第14回　6月10日の授業内容

- §5.　標本分布
  - §5.1　無作為抽出と無作為標本
  - §5.2　統計量の分布
  - §5.3　統計量の分布の近似
    - §5.3.1　大数の法則

§5.1　無作為抽出と無作為標本

- 母集団 population
  - 有限母集団：個体が有限個
  - 無限母集団：個体が無限個
- 無作為抽出 random sampling
  - 復元抽出：sampling with replacement
  - 非復元抽出:sampling without replacement
- 無作為抽出の方法
  - 乱数 (一様乱数)
  - 乱数表

§5.1　無作為抽出と無作為標本(2)

- 無作為標本
  - 無作為抽出された標本
  - 標本の各個体は互いに独立で同じ確率分布にしたがう確率変数
- 統計量 statistics
  - 確率変数の関数
    ⇒ 統計量も確率変数であることに注意!!
  - 標本の各個体の関数であることが多い。
    (例) 標本の算術平均 (標本平均)

§5.2　統計量の分布

- 統計量の分布を知ることの必要性
  - 標本のデータを推測手段とする
    ⇒ 統計量で推測
  - 推測の精度を評価可能。→推測統計の基本
- 統計量g(X_1, X_2, ..., X_n)の分布
  - 平均 $\zeta \equiv \mathrm{E}(g(X_1, X_2, \ldots, X_n))$
    $$= \int \cdots \int_{-\infty}^{\infty} g(x_1, x_2, \ldots, x_n) f(x_1) \times f(x_2) \times \cdots \times f(x_n) dx_1 \cdots dx_n$$
  - 分散 $\mathrm{Var}(g(X_1, X_2, \ldots, X_n)) = \mathrm{E}\{(g(X_1, X_2, \ldots, X_n) - \zeta)^2\}$
    $$= \int \cdots \int_{-\infty}^{\infty} \{g(x_1, x_2, \ldots, x_n) - \xi\}^2 f(x_1) \times f(x_2) \times \cdots \times f(x_n) dx_1 \cdots dx_n$$

§5.2　統計量の分布(2)

- 標本平均 (標本データの算術平均) の分布
  $$\overline{X}_n \equiv \frac{1}{n}\sum_{i=1}^{n} X_i$$
  ただし $\mathrm{E}(X_i) = \mu,\ \mathrm{Var}(X_i) = \sigma^2$
  - 1次モーメント (平均)
    $$\mathrm{E}(\overline{X}_n) \equiv \frac{1}{n}\sum_{i=1}^{n} \mathrm{E}(X_i) = \frac{n\mu}{n} = \mu$$
  - 平均回りの2次モーメント (分散)
    $$\mathrm{Var}(\overline{X}_n) = \mathrm{E}(\overline{X}_n - \mathrm{E}\overline{X}_n)^2 = \frac{1}{n^2}\sum_{i=1}^{n} \mathrm{E}(X_i - \mu)^2 = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

§5.2　統計量の分布(3)

- (例) 二項分布の (平均) 成功比率の分布
  - 確率pで当たりの出るくじをn回引いたときの「当たり」の回数をXとするとX~Bin(n,p)。
  - このとき、「当たり」の比率 X/n の分布を考えよう。
    二項分布の性質より、
    $\mathrm{E}(X/n) = (1/n)\mathrm{E}(X) = np/n = p$
    $\mathrm{Var}(X/n) = (1/n^2)\mathrm{Var}(X) = np(1-p)/n^2 = p(1-p)/n$
    となることがわかる。

§5.2 統計量の分布(4)

- 有限母集団の場合の統計量の分布
  - $B_1, B_2, \ldots, B_N$：母集団の個体
  - 標本の大きさnの非復元抽出を行う場合
  - 標本平均の分布のモーメント

$$\mathrm{E}\left(\overline{X}_n\right)=\frac{{}_{N-1}C_{n-1}}{{}_N C_n}\frac{1}{n}\sum_{j=1}^{N}B_j=\frac{1}{N}\sum_{j=1}^{N}B_j$$

$$\mathrm{Var}\left(\overline{X}_n\right)=\frac{N-n}{N-1}\frac{1}{n}\left(\frac{1}{N}\sum_{j=1}^{N}\left(B_j-\frac{1}{N}\sum_{k=1}^{N}B_k\right)^2\right)$$

§5.3 統計量の分布の近似
§5.3.1 大数の法則 law of large numbers

標本の大きさn→∞のとき、標本平均が母集団平均μに収束するという法則

- チェビシェフの不等式 Chebyshev's inequalityによる説明
  - 平均、分散が $\mathrm{E}(X)=\mu,\ \mathrm{Var}(X)=\sigma^2$である確率変数$X$について任意の定数$c>0$に対して次の式が成立する。

$$P\left(\left|X-\mu\right|\geq c\right)\leq\frac{\sigma^2}{c^2}$$

  - $\mathrm{E}(\overline{X}_n)=\mu,\ \mathrm{Var}(\overline{X}_n)=\sigma^2/n$ であるから、任意の $\varepsilon>0$に対して

$$P\left(\left|\overline{X}_n-\mu\right|\geq\varepsilon\right)\leq\frac{\sigma^2/n}{\varepsilon^2}$$ よって $$\lim_{n\to\infty}P\left(\left|\overline{X}_n-\mu\right|\geq\varepsilon\right)=0$$

§5.3.1 大数の法則(2)

- 大数の法則 (Khintchine's law)
  - 大数の弱法則ともいう

$X_1, X_2, \cdots, X_n$ を平均がμである、互いに独立な確率変数であるとき、すなわち、$\mathrm{E}(X_i)=\mu$ and $X_i \perp X_j$ for $i\neq j$ であるとき、

$$\lim_{n\to\infty}\mathrm{P}\left(\left|\overline{X}_n-\mu\right|<\varepsilon\right)=1 \qquad \forall\varepsilon>0$$

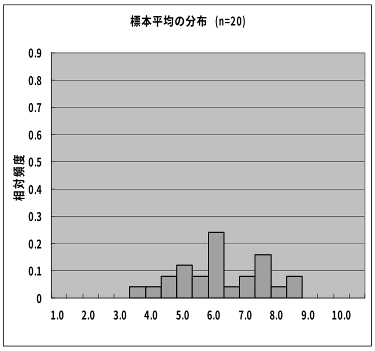
が成立する。このことを $\overline{X}_n$ がμに確率収束するといい、

$\overline{X}_n \xrightarrow{P} \mu$ または、 $\operatorname*{plim}_{n\to\infty}\overline{X}_n=\mu$ と表す。

§5.3.1 大数の法則(3) 実例

- 大数の法則を確認するための実験
  - X~N(6,36)にしたがう確率変数をn個抽出した標本を25ヶ作成。
  - 25ヶのデータからなる標本それぞれについて、標本平均（算術平均）を求め、標本平均の分布（ヒストグラム）を作成。
  - nとして、20, 50, 100, 200, 400, 1000を採用

§5.3.1 大数の法則(4) 実例 n=20



§5.3.1 大数の法則(9) 実例 n=1,000