y_n = X_n - m$ と変数変換すると, $E[Y_n] = E[X_n - m] = E[X_n] - m = m - m = 0$ であるので, Y_n について同様に証明することができる.

5 参考

森真, 入門 確率変数とルベーグ積分, 東京書籍 (2012)

穴太克則, 講義: 確率・統計, 学術図書 (2011)

中島誠, 確率論講義ノート (2019)