第4章　標本分布論と漸近分布論

4.1　標本分布論の枠組み

4.1.1　ランダム標本

定義 4.1　$X_1, X_2, \ldots, X_n$ が母集団分布 $f(x)$ からの標本の大きさが n の/**ランダム標本**であるとは，$X_1, X_2, \ldots, X_n$ は互いに独立な確率変数列であって，各 $X_i, i = 1, 2, \ldots, n$ は確率関数または確率密度関数 $f(x)$ に従うときをいう．また，$X_1, X_2, \ldots, X_n$ は独立同一に確率関数または確率密度関数 $f(x)$ に従うともいう．

注意 4.1　多くの場合は $n > 1$ である．また，$\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ の同時確率関数または同時確率密度関数は

$$f_{\boldsymbol{X}}(x_1, x_2, \ldots, x_n) = \prod_{k=1}^{n} f(x_i)$$

である．

4.1.2　統計量と標本分布

定義 4.2　$X_1, X_2, \ldots, X_n$ をある母集団分布からの標本の大きさが n のランダム標本とし，$T(x_1, x_2, \ldots, x_n)$ をランダム標本 $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ の値域上で定義された実数値または実ベクトル値関数とする．このとき，確率変数または確率ベクトル $Y = T(X_1, X_2, \ldots, X_n)$ を**統計量**という．さらに，統計量 Y の確率分布を Y の**標本分布**とよぶ．

例 4.1　ランダム標本の算術平均は統計量であり，**標本平均**という．通常

$$\bar{X}_n = \frac{X_1 + X_2 + \cdots + X_n}{n} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

と記す.

　また,

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$$

で定義される統計量を**標本分散**という.

補題 4.1 $x_1, x_2, \ldots, x_n$ を実数列とし，$\bar{x}_n = (1/n)(x_1 + x_2 + \cdots + x_n)$ と $s_n^2 = (1/(n-1))\sum_{i=1}^n (x_i - \bar{x}_n)^2$ とおく．このとき，

(1) $\min_a \sum_{i=1}^n (x_i - a)^2 = \sum_{i=1}^n (x_i - \bar{x}_n)^2$.

(2) $(n-1)s_n^2 = \sum_{i=1}^n x_i^2 - n\bar{x}_n^2$.

となる．

証明 (1) を証明するために，$\bar{x}_n$ を加えて引けば，

$$\begin{aligned}\sum_{i=1}^n (x_i - a)^2 &= \sum_{i=1}^n (x_i - \bar{x}_n + \bar{x}_n - a)^2 \\ &= \sum_{i=1}^n (x_i - \bar{x}_n)^2 + 2\sum_{i=1}^n (x_i - \bar{x}_n)(\bar{x}_n - a) + \sum_{i=1}^n (\bar{x}_n - a)^2 \\ &= \sum_{i=1}^n (x_i - \bar{x}_n)^2 + n(\bar{x}_n - a)^2 \qquad (4.1)\end{aligned}$$

となる．最後の等式は

$$\sum_{i=1}^n (x_i - \bar{x}_n)(\bar{x}_n - a) = (\bar{x}_n - a)\sum_{i=1}^n (x_i - \bar{x}_n) = 0$$

よりわかる．(4.1) は $a = \bar{x}_n$ の時に最小になることがわかる．
(2) を示すためには，(4.1) において，$a = 0$ とすればよい． □

定理 4.1 $X_1, X_2, \ldots, X_n$ をある母集団分布からの標本の大きさが n のランダム標本とし，g を X_1 の値域上で定義された実数値関数とする． $g^2(X_2)$ の期待値が存在するとき，

$$\mathbb{E}[\sum_{i=1}^n g(X_i)] = n\mathbb{E}[g(X_1)], \qquad (4.2)$$

$$\mathbb{VAR}[\sum_{i=1}^n g(X_i)] = n\mathbb{VAR}[g(X_1)] \qquad (4.3)$$

が成立する．

証明 (4.2) は期待値の線形性と

$$\mathbb{E}[\sum_{i=1}^n g(X_i)] = \sum_{i=1}^n \mathbb{E}[g(X_i)] = n\mathbb{E}[g(X_1)]$$

からわかる[1].

(4.3) を示すために，分散の定義と期待値の線形性から

$$
\begin{aligned}
\mathbb{VAR}[\sum_{i=1}^{n} g(X_i)] &= \mathbb{E}[\{\sum_{i=1}^{n} g(X_i) - \mathbb{E}[\sum_{i=1}^{n} g(X_i)]\}^2] \\
&= \mathbb{E}[\{\sum_{i=1}^{n}(g(X_i) - \mathbb{E}[g(X_i)])\}^2] \\
&= \mathbb{E}\left[\sum_{i=1}^{n}(g(X_i) - \mathbb{E}[g(X_i)])^2 + \sum_{i\neq j} g(X_i) - \mathbb{E}[g(X_i)])g(X_j) - \mathbb{E}[g(X_j)])\right] \\
&= \sum_{i=1}^{n} \mathbb{E}[(g(X_i) - \mathbb{E}[g(X_i)])^2] \\
&\quad + \sum_{i\neq j} \mathbb{E}[g(X_i) - \mathbb{E}[g(X_i)])g(X_j) - \mathbb{E}[g(X_j)])] \qquad (4.4)
\end{aligned}
$$

となる．しかし，

$$
\sum_{i=1}^{n} \mathbb{E}[(g(X_i) - \mathbb{E}[g(X_i)])^2] = \sum_{i=1}^{n} \mathbb{VAR}[g(X_i)] = n\mathbb{VAR}[g(X_1)]
$$

と $i \neq j$ に対して，

$$
\mathbb{E}[g(X_i) - \mathbb{E}[g(X_i)])(g(X_j) - \mathbb{E}[g(X_j)])] = \mathbb{E}[g(X_i)g(X_j)] - \mathbb{E}[g(X_i)]\mathbb{E}[g(X_j)] = 0
$$

となる．ただし，最後の等号は定理 3.2 からわかる．上のふたつの式を (4.4) に代入すれば，(2) は示される． □

系 4.1 $X_1, X_2, \ldots, X_n$ をある母集団分布からの標本の大きさが n のランダム標本とする．X_1^2 の期待値が存在するとき，

$$
\begin{aligned}
\mathbb{E}[\sum_{i=1}^{n} a_i X_i)] &= \sum_{i=1}^{n} a_i \mathbb{E}[X_i)], \\
\mathbb{VAR}[\sum_{i=1}^{n} a_i X_i] &= \sum_{i=1}^{n} a_i^2 \mathbb{VAR}[X_i]
\end{aligned}
$$

[1]連続型の場合を考える．$X_1, X_2, \ldots, X_n$ の同時確率密度関数を $f_{X_1, X_2, \ldots, X_n}(x_1, x_2, \ldots, x_n)$ としたとき，

$$
\begin{aligned}
\mathbb{E}[\sum_{i=1}^{n} g(X_i)] &= \int \cdots \int \sum_{i=1}^{n} g(x_i) f_{X_1, X_2, \ldots, X_n}(x_1, x_2, \ldots, x_n) \Pi_{i=1}^{n} dx_i \\
&= \sum_{i=1}^{n} \int \cdots \int g(x_i) f_{X_1, X_2, \ldots, X_n}(x_1, x_2, \ldots, x_n) \Pi_{i=1}^{n} dx_i \\
&= \sum_{i=1}^{n} \int g(x_i) f_{X_i}(x_i) dx_i \\
&= \mathbb{E}[g(X_i)]
\end{aligned}
$$

よりわかる．ただし，$f_{X_i}(x_i)$ は X_i の周辺確率密度関数である．

が成立する．ただし，$a_1, a_2, \ldots, a_n$ は定数である．

定理 4.2 $X_1, X_2, \ldots, X_n$ を平均 μ と分散 $\sigma^2 < \infty$ の母集団分布からの標本の大きさが n のランダム標本とする．このとき，

(1) $\mathbb{E}[\bar{X}_n] = \mu,$

(2) $\mathbb{VAR}[\bar{X}_n] = \dfrac{\sigma^2}{n},$

(3) $\mathbb{E}[S_n^2] = \sigma^2.$

となる．

証明 (1) を示すために，定理 において，$g(x) = x$ とすれば，

$$\mathbb{E}\left[\bar{X}_n\right] = \frac{1}{n}\mathbb{E}[\sum_{i=1}^{n} X_i] = \frac{1}{n}n\mathbb{E}[X_1] = \mathbb{E}[X_1] = \mu$$

となることがわかる．

(2) は分散の性質と定理 を同様に利用すれば，

$$\mathbb{VAR}[\bar{X}_n] = \frac{1}{n^2}\mathbb{VAR}[\sum_{i=1}^{n} X_i] = \frac{1}{n^2}n\mathbb{VAR}[X_1] = \frac{1}{n}\mathbb{VAR}[X_1] = \frac{\sigma^2}{n}$$

となることがわかる．

(3) を示すために補題 4.1 を使えば，

$$\begin{aligned} \mathbb{E}[S_n^2] &= \mathbb{E}\left[\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2\right] = \frac{1}{n-1}\mathbb{E}\left[\sum_{i=1}^{n} X_i^2 - n\bar{X}_n^2\right] \\ &= \frac{1}{n-1}\left(n\mathbb{E}[X_1^2] - n\mathbb{E}[\bar{X}_n^2]\right) \end{aligned} \tag{4.5}$$

となる．しかし，

$$\begin{aligned} \mathbb{E}[X_1^2] &= \mathbb{VAR}[X_1] + (\mathbb{E}[X_1])^2 = \sigma^2 + \mu^2, \\ \mathbb{E}[\bar{X}_n^2] &= \mathbb{VAR}[\bar{X}_n] + (\mathbb{E}[\bar{X}_n])^2 = \frac{\sigma^2}{n} + \mu^2 \end{aligned}$$

となる．これらと (4.5) をあわせれば，

$$\mathbb{E}[S_n^2] = \frac{1}{n-1}\left(n(\sigma^2 + \mu^2 - n\left(\frac{\sigma^2}{n} + \mu^2\right)\right) = \sigma^2$$

となることがわかる． □

定理 4.3 $X_1, X_2, \ldots, X_n$ を積率母関数 $M_X(t)$ を持つ母集団分布からの標本の大きさが n のランダム標本とする．このとき，

$$M_{\bar{X}_n}(t) = (M_X(t/n))^n$$

が成立する．

証明

$$M_{\bar{X}_n}(t) = \mathbb{E}[e^{t\bar{X}_n}] = \mathbb{E}\left[\prod_{i=1}^n e^{tX_i/n}\right] = \prod_{i=1}^n \mathbb{E}\left[e^{tX_i/n}\right] = [M_X(t/n)]^n$$

からわかる． □

例 4.2 $X_1, X_2, \ldots, X_n$ を正規分布 $N(\mu, \sigma^2)$ からの標本の大きさが n のランダム標本とする．このとき，標本平均 $\bar{X}_n$ は正規分布 $N(\mu, \sigma^2/n)$ に従うことがわかる．なぜならば，

$$\begin{aligned} M_{\bar{X}_n}(t) &= \left[pe^{t/n} + (1-p)\right]^n = \exp\left[n\left(\frac{t}{n}\mu + \frac{1}{2}\left(\frac{t}{n}\right)^2\sigma^2\right)\right] \\ &= \exp\left[t\mu + \frac{(\sigma^2/n)t^2}{2}\right] \end{aligned}$$

からわかる．

例 4.3 $X_1, X_2, \ldots, X_n$ を母数 p のベルヌーイ分布からの標本の大きさが n のランダム標本とする．ただし，$0 < p < 1$ である．このとき，標本平均 $n\bar{X}_n$ は母数 n と p の二項分布に従うことがわかる．なぜならば，

$$M_{n\bar{X}_n}(t) = \prod_{i=1}^n \mathbb{E}[e^{tX_i}] = \left[pe^t + (1-p)\right]^n$$

からわかる．