第 8 章　確率モデルと推定

1　確率モデル

モデル： 複雑な現象や構造などを単純化して，その仕組みを簡潔に表現したもの

確率モデル：確率変数を導入して母集団を描写したモデル

$$X \ = \ \textbf{系統的部分} \ + \ \textbf{説明不可能な要因}$$

2　単純確率モデル

測定値，身長，体重，試験の点数などのように，一定値のまわりで分布する母集団に対する確率モデルは，**単純確率モデル**と呼ばれる．それは，確率変数

$$X = \mu + \varepsilon \qquad \mu \text{ 未知定数（パラメータ）} \qquad \varepsilon \text{ 誤差項（確率変数）}$$

のすべての実現値からなると想定される．このとき，サイズ n のランダム標本は，

$$X_i = \mu + \varepsilon_i \qquad i = 1, \cdots, \ n \qquad \{\varepsilon_i\} \sim i.i.d.(0, \sigma^2)$$

と表すことができる．

3　推定

推定：標本に基づいて母集団分布のパラメータを推論すること

推定量：パラメータの推定方式を表現する統計量（確率変数）

推定値：推定量の実現値

以下，単純確率モデルの場合で考える．

平均 μ の推定　μ の推定量としては標本平均がよく使われる．この推定量を

$$\hat{\mu} \ = \ \overline{X} \ = \frac{1}{n}\sum_{i=1}^{n} X_i \ = \mu \ + \frac{1}{n}\sum_{i=1}^{n} \varepsilon_i$$

と表す．

$$\left.\begin{aligned} &E(\hat{\mu}) = \mu && \text{（不偏）} \\ &V(\hat{\mu}) = \frac{\sigma^2}{n} && \to \ 0 \end{aligned}\right\} \Rightarrow \quad \hat{\mu} \text{ は } \mu \text{ に確率収束する}$$

標準誤差（SE: Standard Error）：推定量の標準偏差

真のパラメータとのへだたりの目安

$$\hat{\mu} \text{ の } SE \text{ は } \sigma/\sqrt{n} \text{ である．} \Leftrightarrow \quad SE(\hat{\mu}) = \sigma/\sqrt{n}$$

分散の推定

$$\hat{\sigma}^2 \;=\; \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2$$ **：不偏かつ** σ^2 **に確率収束**

$$\overline{\sigma}^2 \;=\; \frac{1}{n}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2 \qquad E\left(\overline{\sigma}^2\right)=\frac{n-1}{n}\sigma^2<\sigma^2$$ **；　下方バイアス**

4　正規母集団の場合の信頼区間

平均　μ　の区間推定（分散既知の場合）

$$P\left(\frac{\sqrt{n}\left|\overline{X}-\mu\right|}{\sigma}\le z_{\alpha/2}\right)=1-\alpha$$ **の括弧内を　μ　に関して解いて**

$$\overline{X}-z_{\alpha/2}\times\frac{\sigma}{\sqrt{n}}\le\mu\le\overline{X}+z_{\alpha/2}\times\frac{\sigma}{\sqrt{n}}$$

これを，μ に関する**信頼係数** 1-α の**信頼区間**という．

信頼区間の意味合い：テキスト 175 ページを参照

平均　μ　の区間推定（分散未知の場合）

$$P\left(\frac{\sqrt{n}\left|\overline{X}-\mu\right|}{\hat{\sigma}}\le t_{\alpha/2}^{(n-1)}\right)=1-\alpha$$ **の括弧内を　μ　に関して解いて**

$$\overline{X}-t_{\alpha/2}^{(n-1)}\times\frac{\hat{\sigma}}{\sqrt{n}}\le\mu\le\overline{X}+t_{\alpha/2}^{(n-1)}\times\frac{\hat{\sigma}}{\sqrt{n}} \qquad (6)$$

〔例題 8.2〕 第 2 章の**表 2-3** の TOPIX 収益率データを正規母集団からのランダム標本の実現値とみなすとき，母集団のリターン μ に関する信頼係数 90% の信頼区間を求めよ．ただし，自由度 83 の t 分布の上側 5% 点は 1.663 である．

（解）(6) 式を適用する．ここで，

$$t_{\alpha/2}^{(n-1)}=t_{0.05}^{(83)}=1.663$$

であり，データのリターン（平均）とリスク（標準偏差）はそれぞれ -0.6325%，5.1232% であるから，これらの値を (6) に代入して，母集団のリターンの信頼区間は

$$\left[-0.6325-1.663\times\frac{5.1232}{\sqrt{84}},\quad -0.6325+1.663\times\frac{5.1232}{\sqrt{84}}\right]=\left[-1.562,\quad 0.297\right]$$

分散の区間推定（平均既知の場合）

$$\tilde{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \mu\right)^2$$ **とおく．**

$$P\left(c_{1-\alpha/2}^{(n)} \le \frac{n\tilde{\sigma}^2}{\sigma^2} \le c_{\alpha/2}^{(n)}\right) = 1-\alpha$$ **の 括弧内を σ^2 に 関して 解い て**

$$\frac{n\tilde{\sigma}^2}{c_{\alpha/2}^{(n)}} \le \sigma^2 \le \frac{n\tilde{\sigma}^2}{c_{1-\alpha/2}^{(n)}}$$

テキスト 177 ページの図 8-3 を参照

分散の区間推定（平均未知の場合）

$$\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$$ **とおく．**

$$P\left(c_{1-\alpha/2}^{(n-1)} \le \frac{(n-1)\hat{\sigma}^2}{\sigma^2} \le c_{\alpha/2}^{(n-1)}\right) = 1-\alpha$$ **の 括弧内を σ^2 に 関して 解い て**

$$\frac{(n-1)\hat{\sigma}^2}{c_{\alpha/2}^{(n-1)}} \le \sigma^2 \le \frac{(n-1)\hat{\sigma}^2}{c_{1-\alpha/2}^{(n-1)}} \qquad (9)$$

テキスト 177 ページの図 8-3 を参照

〔例題 8.3〕 例題 8.2 と同一の収益率データにおいて，母集団分散に関する信頼係数 90% の信頼区間を求めよ．また，リスク σ に関する信頼区間も同様に求めよ．ただし，自由度 83 のカイ 2 乗分布の下側 5% 点は 63.00，上側 5% 点は 105.27 である．

（解）母集団分散の信頼区間を求めるには，(9) 式を適用する．

$$\left[\frac{83\times(5.1232)^2}{105.27},\quad \frac{83\times(5.1232)^2}{63.00}\right] = [20.698,\quad 34.579]$$

リスク σ の信頼区間は，これらの平方根を考えればよいから，

$$\left[\sqrt{20.698},\quad \sqrt{34.579}\right] = [4.55,\quad 5.88]$$

5　正規母集団でない場合の信頼区間

ベルヌーイ母集団の場合

$$P\left(\frac{|\hat{p}-p|}{\sqrt{p(1-p)/n}} \le x_{\alpha/2}\right) = 1-\alpha$$ **の 括弧内を p に関して解くのは複雑**

⇒

$$P\left(\frac{|\hat{p}-p|}{\sqrt{\hat{p}(1-\hat{p})/n}} \le x_{\alpha/2}\right) = 1-\alpha$$ **の 括弧内を p に関して解く.**

$$\hat{p} - z_{\alpha/2} \times \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \le p \le \hat{p} - z_{\alpha/2} \times \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \qquad (10)$$

〔例題 8.4〕 失業率調査で労働力人口としてカウントされる 10,000 人について調査したところ，250 人が失業中であった．このデータをランダム標本の実現値とみなすとき，全労働力人口における失業率の 99% 信頼区間を求めよ．

（解）母集団失業率を p とする．標本サイズは 10,000 と非常に大きいので，p の信頼区間は (10) に基づいて構成することができる．データから

$$\left[0.025 - 2.576 \times \sqrt{\frac{0.025 \times 0.975}{10000}},\quad 0.025 + 2.576 \times \sqrt{\frac{0.025 \times 0.975}{10000}}\right] = [0.021,\quad 0.029]$$

問．1994 年の学校保健統計調査によると，全国約 20,000 人の 17 才男子の身長の平均は 170.7 cm，標準偏差は 5.61 cm である．母集団における身長の平均の 95% 信頼区間を構成せよ．

6　推定量の統一的な構成

第 3 節で平均や分散の推定を考えた際には，推定量を直観的に構成したが，ここでは統一的な観点から構成することを考える．そのための代表的な推定方式としては，最小 2 乗法および最尤（さいゆう）法がある．

7　推定量のよさの基準

同一のパラメータに対してはさまざまな推定量が存在する．そこで，複数の推定量の中での優劣をどのように決めるかが問題となる．例えば，分散を最小にするような推定量が最も望ましいとするならば，標本とは無関係に常にある定数を推定量とすれば，その分散は 0 となる．ここでは，このような場合を排除した上で優劣を比較する際の基準とし不偏性，有効性，一致性について述べる．