

2008 年度後期 情報統計学 第 4 回

大数の法則, 確率分布モデル

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, 前回説明した, 度数分布と確率分布の関係です。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る

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ぎで』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, 今日説明する**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合計は, 『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 の上の例のように², 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 の下のようになり, 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, \dots, X_n を取り出し, 標本平均 $\bar{X}_n = (X_1 + X_2 + \dots + X_n)/n$ 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

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。

² この図のグラフは度数分布や確率分布を表すためのグラフで, **ヒストグラム**といいます。横軸に階級をとり, 各階級の (相対) 度数を階級の上に配置した柱の面積で表します。第 6 回の講義で説明します。

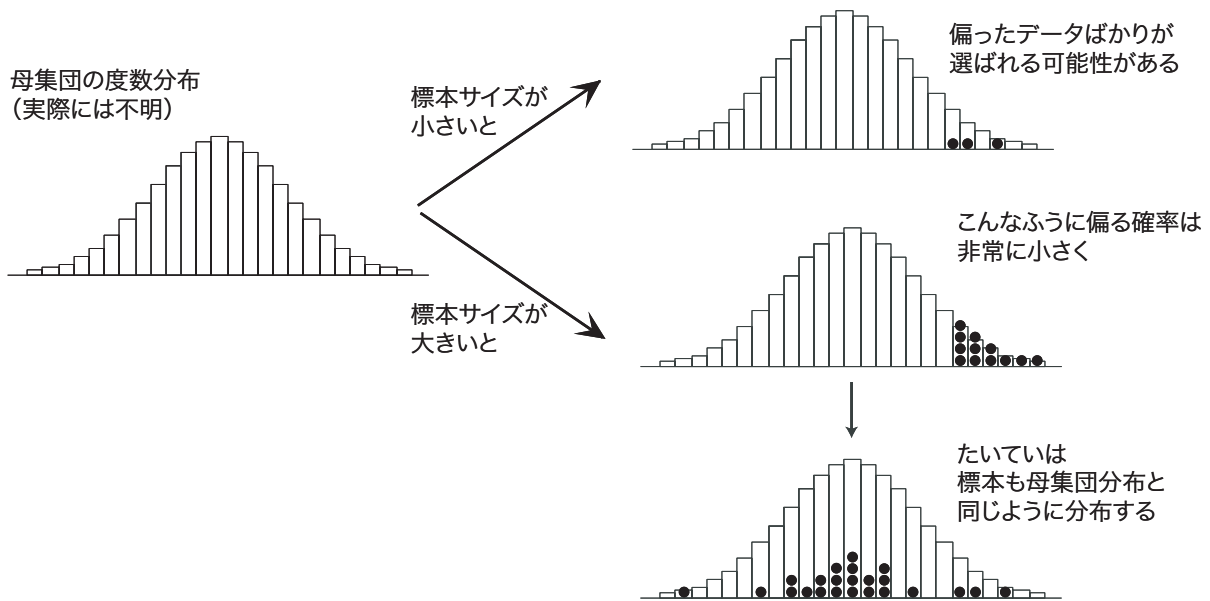


図 1: 標本サイズと標本の分布

な可能性のうちの一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一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 $E(\bar{X}_n)$ と分散 $V(\bar{X}_n)$ はどうなるのでしょうか？ 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 を定数とするととき、

$$\begin{aligned} E(cX) &= cE(X) \\ V(cX) &= c^2V(X) \end{aligned} \tag{1}$$

が期待値と分散の定義（講義第 3 回）から明らかに成り立ちます。また、

$$E(X + Y) = E(X) + E(Y) \tag{2}$$

であり、 X, Y が独立の時は

$$V(X + Y) = V(X) + V(Y) \tag{3}$$

も成り立つことが知られています³。

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

³証明には、多次元確率分布の知識が必要です。第 1 2 回の講義で取り扱います。または、参考書（例えば、講義の案内であげた「基礎統計学 I 統計学入門」）を参照してください。

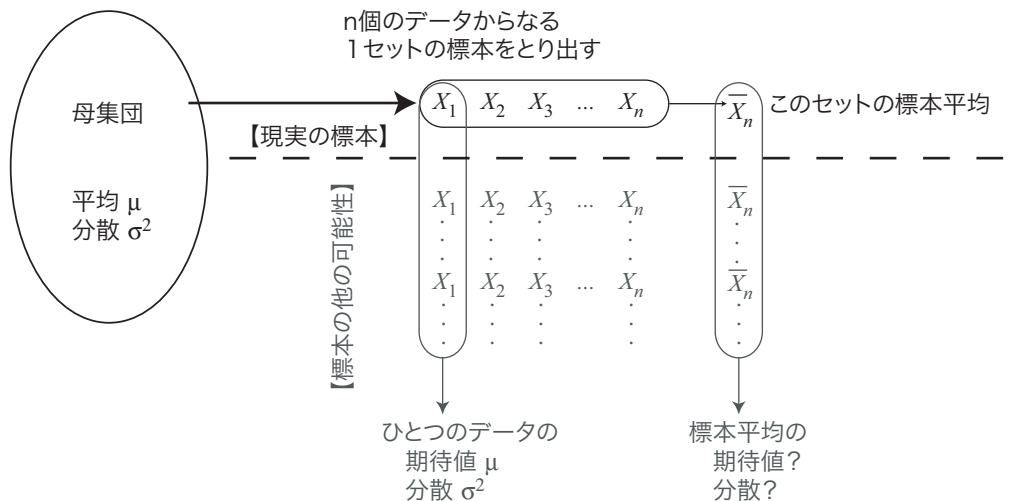


図 2: 標本平均の期待値と分散

わち, X_1, X_2, \dots, 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

$$\begin{aligned}
 E(\bar{X}_n) &= E\left(\frac{X_1 + X_2 + \dots + X_n}{n}\right) \\
 &= \frac{1}{n}\{E(X_1) + E(X_2) + \dots + E(X_n)\} \\
 &= \frac{1}{n}(\mu + \mu + \dots + \mu) = \mu
 \end{aligned} \tag{4}$$

$$\begin{aligned}
 V(\bar{X}_n) &= V\left(\frac{X_1 + X_2 + \dots + X_n}{n}\right) \\
 &= \left(\frac{1}{n}\right)^2\{V(X_1) + V(X_2) + \dots + V(X_n)\} \\
 &= \frac{1}{n^2}(\sigma^2 + \sigma^2 + \dots + \sigma^2) = \frac{\sigma^2}{n}
 \end{aligned} \tag{5}$$

(4) 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, (5) 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, ということの意味しています.

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

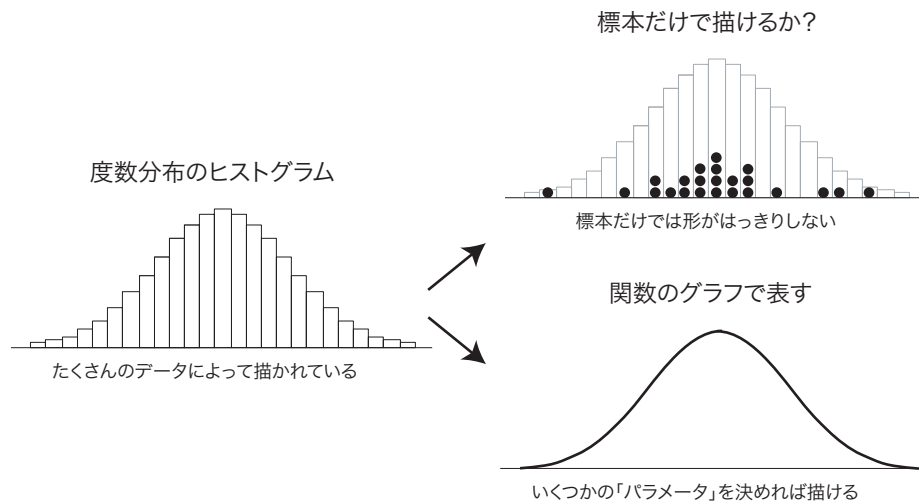


図 3: 確率分布モデル

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つたら、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。

確率分布モデル

上の例では、母平均を推測しましたが、それだけでは「母集団分布を推測した」ことにはならないのはもちろんです。しかし、標本というわずかな手がかりから、母集団分布の全体を完全に誤りなく推測しようというのも、無理な話です。では、母集団分布の何を推測すれば「母集団分布が推測できた」と言えるのでしょうか。

そこで用いるのが、「モデル」の考え方です。この考えでは、母集団分布(=標本がしたがう確率分布)が、ある数式で表されるものだと仮定してしまいます。いわば、母集団分布のヒストグラムが、ある関数のグラフになっていると考えるのです。このようにヒストグラムの形を決めてしまえば、あとはそのグラフの縦横の大きさや位置などを、母集団分布に合うように推測すればよいわけです。この数式を**確率分布モデル**といい、母集団のなりたちに合わせていろいろなものが考えられています。グラフの大きさや位置を決める、これから推測する数値を**パラメータ**といいます。

この講義では、第2部で、さまざまな確率分布モデルの例を説明します。

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」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。大数の法則は、この形で述べるとつぎのようになります。

X_1, X_2, \dots, 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する。その確率分布の期待値を μ とし、 $\bar{X}_n = (X_1 + X_2 + \dots + X_n)/n$ とするとき、任意の正の数 ϵ について⁴

$$\lim_{n \rightarrow \infty} P(|\bar{X}_n - \mu| \geq \epsilon) = 0 \quad (6)$$

が成り立つ。

(6) 式のような収束のしかたを確率収束といいます。これは、「『 \bar{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、**チェビシェフの不等式**を利用します。これは、どんな確率変数 X についても、任意の正の数 k について

$$P(|X - E(X)| \geq k\sqrt{V(X)}) \leq \frac{1}{k^2} \quad (7)$$

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。

この不等式は次のように証明されます。集合 I を

$$I = \{x : |x - E(X)| \geq k\sqrt{V(X)}\} \quad (8)$$

とおくと、 $f(x)$ を X がしたがう確率分布の頻度関数（つまり X が x である確率 $P(X=x)$ ）とすると、

$$\begin{aligned} V(X) &= \sum_x (x - E(X))^2 f(x) \\ &\geq \sum_{x \in I} (x - E(X))^2 f(x) \\ &\quad (\{\sum \text{の中身}\} \geq 0 \text{ だから, } \{\text{すべての } x \text{ についての和}\} \geq \{\text{一部の } x \text{ についての和}\}) \\ &\geq (k\sqrt{V(X)})^2 \sum_{x \in I} f(x) \quad (\text{集合 } I \text{ の定義より}) \\ &= k^2 V(X) P(|X - E(X)| \geq k\sqrt{V(X)}) \quad (\text{集合 } I \text{ の定義より}) \end{aligned} \quad (9)$$

ですから、両辺を $k^2 V(X)$ で割ると (7) 式が得られます。

さて、チェビシェフの不等式を、(4)(5) 式の確率変数 \bar{X}_n に適用します。すると、

$$E(\bar{X}_n) = \mu, \quad V(\bar{X}_n) = \frac{\sigma^2}{n} \quad (10)$$

⁴ 「どんなに小さな正の数 ϵ をもってきて」と読み替えるとわかりやすいと思います。

ですから、これを (7) 式に代入すると

$$P\left(|\bar{X}_n - \mu| \geq k \sqrt{\frac{\sigma^2}{n}}\right) \leq \frac{1}{k^2} \quad (11)$$

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

$$k = \frac{\epsilon}{\sqrt{\frac{\sigma^2}{n}}} = \frac{\sqrt{n}\epsilon}{\sigma} \quad (12)$$

とおいても (11) 式は成り立つので、(12) 式を (11) 式に代入すると

$$P\left(|\bar{X}_n - \mu| \geq \frac{\epsilon}{\sqrt{\frac{\sigma^2}{n}}} \cdot \sqrt{\frac{\sigma^2}{n}}\right) \leq \frac{1}{\left(\frac{\sqrt{n}\epsilon}{\sigma}\right)^2} \quad (13)$$

すなわち

$$P(|\bar{X}_n - \mu| \geq \epsilon) \leq \frac{\sigma^2}{n\epsilon^2} \quad (14)$$

となり、 $n \rightarrow \infty$ のとき上の式の右辺 $\rightarrow 0$ ですから、(6) 式が成り立つことがわかります。