

第0章 なぜ経験過程理論を学ぶか

0.1 なぜ経験分布理論を学ぶか

まず, 最も簡単な経験過程の例をみてみよう. 正規分布からの 10 個の乱数を x_1, x_2, \dots, x_{10} と書く. 経験分布関数 \hat{F}_{10} は

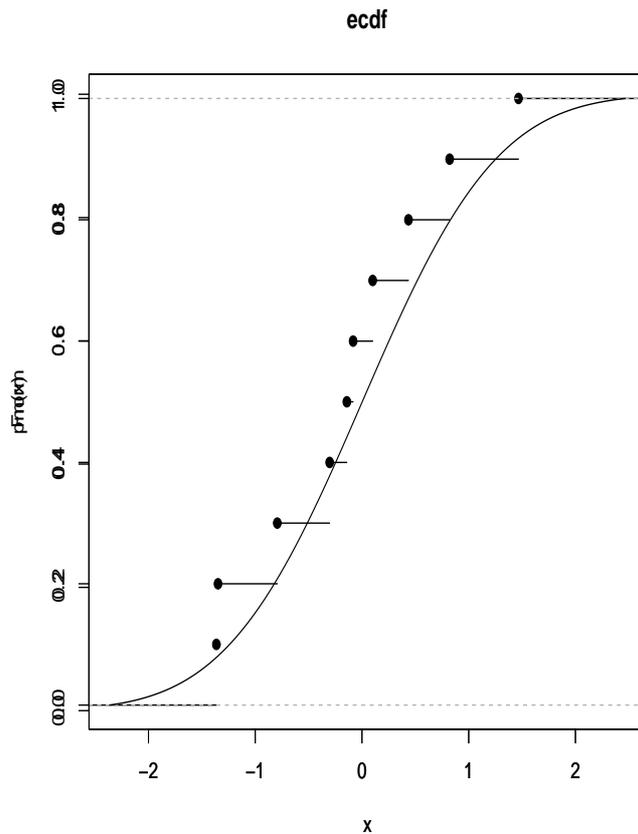
$$\hat{F}_{10}(x) := \frac{\{x_i \leq x (i = 1, 2, \dots, 10)\} \text{ の個数}}{10} \quad (x \in \mathbb{R}) \quad (1)$$

で定義される. また, 標準正規分布の分布関数 F は

$$F(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt \quad (x \in \mathbb{R})$$

である.

下記の図は標準正規分布からの 10 個の乱数から作図した経験分布関数と標準正規分布の分布関数のグラフを並べたものである. 曲線は標準正規分布の分布関数のグラフであり, 階段状の関数は経験分布関数のグラフである. この 2 つの関数のグラフは同じような位置にあることが見て取れる.



R コード

```

> x<-rnorm(10)
> xlimit<-c(min(x)-1,max(x)+1)
> plot(ecdf(x),xlim=xlimit,main="ecdf")
> par(new=T)
> plot(pnorm,min(x)-1,max(x)+1)

```

上記の数値実験の設定を確率変数を用いて表現してみよう. 標準正規分布 $N(\mu, 1)$ ($\mu \in \mathbb{R}$) からの標本の大きさ n のランダム標本を X_1, X_2, \dots, X_n とする¹と (1) のランダムバージョンを考えることができる.

$$\hat{F}_n(x) := \frac{\{X_i \leq x (i = 1, 2, \dots, n)\} \text{ の個数}}{n} \quad (x \in \mathbb{R}).$$

標本平均

$$\bar{X}_n = \frac{1}{n}(X_1 + X_2 + \dots + X_n)$$

¹これらの確率変数たちは確率空間 $(\Omega, \mathcal{A}, \Pr)$ 上で定義されているとする.

が母平均 μ の推定量であるように, \widehat{F}_n を分布関数 F の推定量と考えることができる.

標本平均は \overline{X}_n はよい性質を持っている. すなわち

$$E[\overline{X}_n] = \mu$$

である. さらに, 任意の $\epsilon > 0$ に対して

$$\lim_{n \rightarrow \infty} \Pr(|\overline{X}_n - \mu| > \epsilon) = 0$$

が成立する. 同様に, 固定した x に対して

$$E[\widehat{F}_n(x)] = F(x)$$

であり

$$\lim_{n \rightarrow \infty} \Pr(|\widehat{F}_n(x) - F(x)| > \epsilon) = 0$$

が成立することが標本平均と同じ議論で証明できる. さらに

$$\lim_{n \rightarrow \infty} \Pr\left(\sup_{x \in \mathbb{R}} |\widehat{F}_n(x) - F(x)| > \epsilon\right) = 0 \quad (2)$$

が成り立つことも知られている. (2) から, 標本平均と同じように経験過程の性質も議論できるように見える.

しかし, もうすこし複雑なものを考えてみよう. 正規分布 $N(\mu, 1)$ の p.d.f. を p_μ と書いたとき, それに対応する確率測度を

$$P(B) = \int_B p_\mu(x) dx \quad (B \in \mathcal{B}(\mathbb{R}))$$

と定める. $x \in \mathbb{R}$ に対して, $B = (-\infty, x]$ とすると

$$P(B) = F(x)$$

となる. (1) と同様に経験確率測度を

$$\widehat{P}_n(B) := \frac{\{X_i \in B (i = 1, 2, \dots, n)\} \text{ の個数}}{n} \quad (B \in \mathcal{B}(\mathbb{R}))$$

と定める. 固定した $B \in \mathcal{B}(\mathbb{R})$ と任意の $\epsilon > 0$ に対して

$$\lim_{n \rightarrow \infty} \Pr(|\widehat{P}_n(B) - P(B)| > \epsilon) = 0$$

であるが

$$\sup_{B \in \mathcal{B}(\mathbb{R})} |\widehat{P}_n(B) - P(B)| = 1 \quad (3)$$

となることが知られている. Borel 集合に関して一様にみると (2) のようにはうまくいかない. (2) と (3) の違いがどうして起こるかを理解するには, 経験過程理論が必要になってくる.

いま, 関数 $g: \mathbb{R} \rightarrow \mathbb{R}$ に対して

$$Pg := \int_{\mathbb{R}} g(x) dP(x), \quad \widehat{P}_n g := \frac{1}{n} \sum_{j=1}^n g(X_j)$$

と定める. さらに

$$\nu_n(g) := \sqrt{n}(\widehat{P}_n - P)g$$

とおく. すると関数 g を動かすと, 関数族 $\mathcal{G}()$ 便宜的な記号で添え字付けられた確率過程 $\{\nu_n(g)\}_{g \in \mathcal{G}}$ が得られる. ただし, \mathcal{G} はある関数族である. たとえば, $\mathcal{G}_1 = \{\mathbb{1}_{(-\infty, t]}(x); t \in \mathbb{R}\}$ とおくと

$$\{\nu_n(g); g \in \mathcal{G}_1\} = \{\sqrt{n}(\widehat{F}_n(t) - F(t)); t \in \mathbb{R}\}$$

となる. また, $\mathcal{G}_2 = \{\mathbb{1}_B(x); B \in \mathcal{B}(\mathbb{R})\}$ とすると

$$\{\nu_n(g); g \in \mathcal{G}_2\} = \{\sqrt{n}(\widehat{P}_n(B) - P(B)); B \in \mathcal{B}(\mathbb{R})\}$$

となる. \mathcal{G}_1 で添え字付けられた経験過程はうまく扱うことが期待されるが, \mathcal{G}_2 で添え字付けられた経験過程はうまく扱うことができないようである.

このことから, 経験過程を添え字付けられている関数族の大きさが問題であることが想像される. さらに, 関数族で添え字づけられた経験過程を考えることで, 統計推測において有効になる極限定理を扱うこともできることに注意しておく.