2011 年度秋学期　解析応用　第１０回

確率変数の収束と大数の法則

大数の法則とは，ある試行によって値の決まる確率変数の期待値が μ であるとき，この試行を独立に何度も繰り返すと，それらの確率変数の平均が μ に収束する，という定理です。つまり，多数の独立な確率変数の平均は，ほぼその期待値と同じになる，ということで，統計的推測の基盤になっている定理です。今回は，まず確率変数の収束について説明し，それから大数の法則の意味を説明します。

確率変数の収束

確率変数の列 $X_1, X_2, \ldots$ があるとき，それが確率変数 X に収束するとはどういうことかを考えます。確率変数は，標本空間 Ω の要素 ω の関数ですから，これは関数列です。その収束の定義にはいろいろなものがありますが，そのうちいくつかを下で説明します。

概収束

確率変数の列 X_n が X に**概収束**するとは，確率 1 でおきる事象 Ω_0 に対して，

$$\lim_{n\to\infty} X_n(\omega) = X(\omega) \ (\forall \omega \in \Omega_0) \tag{1}$$

がなりたつことをいい，これを $X_n \xrightarrow{\text{a. s.}} X \ (n \to \infty)$ とかきます。a. s. は “almost surely” の略です。

「確率 1 でおきる事象」とは，全事象 Ω だけではありません。分布関数が連続である確率変数（いわゆる連続型確率変数）を考えます。この場合，確率変数 X が区間 $(a, b]$ をとる確率測度 $P((a, b])$ が，分布関数 F を使って $P((a, b]) = F(b) - F(a)$ で定められます。$a \to b$ のときを考えると，分布関数が連続なので，$P(X = a) = 0$ となります。

したがって，Ω から $X = a$ となる事象を除いた事象を考えても，その確率は 1 です。$X = a$ のひとつだけでなく，可算無限個の点（たとえば「X が有理数」となる事象）を除いても，確率測度は完全加法的ですから，やはりその確率は 1 です。

このような「測度が 0 となる点の集合」を**零集合**といいます。また，零集合を除いた全体で何かがなりたつことを**ほとんどいたるところ** (almost everywhere) でなりたつ，といい，a. e. と書きます。概収束は，この言い方でいうと $X_n \to X$ a. e. $(n \to \infty)$ となります。

確率収束

確率変数の列 X_n が X に**確率収束**するとは，任意の正の数 ε について

$$\lim_{n\to\infty} P(|X_n - X| \geqq \varepsilon) = 0 \tag{2}$$

がなりたつ，という意味で，$X_n \xrightarrow{\text{P}} X \ (n \to \infty)$ とも書きます。「任意の正の数 ε について」というのは，「どんなに小さな正の数 ε をもってきても」と読み替えるとわかりやすいと思います。「 X_n と X との隔たりがわずかでもある」確率は，n が大きいと 0 に近づく」ということになります。

法則収束

確率変数の列 X_n が X に**法則収束**するとは，確率変数 X_n, X に分布関数 F_n, F が対応しているとき，$F(x)$ が連続であるすべての点 x で

$$\lim_{n\to\infty} F_n(x) = F(x) \tag{3}$$

がなりたつ，という意味で，$X_n \xrightarrow{\mathrm{D}} X \ (n \to \infty)$ とも書きます。つまり，「分布が収束する」ということを表しています。

なお，これら３つの収束の間には，「概収束するならば確率収束する，確率収束するならば法則収束する」という関係があります（詳細は参考文献を見てください）。

統計的推測と大数の法則

統計的推測とは，対象のデータがものすごく多いなどの理由で，すべてのデータを調べることができないときに，限られた数のデータ（標本）だけを調べて，データ全体（母集団）の相対度数分布（母集団分布）を推測しようとするものです。

統計的推測の原理は，母集団分布と標本の確率分布の関係にあります。すなわち，

- **母集団の，ある階級の相対度数　＝　無作為抽出された標本がその階級に属する確率**
- **無作為抽出された標本は互いに独立な確率変数で，母集団分布と同じ確率分布にしたがう**

という関係です。簡単に言えば，「母集団の階級値 172.5cm の階級の相対度数が 10%なら，そこから標本を一人無作為抽出すると，その人が階級値 172.5cm の階級に入っている確率も 10%」という関係です。もっと簡単に言えば，「母集団のうち男性が 60%なら，そこからだれかひとりを『ズルをしないくじびきで』選ぶと，その人が男性である確率は 60%」という，ごく当たり前の関係です。

しかし，標本としてひとつのデータだけを取り出しても，母集団分布を知る手がかりには，ほとんどなりません。くじを１回だけひいて当たったからといって，そのくじの当たり確率はわからないのと同じです。では，どうすればよいのでしょうか？

その答えは，標本としてひとつのデータだけを取り出すのではなく，いくつものデータのセットになった標本を取り出すことです。その理由が**大数の法則**です。大数の法則とは，

> たくさんの独立な確率変数の平均は，『たいてい，ほぼ』その期待値と同じであり，期待値からかけ離れた値になる確率は非常に小さい

ということです。いくつかのデータのセットになった標本を取り出すとき，その個数を標本の**サイズ**といい，個数が多いことを**標本サイズが大きい**といいます[1]。

例えば，母集団の平均を，標本を取り出して推測するとしましょう。標本サイズが小さいと，図１のヒストグラムの上の例のように，たまたま極端に大きなデータばかり，あるいは極端に小さなデータば

[1]「標本サイズ」と「標本数」は，別の意味の言葉です。「標本数」は，標本のセットの数をさします。例えば，２つの母集団から，それぞれ n 個のデータのセットを標本として取り出すと，標本数は 2，標本サイズはそれぞれ n となります。

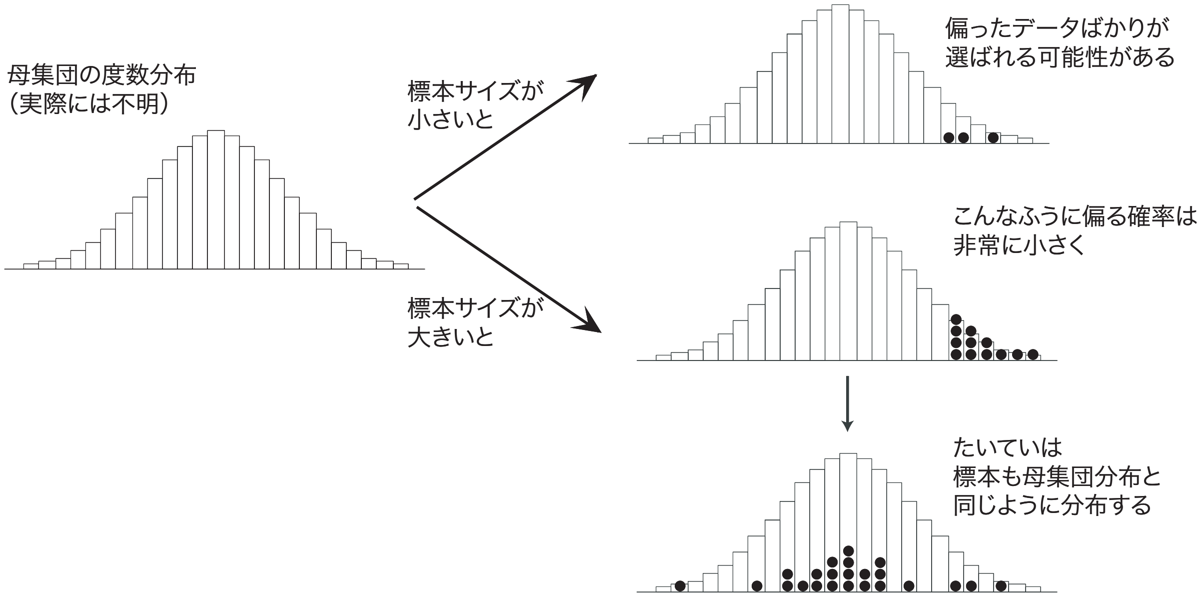

図 1: 標本サイズと標本の分布

かりが標本に選ばれる可能性があります。このような「偏った標本」の平均を求めても，それは母集団の平均とはかけ離れたものになっていますから，この推測は失敗です。ところが，標本サイズが大きくなると，図１の下のように，偏った標本が得られる確率は大変小さいので，上のような失敗をする確率は小さくなります。

ちょっと計算をしてみましょう。母集団の平均（母平均）が μ，分散が σ^2 であるとします。母集団から，n 個のデータからなる標本 $X_1, X_2, \ldots, X_n$ を取り出し，標本平均 $\overline{X_n} = (X_1 + X_2 + \cdots + X_n)/n$ を計算して，この値を母平均 μ の推測結果とする，としましょう。標本としてどんなデータが取り出されるかは，偶然によって決まります。つまり，図２のようにいま標本として取り出されたデータは，さまざまな可能性のうちのひとつでしかありません。ですから，その標本から計算された標本平均も，さまざまな可能性のうちのひとつということになります。

上のように，標本平均は，現に求められた値以外にも，さまざまな値になる可能性があるわけですが，では「標本平均のさまざまな可能性」の期待値 $E(\overline{X_n})$ と分散 $V(\overline{X_n})$ はどうなるでしょうか？ つまり，サイズが n の標本のセットを取り出す機会が何度も何度もあったとしたとき，標本平均の値は，長い目でみて平均いくらで，どのくらいのばらつきがあるのでしょうか？

ここで，X, Y を確率変数，c を定数とするとき，

$$\begin{aligned} E(cX) &= cE(X) \\ V(cX) &= c^2 V(X) \end{aligned} \tag{4}$$

が期待値と分散の定義から明らかに成り立ちます。また，

$$E(X+Y) = E(X) + E(Y) \tag{5}$$

であり，X, Y が独立の時は

$$V(X+Y) = V(X) + V(Y) \tag{6}$$

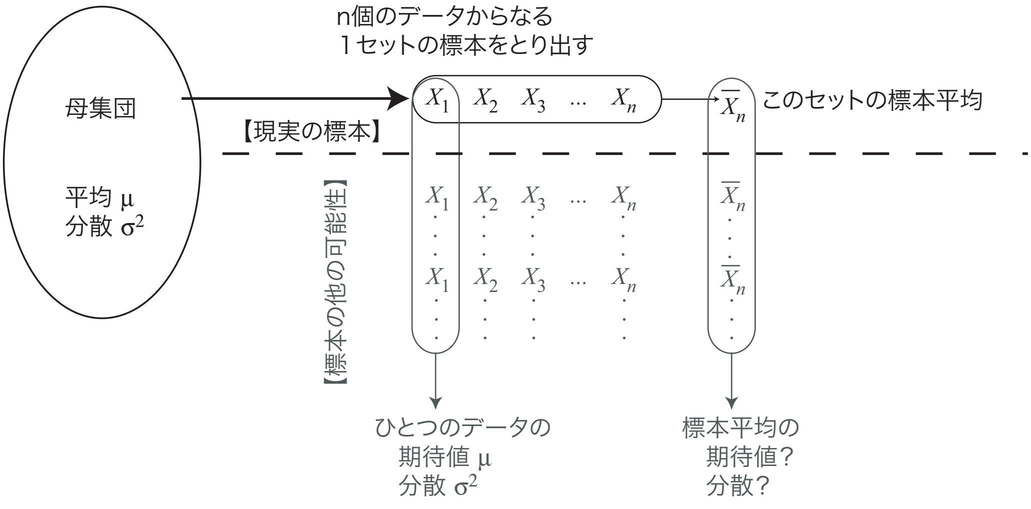

図 2: 標本平均の期待値と分散

も成り立つことが知られています[2]。

標本は，母集団と同じ確率分布にしたがう確率変数で，また標本の各データは互いに独立です。すなわち，$X_1, X_2, \ldots, X_n$ は独立で，各々が平均 μ，分散 σ^2 の確率分布にしたがうので，以上のことから

$$\begin{aligned} E(\overline{X_n}) &= E(\frac{X_1 + X_2 + \cdots + X_n}{n}) \\ &= \frac{1}{n}\{E(X_1) + E(X_2) + \cdots + E(X_n)\} \\ &= \frac{1}{n}(\mu + \mu + \cdots + \mu) = \mu \end{aligned} \tag{7}$$

$$\begin{aligned} V(\overline{X_n}) &= V(\frac{X_1 + X_2 + \cdots + X_n}{n}) \\ &= (\frac{1}{n})^2\{V(X_1) + V(X_2) + \cdots + V(X_n)\} \\ &= \frac{1}{n^2}(\sigma^2 + \sigma^2 + \cdots + \sigma^2) = \frac{\sigma^2}{n} \end{aligned} \tag{8}$$

となります。(7) 式は，標本平均の期待値が母平均と同じであることを表しています。しかし，n が小さい，つまり標本サイズが小さいときは，(8) 式から分散が大きくなります。つまり，期待値（＝母平均）とはかけ離れた値が標本平均となることもしばしばある，ということになります。ところが，n が大きくなると，標本平均の分散は小さくなります。このことは，標本平均は「たいてい，ほぼ」期待値と同じ値，つまり母平均と同じ値になり，期待値からかけ離れた値になる可能性はほとんどない，ということを意味しています。

大数の法則は日常生活でも経験的に知られていることであり，世の中はこのことを承知のうえで回っている，ということもできます。例えば，損害保険の例を考えてみましょう。損害保険会社は，自社の保険加入者が事故にあった場合，保険料に比べてはるかに多額の保険金を支払わなければなりません。自社の保険加入者が事故にあうかどうかは偶然に左右されますから，１年あたりに支払わなければならな

[2]証明には，多次元確率分布の知識が必要です。第１３回の講義で取り扱います。

い保険金額も偶然に左右されます。にもかかわらず，保険会社は，常にほぼ一定額の保険料を受け取って経営を続けています。どうしてそういうことができるのでしょうか？

ある保険会社に保険加入者がひとりしかいないとしましょう。ひとりの保険加入者が事故にあうかどうかは偶然に左右されますから，そのひとりに 1 年間に支払わなければならない保険金の額は確率変数です。ふつう事故にあう確率は小さいですから，保険金額の期待値はそれほど大きくありません。しかし，だからといって保険金額の期待値程度の保険料しか 1 年間に受け取らなかったら，いざ事故が起こったときに保険金を支払うことができません。

ところが，保険加入者がたくさんいて，それぞれが独立に事故にあうとすると，加入者全員が同時に事故にあうなどという事態はほぼ起こりえず，「保険会社からみた，1 年間の支払い保険金額の合計」は，いつもだいたいその期待値程度になるということが，大数の法則からわかります。つまり，各加入者が独立に事故にあうのならば，各加入者から一人当たりの保険金額の期待値程度の保険料を受け取っておけば，事故の時に保険金を支払うことができるというわけです。

チェビシェフの不等式と大数の法則の証明

ところで，ここまでの大数の法則の説明では，ひとつごまかしていることがあります。それは，「標本平均の分散が小さい」ことと，「標本平均が母平均からかけ離れた値になる確率は小さい」ことは本当に同じか，という問題です。これは同じだとはまだ言っていません。ですから，「標本サイズが大きいとき，標本平均が母平均からかけ離れた値になる確率は小さい」ことを示す必要があります。

これが本来の意味での大数の法則で，つぎのようになります。

> $X_1, X_2, \ldots, X_n$ を互いに独立で同じ確率分布にしたがう確率変数とします。その確率変数の期待値 μ，分散 σ^2 がどちらも有限な値であれば，$\overline{X_n} = (X_1 + X_2 + \cdots + X_n)/n$ とするとき，$\overline{X_n}$ は μ に確率収束します。すなわち
>
> $$\lim_{n\to\infty} P(|\overline{X_n} - \mu| \geqq \varepsilon) = 0 \tag{9}$$
>
> がなりたちます[3]。

これは，「『 $\overline{X_n}$ と μ の隔たりがわずかでもある』という確率は，n が大きいと 0 に近づく」ということで，上の「標本サイズが大きいとき，標本平均が母平均からかけ離れた値になる確率は小さい」という記述に対応しています。

この形の大数の法則を証明するために，**チェビシェフ (Chebychev) の不等式**を利用します。これは，期待値と分散の存在するどんな確率変数 X についても，任意の正の数 k について

$$P\left(|X - E(X)| \geqq k\sqrt{V(X)}\right) \leqq \frac{1}{k^2} \tag{10}$$

が成り立つ，というものです。簡単に言えば，上で問題にした「確率変数と期待値との隔たりが，分散に比べて極端に大きくなる確率は，小さい」というのは本当だ，と言っているのです。

[3]確率収束の形の大数の法則は，正式には**大数の弱法則**といいます。他に，概収束の形で表される，すなわち $\overline{X_n} \xrightarrow{\text{a. s.}} \mu \ (n \to \infty)$ という**大数の強法則**があります。

この不等式は次のように証明されます。集合 I を

$$I = \{x : |x - E(X)| \geqq k\sqrt{V(X)}\} \tag{11}$$

とおくと，$F(x)$ を X のしたがう分布関数とするとき

$$\begin{aligned}
V(X) &= \int_R (x - E(X))^2 dF(x) \\
&\geq \int_{x\in I} (x - E(X))^2 dF(x) \\
&\qquad (\{\text{ 被積分関数 }\} \geqq 0 \text{ だから，} \{\text{ 全実数での積分 }\} \geqq \{\text{ 一部での積分 }\}) \\
&\geqq (k\sqrt{V(X)})^2 \int_{x\in I} dF(x) \quad \text{（集合 } I \text{ の定義より）} \\
&= k^2 V(X) P\left(|X - E(X)| \geqq k\sqrt{V(X)}\right) \quad \text{（集合 } I \text{ の定義より）}
\end{aligned} \tag{12}$$

ですから，両辺を $k^2V(X)$ で割ると (10) 式が得られます。

さて，チェビシェフの不等式を，(7)(8) 式の確率変数 $\overline{X_n}$ に適用します。すると，

$$E(\overline{X_n}) = \mu, \quad V(\overline{X_n}) = \frac{\sigma^2}{n} \tag{13}$$

ですから，これを (10) 式に代入すると

$$P\left(|\overline{X_n} - \mu| \geqq k\sqrt{\frac{\sigma^2}{n}}\right) \leqq \frac{1}{k^2} \tag{14}$$

が任意の正の数 k について成り立ちます。ですから，任意の正の数 ε をもってきて

$$k = \frac{\varepsilon}{\sqrt{\frac{\sigma^2}{n}}} = \frac{\sqrt{n}\varepsilon}{\sigma} \tag{15}$$

とおいても (14) 式は成り立つので，(15) 式を (14) 式に代入すると

$$P\left(|\overline{X_n} - \mu| \geqq \frac{\varepsilon}{\sqrt{\frac{\sigma^2}{n}}} \cdot \sqrt{\frac{\sigma^2}{n}}\right) \leqq \frac{1}{\left(\frac{\sqrt{n}\varepsilon}{\sigma}\right)^2} \tag{16}$$

すなわち

$$P\left(|\overline{X_n} - \mu| \geqq \varepsilon\right) \leqq \frac{\sigma^2}{n\varepsilon^2} \tag{17}$$

となり，$n \to \infty$ のとき上の式の右辺 $\to 0$ ですから，(9) 式が成り立つことがわかります。

今日の演習

確率変数の列 X_n について，$X_n \xrightarrow{\mathrm{P}} X$, $X_n \xrightarrow{\mathrm{P}} Y$ $(n \to \infty)$ がなりたつとき，$P(X = Y) = 1$ を証明してください。（ヒント：$|X - Y| = |(X - X_n) - (Y - X_n)| \leqq |X - X_n| + |Y - X_n|$ より，任意の $\varepsilon > 0$ について $P(|X - Y| > \varepsilon) \leqq P(|X - X_n| + |Y - X_n| > \varepsilon)$ であることを用いる）

この定理は，確率収束するときの収束先は確率の意味で一意であることを意味しています。