

第 162 講 標本平均の分布

学習目標

この講では、標本平均を新しい確率変数として見る考え方を理解する。

具体的には、次のことができるようになる。

- ・ 標本平均 $\bar{X}$ の期待値・分散・標準偏差を求めることができる。
- ・ 標本平均のばらつきが標本の大きさ n によってどのように変化するか説明できる。
- ・ 正規母集団からの標本平均の分布と、中心極限定理による近似を区別できる。
- ・ 観測後の値 $\bar{x}$ と、抽出前の確率変数 $\bar{X}$ の違いを説明できる。

標本平均は新しい確率変数である

標本平均を考えるときは、抽出する前の状態と、実際に観測した後の状態を区別する必要がある。

以下では、標本の大きさを正整数 n とし、確率変数 $X_1, X_2, \dots, X_n$ が独立同分布に従うとする。

各 X_i の平均を μ 、分散を σ^2 とする。ただし、平均と分散は有限であり、標準偏差については $\sigma \geq 0$ である。

標本平均を、抽出前の確率変数として

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

と定める。

この $\bar{X}$ は、標本値の関数として定まる確率変数である。なお、 $\sigma = 0$ の場合のように確率 1 で一定となる場合も含まれる。

実際に標本を取り出した後、得られた具体的な値を $\bar{x}$ と表す。

つまり、

- ・ $\bar{X}$ は抽出前の確率変数
- ・ $\bar{x}$ は観測後に得られた数値

である。

標本を取り出すたびに $\bar{x}$ が変わることはあるが、「毎回必ず変化する」という意味ではない。同じ値になる場合もある。特に $\sigma = 0$ の場合には、確率 1 で $X_i = \mu$ となり、確率 1 で $\bar{X} = \mu$ となる。

以下で扱う標本や母集団は、数学的に考えるための架空のモデルである。実際の調査では、抽出方法や条件がこのモデルと一致するかを確認する必要がある。

標本平均の期待値

標本平均の期待値とは、確率変数 $\bar{X}$ の分布に対する平均である。

独立同分布に従う確率変数 $X_1, X_2, \dots, X_n$ の平均を μ とすると、 $E(\bar{X}) = \mu$ となる。

これは、標本平均の分布の中心となる値が母平均 μ であることを意味する。

しかし、これは 1 回ごとの観測後の値 $\bar{x}$ が必ず μ になるという意味ではない。

また、 μ は最も多く現れる値であるとは限らず、分布の対称軸を意味するとも限らない。

有限個の可能な値を持つ場合、期待値は各値にその確率を掛けて合計した値として定義される。

したがって、期待値とは確率分布全体から決まる平均的な位置であり、1 回の結果を直接表す値ではない。

標本平均の分散を考える

標本平均

$$\bar{X} = \frac{X_1 + \cdots + X_n}{n}$$

の分散を考える。

まず、期待値の線形性より、 $E(X_1 + \cdots + X_n) = E(X_1) + \cdots + E(X_n)$ である。

また、標本平均の期待値について、

$$\begin{aligned} E(\bar{X}) &= E\left(\frac{1}{n} \sum_{i=1}^n X_i\right) \\ &= \frac{1}{n} \sum_{i=1}^n E(X_i) \\ &= \frac{1}{n} \cdot n\mu \\ &= \mu \end{aligned}$$

となる。

つまり、標本平均の分布の期待値は母平均と一致する。

次に、定数倍の分散について考える。

任意の有限分散を持つ確率変数 U と実定数 c に対して、 $\text{Var}(cU) = c^2 \text{Var}(U)$ である。

これは分散の定義

$\text{Var}(U) = E((U - E(U))^2)$ から、 $\text{Var}(cU) = E((cU - cE(U))^2) = c^2 E((U - E(U))^2)$ となるためである。

ここで、 $Y_i = X_i - \mu$

とおく。

すると、 $E(Y_i) = 0$ である。

和の分散を考えると、 $\text{Var}(X_1 + \cdots + X_n) = \text{Var}(Y_1 + \cdots + Y_n)$ である。

平方を展開すると、

$$(Y_1 + \cdots + Y_n)^2 = \sum_{i=1}^n Y_i^2 + 2 \sum_{1 \leq i < j \leq n} Y_i Y_j$$

となる。

ここで、 $i \neq j$ のとき Y_i と Y_j は独立である。

独立な確率変数 U, V について、有限個の値をとる場合を考える。

U が値 u_r を、 V が値 v_s をとるとする。

ここで、 l, m は正整数であり、 u_r, v_s はそれぞれの可能値、 p_r, q_s はそれぞれの確率である。 U と V が独立ならば、同時に $U = u_r$ 、 $V = v_s$ となる確率は $p_r q_s$ になる。

$r = 1, \dots, l$ 、 $s = 1, \dots, m$ とすると、

$$\begin{aligned} E(UV) &= \sum_{r=1}^l \sum_{s=1}^m u_r v_s p_r q_s \\ &= \left(\sum_{r=1}^l u_r p_r \right) \left(\sum_{s=1}^m v_s q_s \right) \\ &= E(U)E(V) \end{aligned}$$

となる。

一般の有限分散をもつ独立な確率変数でも、この独立な積の期待値が積になる性質を用いる。

したがって、 $E(Y_i Y_j) = E(Y_i)E(Y_j) = 0$ となり、交差項は消える。

各 Y_i は $Y_i = X_i - \mu$

で定義されている。

したがって、 $E(Y_i) = E(X_i - \mu) = E(X_i) - \mu = \mu - \mu = 0$ となる。

その結果、和についても、 $E(Y_1 + \dots + Y_n) = E(Y_1) + \dots + E(Y_n) = 0$ となる。

平方を展開すると、

$$\begin{aligned}\text{Var}(X_1 + \dots + X_n) &= E((Y_1 + \dots + Y_n)^2) \\ &= E\left(\sum_{i=1}^n Y_i^2 + 2 \sum_{1 \leq i < j \leq n} Y_i Y_j\right) \\ &= \sum_{i=1}^n E(Y_i^2) + 2 \sum_{1 \leq i < j \leq n} E(Y_i Y_j)\end{aligned}$$

である。

各 Y_i について、 $E(Y_i^2) = \sigma^2$ である。

また、 $i \neq j$ のとき独立性より、 $E(Y_i Y_j) = E(Y_i)E(Y_j) = 0$ となる。

したがって、

$$\begin{aligned}\text{Var}(X_1 + \dots + X_n) &= \sum_{i=1}^n E(Y_i^2) + 2 \sum_{1 \leq i < j \leq n} E(Y_i Y_j) \\ &= \sum_{i=1}^n \sigma^2 + 0 \\ &= n\sigma^2\end{aligned}$$

となる。

したがって、

$$\begin{aligned}\text{Var}(\bar{X}) &= \text{Var}\left(\frac{X_1 + \dots + X_n}{n}\right) \\ &= \frac{1}{n^2} \text{Var}(X_1 + \dots + X_n) \\ &= \frac{1}{n^2} n\sigma^2 \\ &= \frac{\sigma^2}{n}\end{aligned}$$

となる。

以上より、 $E(\bar{X}) = \mu$

$$\begin{aligned}\text{Var}(\bar{X}) &= \frac{\sigma^2}{n} \\ \text{SD}(\bar{X}) &= \frac{\sigma}{\sqrt{n}}\end{aligned}$$

である。

この標準偏差は標本平均の標準誤差とも呼ばれるが、標本平均と母平均との差の上限を表すものではない。

標本平均の分布で覚える関係

独立同分布に従う確率変数 $X_1, \dots, X_n$ について、平均 μ と分散 σ^2 が有限であるとする。
標本平均を

$$\bar{X} = \frac{X_1 + \dots + X_n}{n}$$

とすると、 $E(\bar{X}) = \mu$ である。

また、分散を $\text{Var}(\bar{X})$ 、標準偏差を $\text{SD}(\bar{X})$ と書くと、

$$\text{Var}(\bar{X}) = \frac{\sigma^2}{n}$$

$$\text{SD}(\bar{X}) = \frac{\sigma}{\sqrt{n}}$$

となる。

標本の大きさ n を 4 倍にすると、

$$\frac{\sigma^2}{4n} = \frac{1}{4} \frac{\sigma^2}{n}$$

より、分散は 4 分の 1 になる。

一方、

$$\frac{\sigma}{\sqrt{4n}} = \frac{1}{2} \frac{\sigma}{\sqrt{n}}$$

より、標準偏差は 2 分の 1 になる。

したがって、標本の大きさを 4 倍にすると、標本平均の分散は 4 分の 1、標準偏差は 2 分の 1 になる。

抽出方法に注意する

有限母集団から標本を取り出す場合、独立な復元抽出を基本モデルとして考えることが多い。

非復元抽出では、同じ標本内の値は一般には独立とはみなせない。

ただし、特殊な場合には独立になることもあるため、「非復元抽出なら必ず独立ではない」と断定することはできない。

独立性がない場合、独立な復元抽出を前提として導いた

$$\text{Var}(\bar{X}) = \frac{\sigma^2}{n}$$

を無条件に適用してはいけない。

また、 $\sigma = 0$ の場合、確率 1 で

$X_i = \mu$ である。

したがって、 $\bar{X} = \mu$ となり、 $\text{Var}(\bar{X}) = 0$ である。

このとき、

$$\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

は分母が 0 になるため未定義であり、標準化できない。

正規性がある場合と近似の場合

母集団が正規分布

$$N(\mu, \sigma^2)$$

に従い、 $\sigma > 0$ であるとする。

このとき、独立に取り出した標本平均は任意の n について正規分布になる。

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

これは正規分布の閉性による性質として用いる。

したがって、

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

は標準正規分布に従う。

一方、母集団が正規分布とは限らない場合でも、平均と分散が有限で、分散が正である独立同分布の場合には、大きい n で標準化した標本平均の分布は標準正規分布で近似できる。

これが中心極限定理である。

ただし、これは近似であり、正規分布であると厳密に断定するものではない。

また、 $n = 30$ なら必ず十分というような普遍的な条件や精度保証はない。

中心極限定理によって、元の観測値そのものが正規分布になるわけでもない。

例題 1

確率変数 X が 0 と 2 をそれぞれ確率 $\frac{1}{2}$ でとる架空の母集団を考える。

この母集団から独立な復元抽出を 2 回行い、

$$\bar{X} = \frac{X_1 + X_2}{2}$$

とする。

(1) $\bar{X}$ の分布を求めなさい。

(2) $\bar{X}$ の平均、分散、標準偏差を求めなさい。

(3) 母平均と母分散から求めた公式と一致することを確認しなさい。

例題 1 の解答・解説

2 回の抽出結果は順序を区別すると、

結果	確率
(0, 0)	$\frac{1}{4}$
(0, 2)	$\frac{1}{4}$
(2, 0)	$\frac{1}{4}$
(2, 2)	$\frac{1}{4}$

である。

それぞれについて標本平均を求める。

(X_1, X_2)	$\bar{X}$	確率
(0, 0)	0	$\frac{1}{4}$
(0, 2)	1	$\frac{1}{4}$

(X_1, X_2)	$\bar{X}$	確率
$(2, 0)$	1	$\frac{1}{4}$
$(2, 2)$	2	$\frac{1}{4}$

したがって、

$\bar{X}$	確率
0	$\frac{1}{4}$
1	$\frac{1}{2}$
2	$\frac{1}{4}$

となる。

平均は定義から、

$$E(\bar{X}) = 0 \cdot \frac{1}{4} + 1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{4} = 1$$

である。

分散は、

$$\text{Var}(\bar{X}) = (0 - 1)^2 \frac{1}{4} + (1 - 1)^2 \frac{1}{2} + (2 - 1)^2 \frac{1}{4} = \frac{1}{2}$$

となる。

したがって標準偏差は、

$$\sqrt{\frac{1}{2}} = \frac{1}{\sqrt{2}}$$

である。

母集団について、 $E(X) = 1$ であり、

$$\text{Var}(X) = (0 - 1)^2 \frac{1}{2} + (2 - 1)^2 \frac{1}{2} = 1$$

である。

したがって公式より、 $E(\bar{X}) = 1$

$$\text{Var}(\bar{X}) = \frac{1}{2}$$

となり、定義から求めた結果と一致する。

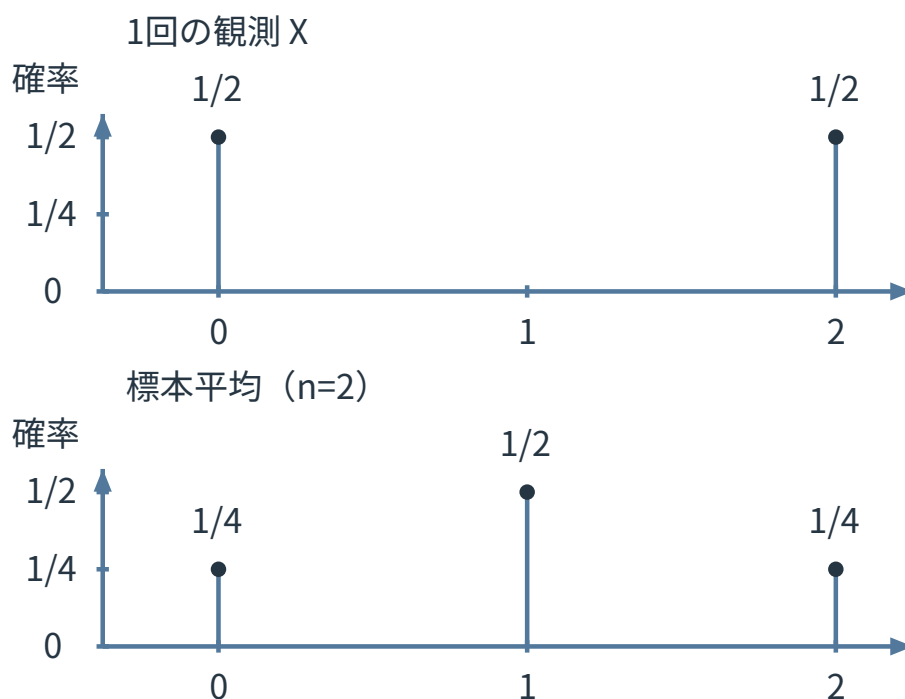


図1 上は0と2を等確率でとる1回の観測 X 、下は独立な復元抽出2回の標本平均である。横軸の値と縦軸の確率は上下で同じ尺度にそろえている。平均はいずれも1で、標本平均の分布では平均に近い値1へ確率が集まり、ばらつきが小さくなる。棒の高さが各値の確率を表し、区間の面積としては読まない。

例題 2

独立な確率変数

$X_i \sim N(50, 100)$ を25個取り出し、標本平均を $\bar{X}$ とする。

(1) $\bar{X}$ の分布を求めなさい。

(2) 標本平均の標準偏差を求めなさい。

(3) $P(48 \leq \bar{X} \leq 52)$ を求めなさい。

ただし、 $A(1) = P(0 \leq Z \leq 1) \approx 0.3413$ を用いてよい。

例題 2 の解答・解説

母分布は正規分布であるため、正規分布の閉性により、標本平均の分布は厳密に正規分布になる。

ここで、 $\mu = 50$, $\sigma^2 = 100$ より、 $\sigma = 10$ である。

標本の大きさは $n = 25$ なので、

$$\bar{X} \sim N\left(50, \frac{100}{25}\right)$$

すなわち、 $\bar{X} \sim N(50, 4)$ となる。

標本平均の標準偏差は、

$$\frac{\sigma}{\sqrt{n}} = \frac{10}{\sqrt{25}} = 2$$

である。
標準化すると、

$$Z = \frac{\bar{X} - 50}{2}$$

となる。
したがって、 $48 \leq \bar{X} \leq 52$ は、 $-1 \leq Z \leq 1$ に対応する。
よって、

$$\begin{aligned} P(48 \leq \bar{X} \leq 52) &= P(-1 \leq Z \leq 1) \\ &= 2P(0 \leq Z \leq 1) \\ &\approx 2 \times 0.3413 \\ &\approx 0.6826 \end{aligned}$$

である。
したがって、 $P(48 \leq \bar{X} \leq 52) \approx 0.6826$ となる。
正規分布になることは厳密であり、0.3413 や 0.6826 の値は表の値を用いた近似である。

例題 3

母平均が 20、母標準偏差が 6 である母集団から、独立同分布に標本を取り出す。
標本の大きさが $n = 9$ の場合と $n = 36$ の場合について、標本平均の平均、分散、標準偏差を求めなさい。
また、標本の大きさを増やすと何が変化するか説明しなさい。
ただし、母集団は正規分布であるとは限らない。

例題 3 の解答・解説

母平均は

$$\mu = 20$$

母標準偏差は $\sigma = 6$ なので、母分散は $\sigma^2 = 36$ である。

標本平均の平均はどちらの場合も、 $E(\bar{X}) = 20$ である。

$n = 9$ の場合、

$$\text{Var}(\bar{X}) = \frac{36}{9} = 4$$

標準偏差は、 $\sqrt{4} = 2$ である。

$n = 36$ の場合、

$$\text{Var}(\bar{X}) = \frac{36}{36} = 1$$

標準偏差は、1 である。

標本の大きさを 4 倍にすると、標本平均の標準偏差は半分になる。

ただし、これは分布のばらつきが小さくなることを意味するだけであり、毎回の標本平均が必ず母平均に近づく保証ではない。

また、母集団が正規分布であるという条件がないため、標本平均の分布の形を正規分布と断定することはできない。

演習問題

練習問題 1

独立同分布に従う確率変数 $X_1, \dots, X_n$ を考える。

母平均が 12、母標準偏差が 6、標本の大きさが $n = 9$ であるとする。

標本平均 $\bar{X}$ の期待値、分散、標準偏差を求めなさい。

ただし、母集団が正規分布に従うとは限らない。

練習問題 1 の解答・解説

母平均は

$$\mu = 12$$

母標準偏差は $\sigma = 6$ なので、母分散は $\sigma^2 = 36$ である。

標本平均の期待値は、 $E(\bar{X}) = \mu$ より、 $E(\bar{X}) = 12$ である。

分散は、

$$\text{Var}(\bar{X}) = \frac{\sigma^2}{n}$$

より、

$$\text{Var}(\bar{X}) = \frac{36}{9} = 4$$

である。

標準偏差は、

$$\frac{\sigma}{\sqrt{n}} = \frac{6}{\sqrt{9}} = 2$$

となる。

したがって、 $E(\bar{X}) = 12$

$$\text{Var}(\bar{X}) = 4$$

$\text{SD}(\bar{X}) = 2$ である。

ただし、母集団が正規分布であるという条件はないため、標本平均の分布の形を正規分布と断定することはできない。

練習問題 2

独立な確率変数

$$X_i \sim N(30, 16)$$

について、標本の大きさを $n = 16$ とする。

(1) 標本平均 $\bar{X}$ の分布を求めなさい。

(2) $P(29 \leq \bar{X} \leq 31)$ を求めなさい。

ただし、 $A(1) = P(0 \leq Z \leq 1) \approx 0.3413$ を用いてよい。

練習問題 2 の解答・解説

母分布は正規分布であるため、標本平均も正規分布になる。

ここで、 $\mu = 30$

$\sigma^2 = 16$ より、 $\sigma = 4$ である。

標本の大きさは $n = 16$ なので、

$$\bar{X} \sim N\left(30, \frac{16}{16}\right)$$

したがって、 $\bar{X} \sim N(30, 1)$ である。

標準偏差は、

$$\frac{\sigma}{\sqrt{n}} = \frac{4}{\sqrt{16}} = 1$$

である。
標準化すると、

$$Z = \frac{\bar{X} - 30}{1}$$

となる。

よって、 $29 \leq \bar{X} \leq 31$ は、 $-1 \leq Z \leq 1$ に対応する。

したがって、 $P(29 \leq \bar{X} \leq 31) = P(-1 \leq Z \leq 1)$ である。

標準正規分布の対称性より、

$$\begin{aligned} P(29 \leq \bar{X} \leq 31) &= P(-1 \leq Z \leq 1) \\ &= 2A(1) \\ &\approx 2 \times 0.3413 \\ &\approx 0.6826 \end{aligned}$$

したがって、 $P(29 \leq \bar{X} \leq 31) \approx 0.6826$ である。

正規分布になることは母集団が正規分布であるため厳密であり、0.3413 という値は正規分布表による近似値である。

練習問題 3

独立同分布に従う確率変数 $X_1, \dots, X_n$ について、標本の大きさを $n = 1$ とする。

(1) $\bar{X} = X_1$ となることを示しなさい。

(2) $\bar{X}$ の平均、分散、標準偏差を答えなさい。

(3) 元の母集団が正規分布でない場合、標本平均が正規分布になるとはいえない理由を説明しなさい。

練習問題 3 の解答・解説

標本平均は、

$$\bar{X} = \frac{X_1 + \dots + X_n}{n}$$

で定義される。

$n = 1$ のとき、

$$\bar{X} = \frac{X_1}{1} = X_1$$

となる。

したがって、標本平均は元の確率変数そのものである。

母平均を μ 、母分散を σ^2 とすると、 $E(\bar{X}) = E(X_1) = \mu$ である。

また、 $\text{Var}(\bar{X}) = \text{Var}(X_1) = \sigma^2$ である。

標準偏差は、 σ である。

つまり、標本の大きさが 1 のとき、標本平均の分布は元の母集団の分布そのものになる。

$n = 1$ のとき、標本平均は元の確率変数そのものであるため、標本平均の分布は元の母集団の分布そのものである。したがって、母集団が正規分布ではないと仮定したなら、標本平均も正規分布ではない。

練習問題 4

独立な確率変数

$$X_i \sim N(100, 64)$$

について、標本の大きさを $n = 64$ とする。

次の確率を求めなさい。

$$P(\bar{X} \geq 102)$$

ただし、 $A(2) = P(0 \leq Z \leq 2) \approx 0.4772$ を用いてよい。

練習問題 4 の解答・解説

母平均は、

$$\mu = 100$$

母分散は、 $\sigma^2 = 64$ なので、 $\sigma = 8$ である。

標本平均の標準偏差は、

$$\frac{\sigma}{\sqrt{n}} = \frac{8}{\sqrt{64}} = 1$$

である。

したがって、 $\bar{X} \sim N(100, 1)$ となる。

標準化すると、

$$Z = \frac{\bar{X} - 100}{1}$$

である。

よって、 $P(\bar{X} \geq 102) = P(Z \geq 2)$ となる。

標準正規分布の対称性より、

$$\begin{aligned} P(\bar{X} \geq 102) &= P(Z \geq 2) \\ &= \frac{1}{2} - A(2) \\ &\approx 0.5 - 0.4772 \\ &\approx 0.0228 \end{aligned}$$

したがって、 $P(\bar{X} \geq 102) \approx 0.0228$ である。

練習問題 5

母標準偏差が 3 の母集団から独立同分布に標本を取り出す。

標本の大きさが

$n = 4$ の場合と

$n = 16$ の場合について、標本平均の分散と標準偏差をそれぞれ求め、比較しなさい。

練習問題 5 の解答・解説

母標準偏差は、

$\sigma = 3$ なので、母分散は、 $\sigma^2 = 9$ である。

まず $n = 4$ の場合、

$$\text{Var}(\bar{X}) = \frac{9}{4}$$

標準偏差は、

$$\frac{3}{\sqrt{4}} = \frac{3}{2}$$

である。

次に $n = 16$ の場合、

$$\text{Var}(\bar{X}) = \frac{9}{16}$$

標準偏差は、

$$\frac{3}{\sqrt{16}} = \frac{3}{4}$$

である。

分散を比較すると、

$$\frac{9}{16} = \frac{1}{4} \times \frac{9}{4}$$

なので、標本の大きさを 4 倍にすると分散は 4 分の 1 になる。

一方、標準偏差は、

$$\frac{3}{4} = \frac{1}{2} \times \frac{3}{2}$$

なので、標本の大きさを 4 倍にすると 2 分の 1 になる。

分散と標準偏差では変化の割合が異なることに注意する。

練習問題 6

確率変数 X が

$P(X = 7) = 1$ を満たすとする。

この母集団から独立に標本を取り、標本の大きさを $n = 5$ とする。

標本平均 $\bar{X}$ の期待値、分散、標準偏差を求めなさい。

また、

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

による標準化ができない理由を説明しなさい。

練習問題 6 の解答・解説

確率 1 で

$X = 7$ なので、標本を何回取り出してもすべての値は 7 である。

したがって、 $P(\bar{X} = 7) = 1$ である。

よって、 $E(\bar{X}) = 7$ となる。

また、標本平均は常に 7 なので、 $\text{Var}(\bar{X}) = 0$ である。

標準偏差は、 $\text{SD}(\bar{X}) = 0$ である。

この場合、母標準偏差は $\sigma = 0$ である。

したがって、標準化の式

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

では分母が

$$\frac{\sigma}{\sqrt{n}} = 0$$

となる。

そのため、 Z は未定義であり、標準化を行うことはできない。

練習問題 7

値 0 と 2 を持つ 2 つの対象から、非復元抽出で標本を取り出す。

1 回目は 2 対象から等確率で 1 つを選び、2 回目は残った 1 対象を選ぶ無作為な非復元抽出とする。

標本の大きさを $n = 2$ とし、標本平均 $\bar{X}$ を考える。

(1) 母平均と母分散を求めなさい。

(2) 標本平均 $\bar{X}$ の分散を求めなさい。

(3) 独立な復元抽出の場合の公式

$$\text{Var}(\bar{X}) = \frac{\sigma^2}{n}$$

をそのまま使ってはいけない理由を、この例に即して説明しなさい。

練習問題 7 の解答・解説

母集団の値は 0, 2 である。

母平均は、

$$\mu = \frac{0 + 2}{2} = 1$$

である。

母分散は、

$$\sigma^2 = (0 - 1)^2 \frac{1}{2} + (2 - 1)^2 \frac{1}{2} = 1$$

である。

非復元抽出では、必ず 0 と 2 を 1 回ずつ選ぶ。

したがって、

$$\overline{X} = \frac{0+2}{2} = 1$$

であり、毎回一定である。

よって、 $\text{Var}(\overline{X}) = 0$ である。

一方、独立な復元抽出を仮定した公式を使うと、

$$\frac{\sigma^2}{n} = \frac{1}{2}$$

となる。

この例では、1 回目に選んだ値を X_1 、2 回目に選んだ値を X_2 とすると、 $X_2 = 2 - X_1$ となる。

1 回目は 2 対象から等確率で選ぶため、

$$P(X_1 = 0) = P(X_2 = 0) = \frac{1}{2}$$

である。

しかし、両方が 0 になることはないため、 $P(X_1 = 0, X_2 = 0) = 0$ である。

独立ならば、

$$P(X_1 = 0, X_2 = 0) = P(X_1 = 0)P(X_2 = 0) = \frac{1}{4}$$

となるはずである。

実際には一致しないため、この例では X_1 と X_2 は独立ではない。

したがって、独立な復元抽出を前提として導いた分散公式

$$\text{Var}(\overline{X}) = \frac{\sigma^2}{n}$$

を、この非復元抽出へそのまま適用することはできない。

したがって、非復元抽出へ独立復元抽出の場合の分散公式を無条件に使ってはいけない。

練習問題 8

次の 2 つの主張について、正しいか誤っているかを判定し、理由を説明しなさい。

- (1) 正規分布でない母集団でも、標本の大きさが $n \geq 30$ なら標本平均は必ず厳密な正規分布になる。
- (2) 中心極限定理によって、元の確率変数 X_i 自体が正規分布になる。

練習問題 8 の解答・解説

(1) は誤りである。

中心極限定理では、独立同分布で、平均が有限、分散が有限かつ正である場合、大きな標本の大き
きで標準化した標本平均の分布を標準正規分布で近似できる。

しかし、これは近似であり、厳密に正規分布になるという意味ではない。

また、 $n \geq 30$ のような条件は普遍的な十分条件ではなく、精度保証でもない。

したがって、「必ず厳密な正規分布になる」という部分が誤りである。

(2) も誤りである。

中心極限定理が述べているのは、標本平均の分布についてである。

元の確率変数 X_i の分布が正規分布になるわけではない。

例えば、元の母集団が左右非対称な分布であっても、標本平均を標準化した分布が大きな標本の
大ききで標準正規分布に近づく場合がある。

したがって、中心極限定理は元の観測値ではなく、標本平均の分布に関する性質である。