数理統計学まとめ（その 6）：第 4 章 母集団と標本

1 標本抽出

調査対象の集団を**母集団** という.（日本人全体，兵庫県在住の成人男子，神戸市の女子学生全体など）

調べ方

- 全部調べる（**悉皆調査**）：金も時間もかかる．(例：国勢調査）
- 標本の小規模な集団を抜き出して調べる（**標本調査**）：全体の動向を反映しているかが問題.（例：アンケート調査）

標本抽出（標本を母集団から取り出すこと.）

- **復元抽出**： 取り出した標本を元に戻してから次の標本を取り出す.
- **非復元抽出**：標本を元に戻さない.
- **有意抽出法**：調査者が自分の経験や知識によって標本を選ぶ．統計的推測はできない.
- **無作為抽出法**：標本を偶然性に基づいて取り出す.

統計的推測　無作為抽出法で取り出された標本 $X_1, \ldots, X_n$ はそれぞれ母集団の分布に従う確率変数で，互いに独立と考える．標本の値を使って母集団の分布を推測することを統計的推測と言う.

2 標本分布

標本のデータの関数 $T(X_1, \ldots, X_n)$ を（標本）**統計量**という．これは母集団の特性値（平均，分散など．**母数**と呼ぶ.）を推測するために使う.
標本統計量 $T(X_1, \ldots, X_n)$ の分布を標本分布と言う．標本の総数 n を標本の**大きさ**という.

- 母平均 μ に対して 標本平均 $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$
- 母分散 σ^2 に対して標本分散 $S^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2$

3　標本平均の分布

標本平均 $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ の標本分布については次が言える．

定理 3.1（教科書 p.83 定理 4.1）
母平均 μ 母分散 σ^2 の母集団から無作為に抽出された大きさ n の標本を $(X_1, \ldots, X_n)$ とする．このとき標本平均 $\bar{X}$ の分布について次が成立.

$$E(\bar{X}) = \mu, \qquad (\bar{X} \text{ の平均})$$
$$V(\bar{X}) = \frac{\sigma^2}{n}. \qquad (\bar{X} \text{ の分散})$$

証明には次の定理が基本的

定理 3.2（教科書 p.73 定理 3.5 と定理 3.7）
X, Y を確率変数とする．

(1) X, Y が独立ならば，その共分散 $\mathrm{Cov}(X, Y)$ は 0.

(2) $V(X+Y) = V(X) + V(Y) - 2\mathrm{Cov}(X, Y)$ でしたがって X, Y が独立ならば $V(X+Y) = V(X) + V(Y)$.

証明　(1) 第 3 章定理 5.1 教科書 p.73 定理 3.5 として前回紹介した．証明しておく.

X, Y の平均をそれぞれ μ_X, μ_Y と書く.

$$\mathrm{Cov}(X, Y) = E(XY) - \mu_X\mu_Y$$

X, Y が独立なので，$E(XY)$ は

$$\sum_{i=1}^{n}\sum_{j=1}^{m} x_i y_j P(X = x_i)P(Y = y_j) = \left(\sum_{i=1}^{n} x_i P(X = x_i)\right)\left(\sum_{j=1}^{m} y_j P(Y = y_j)\right)$$

に等しく，これは $E(X)E(Y) = \mu_X\mu_Y$ と等しい．$\therefore \mathrm{Cov}(X, Y) = 0$.
(2) については,

$$\begin{aligned} V(X+Y) &= E\left(\{(X+Y) - (\mu_X + \mu_Y)\}^2\right) \\ &= E\left((X-\mu_X)^2 + 2(X-\mu_X)(Y-\mu_Y) + (Y-\mu_Y)^2\right) \\ &= V(X) + 2\mathrm{Cov}(X, Y) + V(Y). \end{aligned}$$

定理 3.1 の証明をしよう．

$$E(\bar{X}) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{n\mu}{n} = \mu.$$

だから $\bar{X}$ の平均は μ になる．一方，$X_1, \ldots, X_n$ が独立なので，定理 3.2 により，

$$\begin{aligned} V(\bar{X}) &= E\left(\left\{\frac{1}{n}\sum_{j=1}^{n}(X_j - \mu)\right\}^2\right) \\ &= \frac{1}{n^2}\sum_{j=1}^{n} V(X_j) + \frac{2}{n^2}\sum_{1\le i<j\le n}\mathrm{Cov}(X_i, X_j) \\ &= \frac{1}{n}\sigma^2. \end{aligned}$$

定理 3.3 （**大数の法則**, 教科書 p.84 定理 4.2）
母集団の平均を μ 分散を σ^2 とする．このときこの母集団からとった標本 $X_1, \ldots, X_n$ について，任意の小さな $\varepsilon > 0$ に対して

$$P(|\bar{X} - \mu| > \varepsilon)$$

は $n \to \infty$ のとき 0 に近づく．

定理 3.4 （教科書 p.84, 定理 4.3）
母集団が正規分布に従えば，無作為標本の標本平均 $\bar{X}$ も正規分布に従う

定理 3.5 （**中心極限定理**, 教科書 p.85 定理 4.4）
母集団の平均を μ，分散を σ^2 とする．どちらも有限な値とする．このときこの母集団からとった標本 $X_1, \ldots, X_n$ について，

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

の分布は $n \to \infty$ のとき，標準正規分布に限りなく近づく．

例 3.1 *1.* Z が標準正規分布に従うとき，$Z \geq 1.88$ となる確率を求める．

標準正規分布表で縦の左端が 1.8, 横の上端が .08 に対応したところを見ると 0.4699 が得られる．これは $0 \leq Z \leq 1.88$ となる確率なので，求める確率は $P(0 \leq Z) = 0.5$ からこの確率を引けば

$$P(0 \leq Z) - P(0 \leq Z \leq 1.88) = P(Z \geq 1.88) = 0.5 - 0.4699 = 0.0301$$

となり，求める確率は 0.0301.

2. X が期待値 (平均) 1.5, 分散 9 の正規分布のとき，$X \geq 7.48$ となる確率を求める．

   $Z = (X - 1.5)/\sqrt{9}$ とおくと，これは定理 3.1 と 教科書 $p.64$ 定理 3.3 により期待値 0, 分散 1 の標準正規分布に従うので，

   $$P(X \geq 7.48) = P(Z \geq \frac{7.48 - 1.5}{3}) = P(Z \geq 1.99)$$

   上と同じように計算して

   $$P(Z \geq 1.99) = 0.5 - P(0 \leq Z \leq 1.99) = 0.5 - 0.4767 = 0.0233.$$

   よって求める確率は 0.0233.

3. Z が標準正規分布のとき，

   $$P(-1.04 \leq Z \leq 1.008)$$

   を求める．

   $$P(-1.04 \leq Z \leq 1.008) = P(-1.04 \leq Z \leq 0) + P(0 \leq Z \leq 1.008)$$

   標準正規分布は左右対象なので，

   $$P(-1.04 \leq Z \leq 0) = P(0 \leq Z \leq 1.04)$$

   標準正規分布表からこの値は 0.3508.

   一方，標準正規分布表から $P(0 \leq Z \leq 1.00) = 0.3413$, $P(0 \leq Z \leq 1.01) = 0.3438$ なので比例配分して (0.02 : 0.08 = 0.2 : 0.8 に内分)

   $$P(0 \leq Z \leq 1.008) = 0.2 \times 0.3413 + 0.8 \times 0.3438 = 0.3433$$