1　標本分布

1.1　標本抽出

母集団と標本　調査することが想定される対象全体のことを母集団（population）という。有限母集団と無限母集団に分けられるが，有限の場合も非常に大きな集団であることが多い[*i]。母集団から無作為に取り出した一部分のことを標本（sample）という。母集団全体を調査する**全数調査**（census），標本のみを調査する**標本調査**（sample survey）がある。費用や時間に制約がある場合や，破壊を伴う調査を行う場合は，全数調査でなく標本調査が行われる。

復元抽出と非復元抽出　母集団の要素を 1 個ずつ取り出すとき，一度取り出したものを元に戻してから次を取り出す方法を復元抽出（sampling with replacement）という。復元抽出では同じ要素が複数回取り出されることがある。反対に，一度取り出したものを元に戻さずに次を取り出す方法を非復元抽出（sampling without replacement）という。標本 $X_1, X_2, \ldots, X_n$ の要素 X_i は確率変数である。有限母集団で非復元抽出をすると X_i は独立にならないが，母集団は非常に大きいため，近似的には独立と見なせる。

有限母集団	復元抽出	$\to$	独立
有限母集団	非復元抽出	$\to$	独立でない（近似的に独立）
無限母集団	復元抽出	$\to$	独立
無限母集団	非復元抽出	$\to$	独立

母集団分布　母集団から取り出した 1 個の要素 X は確率変数であり，X の確率分布のことを母集団分布という。母集団分布の特性を表す定数のことを**母数**（parameter）という[*ii]。X の期待値，分散，標準偏差のことを，母平均，母分散，母標準偏差といい，これらは母数の例である。

標本分布　標本 $X_1, X_2, \ldots, X_n$ の関数のことを**統計量**（statistic）といい，統計量は確率変数になる。統計量の確率分布のことを標本分布という。標本平均，標本分散，不偏分散等は統計量の例である。未知の母数を，統計量を用いて推測することが統計学の主題となっている。

[*i] 硬貨投げ等，何度でも繰り返すことができる試行では，無限母集団を考える。
[*ii] 標本の大きさ，つまり分母の値という意味で「母数」を使う人がいるが，誤用である。

1.2　標本平均の分布

母平均が μ，母分散が σ^2 である母集団から取り出した標本を $X_1, X_2, \ldots, X_n$ とすると，標本の各要素 X_i は確率変数で，互いに独立と考えられる。X_i の分布は母集団分布と同一である。

$$E\left[X_i\right] = \mu, \qquad V\left[X_i\right] = \sigma^2 \tag{1}$$

標本平均の分布　標本 $X_1, X_2, \ldots, X_n$ の平均を**標本平均**（sample mean）といい，$\bar{X}$ と表す。標本平均は統計量のひとつである。

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

各 X_i は (1) の分布をもつから，$\bar{X}$ の期待値は母平均 μ に等しい。

$$\begin{aligned}
E\left[\bar{X}\right] &= E\left[\frac{1}{n}(X_1 + X_2 + \cdots + X_n)\right] \\
&= \frac{1}{n}(E\left[X_1\right] + E\left[X_2\right] + \cdots + E\left[X_n\right]) \\
&= \frac{1}{n}(\mu + \mu + \cdots + \mu) = \frac{n\mu}{n} = \mu
\end{aligned}$$

各 X_i は互いに独立で (1) の分布をもつから，$\bar{X}$ の分散は σ^2/n に等しい。

$$\begin{aligned}
V\left[\bar{X}\right] &= V\left[\frac{1}{n}(X_1 + X_2 + \cdots + X_n)\right] \\
&= \frac{1}{n^2}(V\left[X_1\right] + V\left[X_2\right] + \cdots + V\left[X_n\right]) \\
&= \frac{1}{n^2}(\sigma^2 + \sigma^2 + \cdots + \sigma^2) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}
\end{aligned}$$

以上のことから標本平均 $\bar{X}$ は次の分布をもつ。

$$E\left[\bar{X}\right] = \mu, \qquad V\left[\bar{X}\right] = \frac{\sigma^2}{n}, \qquad \sigma\left[\bar{X}\right] = \frac{\sigma}{\sqrt{n}} \tag{2}$$

1.3　チェビシェフの不等式

Markov の不等式　離散型または連続型確率変数 X は非負（$X \geqq 0$），定数 a は正（$a > 0$）とする。ここでは X は連続型とする。

$$\begin{aligned}
E\left[X\right] &= \int_{-\infty}^{\infty} xf(x)dx = \int_{0}^{\infty} xf(x)dx \\
&\geqq \int_{a}^{\infty} xf(x)dx \\
&\geqq \int_{a}^{\infty} af(x)dx = a\int_{a}^{\infty} f(x)dx = aP\left(X \geqq a\right)
\end{aligned}$$

となるから，次の Markov の不等式が成り立つ。X が離散型の場合も同様である。

$$P(X \geqq a) \leqq \frac{E[X]}{a}$$

Chebyshev の不等式　Y についての Markov の不等式は $P(Y \geqq a) \leqq E[Y]/a$ と表せる。確率変数 X が期待値 μ，標準偏差 $\sigma > 0$ の分布をもつとき，

$$Y = (X-\mu)^2 \geqq 0, \qquad a = k^2\sigma^2 > 0 \qquad (k > 0)$$

とおくと，Markov の不等式から，

$$P\left((X-\mu)^2 \geqq k^2\sigma^2\right) \leqq \frac{E\left[(X-\mu)^2\right]}{k^2\sigma^2}$$
$$P(|X-\mu| \geqq k\sigma) \leqq \frac{\sigma^2}{k^2\sigma^2} = \frac{1}{k^2}$$

となる。期待値 μ，標準偏差 $\sigma > 0$ である確率変数 X と，任意の $k > 0$ に対して，次の Chebyshev の不等式が成り立つ。

$$P(|X-\mu| \geqq k\sigma) \leqq \frac{1}{k^2}, \qquad P(|X-\mu| < k\sigma) \geqq 1 - \frac{1}{k^2} \tag{3}$$

1.4　大数の法則

Chebyshev の不等式　1.3 節の結果から，期待値 μ，標準偏差 $\sigma > 0$ である確率変数 X と，任意の $k > 0$ に対して，次のことが成り立つ。

$$P(|X-\mu| \geqq k\sigma) \leqq \frac{1}{k^2}, \qquad P(|X-\mu| < k\sigma) \geqq 1 - \frac{1}{k^2}$$

大数の法則（law of large numbers）　平均 μ，分散 σ^2 の母集団から大きさ n の標本を取り出し，標本平均を $\bar{X} = (X_1 + X_2 + \cdots + X_n)/n$ とすると，任意の正数 ε に対して，次のことが成り立つ。これを「大数の弱法則」という。

$$\lim_{n\to\infty} P\left(|\bar{X}-\mu| \geqq \varepsilon\right) = 0, \qquad \lim_{n\to\infty} P\left(|\bar{X}-\mu| < \varepsilon\right) = 1 \tag{4}$$

なぜなら，$\bar{X}$ についての Chebyshev の不等式 $P\left(|\bar{X} - E\left[\bar{X}\right]| \geqq k \cdot \sigma\left[\bar{X}\right]\right) \leqq 1/k^2$ は，1.2 節の結果から，

$$P\left(|\bar{X}-\mu| \geqq k\frac{\sigma}{\sqrt{n}}\right) \leqq \frac{1}{k^2}$$

と表すことができ，$\varepsilon = k\sigma/\sqrt{n}$ あるいは $k^2 = \varepsilon^2 n/\sigma^2$ と置き換えると，

$$P\left(|\bar{X}-\mu| \geqq \varepsilon\right) \leqq \frac{\sigma^2}{\varepsilon^2 n} \to 0 \qquad (n \to \infty)$$

(4) の第 1 式が得られる。この「大数の弱法則」は確率収束で表現されているが，概収束で表現された次の「大数の強法則」も成り立つ。[*iii] *iii

$$P\left(\lim_{n\to\infty}\bar{X}=\mu\right)=1 \tag{5}$$

大数の弱法則や大数の強法則の内容はどちらも，$n\to\infty$ のとき，標本平均 $\bar{X}$ が母平均 μ に限りなく近づくことを表している。

中心極限定理（central limit theorem）　平均 μ，分散 σ^2 である母集団から大きさ n の標本を取り出し，標本平均を $\bar{X}=(X_1+X_2+\cdots+X_n)/n$，その標準化を $Z=(\bar{X}-\mu)/(\sigma/\sqrt{n})$ とすると，次のことが成り立つ。これを中心極限定理という。

$$\lim_{n\to\infty}P\left(Z\leqq z\right)=\int_{-\infty}^{z}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx$$

左辺は Z の分布関数の極限，右辺は標準正規分布 $N(0,1)$ の分布関数である。n が十分大きいとき，Z は近似的に $N(0,1)$ に，$\bar{X}$ は近似的に $N(\mu,\sigma^2/n)$ に従う。

$$Z=\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}\sim N(0,1),\qquad \bar{X}\sim N\left(\mu,\frac{\sigma^2}{n}\right) \tag{6}$$

1.5　点推定

*iv **推定量**　母数を推定するために用いる統計量のことを推定量という [*iv]。母数 θ に対して，$E\left[\hat{\Theta}\right]=\theta$ を満たす統計量 $\hat{\Theta}$ のことを，θ の不偏推定量という。

*v 　推定量の実現値を推定値という [*v]。推定量を大文字で，推定値を小文字で表すことが多い。たとえば標本平均なら，推定量は $\bar{X}$，推定値は $\bar{x}$ と表す。

母平均の推定値　平均 μ，分散 σ^2 である母集団から取った標本を $X_1,X_2,\ldots,X_n$，標本平均を $\bar{X}$ とすると，1.2 節の結果から，$\bar{X}$ の期待値は母平均 μ に等しい。

$$E\left[\bar{X}\right]=\mu,\qquad \bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i \tag{7}$$

よって，母平均 μ の推定値は標本平均 $\bar{x}$ である。

母分散の推定値　次の S^2 を**標本分散**（sample variance）という。

$$S^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2,\qquad S^2=\frac{1}{n}\sum_{i=1}^{n}X_i^2-\bar{X}^2$$

[*iii] 確率収束や概収束は難解なので区別できなくてもよい。
[*iv] 母数 θ は，母平均，母分散等。
[*v] 推定量：estimator，推定値：estimate

公式 $V[X]=E[X^2]-\{E[X]\}^2$ の変形 $E[X^2]=\{E[X]\}^2+V[X]$ より

$$E\left[X_i^2\right]=\{E[X_i]\}^2+V[X_i]=\mu^2+\sigma^2$$
$$E\left[\bar{X}^2\right]=\{E\left[\bar{X}\right]\}^2+V\left[\bar{X}\right]=\mu^2+\frac{\sigma^2}{n}$$

であるから，S^2 の期待値は

$$\begin{aligned}
E\left[S^2\right]&=E\left[\frac{1}{n}\sum X_i^2-\bar{X}^2\right]=\frac{1}{n}\sum E[X_i^2]-E\left[\bar{X}^2\right]\\
&=\frac{1}{n}\sum(\mu^2+\sigma^2)-\left(\mu^2+\tfrac{\sigma^2}{n}\right)\\
&=\sigma^2-\frac{\sigma^2}{n}=\frac{n-1}{n}\sigma^2
\end{aligned}$$

S^2 の期待値が母分散 σ^2 と異なるから，S^2 は不偏推定量ではない。そこで U^2 を次のように定める。

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

U^2 の期待値は母分散 σ^2 に等しい。

$$E\left[U^2\right]=E\left[\frac{n}{n-1}S^2\right]=\frac{n}{n-1}E\left[S^2\right]=\frac{n}{n-1}\cdot\frac{n-1}{n}\sigma^2=\sigma^2$$

U^2 のことを母分散 σ^2 の不偏推定量あるいは簡単に**不偏分散**（unbiased variance）という [*vi]。母分散 σ^2 の推定値は不偏分散 u^2 である。 *vi

$$\begin{aligned}
E\left[U^2\right]=\sigma^2,\qquad U^2&=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2\\
U^2&=\frac{n}{n-1}\left(\frac{1}{n}\sum_{i=1}^{n}X_i^2-\bar{X}^2\right)\\
U^2&=\frac{1}{n(n-1)}\left\{n\sum X_i^2-\left(\sum X_i\right)^2\right\}
\end{aligned}\tag{8}$$

[*vi] 不偏分散 U^2 のことを「標本分散」とよぶ文献もある。

参考文献

統計学入門（基礎統計学）		
東京大学教養学部統計学教室（編）	東京大学出版会	978-4-13-042065-5
統計学		
久保川 達也 （著）	東京大学出版会	978-4-13-062921-8
はじめての統計学		
道家 暎幸（共著）	コロナ社	978-4-339-06113-0
確率統計 新版 （新版数学シリーズ）		
岡本 和夫 （ほか著）	実教出版	978-4-407-32171-5
統計学序論 改訂版		
山本 義郎 （著）	東海大学出版部	978-4-486-02133-9
確率統計（高専テキストシリーズ）		
上野 健爾（監修）	森北出版	978-4-627-05561-2
基本統計学 第４版		
宮川 公男 （著）	有斐閣	978-4-641-16455-0
新統計入門		
小寺 平治 （著）	裳華房	978-4-7853-1099-8
Schaum's Outline of Introduction to Probability and Statistics		
Seymour Lipschutz （著）	McGraw-Hill Education	978-0-07-176249-6
A Dictionary of Statistics		
Graham Upton （著）	Oxford Univ Pr	978-0-19-967918-8

`www5e.biglobe.ne.jp/~emm386/statistics/`