統計推測特論（稲葉） <u>No.D1</u>

2021.10.26(火２)

ZOOM

D.0. ３つの基本定理（統計学の考え方を導く定理）

（データ数を多くすると、情報が増え正規分布に近くなる。）

定理 D.1. <u>大数の法則</u>（データ数を増やすと情報量が増える）

$X_1, X_2, \cdots, X_n$：互いに独立で同じ分布に従う確率変数

[$E(X_i) = \mu,\ \ V(X_i) = \sigma^2$ とする。]

$\Rightarrow$ 任意の $\varepsilon > 0$ に対し、$\lim_{n\to\infty} P(|\overline{X} - \mu| > \varepsilon) = 0$

（標本の大きさ n（データの数）を大きくしていくと、$\overline{X}$ と μ は近くなる可能性が高い。(ε 以上離れていることはできない。)）

（略証）まず、標本平均 $\overline{X}$ の期待値と分散を計算すると

$$E(\overline{X}) = \mu,\ V(\overline{X}) = \frac{\sigma^2}{n}$$

と分る。これに、チェビシェフの不等式を適用すると

$$P(|\overline{X} - \mu| \geqq c\tfrac{\sigma}{\sqrt{n}}) \leqq \frac{1}{c^2}$$

ここで、$\frac{c\sigma}{\sqrt{n}} = \varepsilon$ となるように c を決めると c は n 毎に変えないといけない。つまり、

$$c = c_n = \frac{\sqrt{n}\varepsilon}{\sigma}$$

とおくことになる。したがって

$$P(|\overline{X} - \mu| \geqq \varepsilon) \leqq \frac{\sigma^2}{n\varepsilon^2}$$

の右辺は 0 に近づいていく。（略証終）

（準備）定理 D.2. <u>チェビシェフの不等式</u>

母平均 μ、母分散 σ^2 をもつ確率変数 X に対し、

$$P(|X - \mu| \geqq c\sigma) \leqq \frac{1}{c^2},\ \ c > 0.$$

（確率変数が母平均から母標準偏差の定数倍以上離れる確率は、その定数を用いて、評価できる。）

（略証）まず、母分散 $\sigma^2 = E\{(X-\mu)^2\}$ である。これを、３つの場合（① $k \leqq \mu - c\sigma$ の場合、② $\mu - c\sigma < k < \mu + c\sigma$ の場合、③ $k \geqq \mu + c\sigma$ の場合）に分けて考える。

$$\begin{aligned}
\sigma^2 &= \sum_{k\leq\mu-c\sigma} (k-\mu)^2 f_k + \sum_{\mu-c\sigma<k<\mu+c\sigma} (k-\mu)^2 f_k + \sum_{k\geq\mu+c\sigma} (k-\mu)^2 f_k \\
&\geqq \sum_{k\leq\mu-c\sigma} (c\sigma)^2 f_k + \sum_{\mu-c\sigma<k<\mu+c\sigma} 0^2 f_k + \sum_{k\geq\mu+c\sigma} (c\sigma)^2 f_k \\
&= c^2\sigma^2 Pr\{|X-\mu| \geqq c\sigma\}
\end{aligned}$$

よって、

$$P(|X - \mu| \geqq c\sigma) \leqq \frac{1}{c^2},\ \ c > 0.$$ （略証終）

2021.10.26(火２)

ZOOM

定理 D.3. 中心極限定理（データ数が増えると、正規分布が近似的に成り立つと考えるうる）

母平均 μ、母分散 σ^2 のある分布からの互いに独立な標本 $X_1, X_2, \cdots, X_n$ に対し、
$S_n = X_1 + X_2 + \cdots + X_n$ とおくと

$$\lim_{n\to\infty} P\left(\frac{S_n - n\mu}{\sqrt{n}\sigma} \leqq a\right) = \lim_{n\to\infty} P\left(\frac{\sqrt{n}(\overline{X}-\mu)}{\sigma} \leqq a\right) = \Phi(a)$$

ここで、$\Phi(a) = \displaystyle\int_{-\infty}^{a} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt.$（標準正規分布の分布関数）

（略証）まず、各 X_i に母平均 μ と母分散 σ^2 があるので、X_i の積率母関数は以下のように表せる。

$$M_{X_i}(t) = \exp(\mu t + \tfrac{1}{2}\sigma^2 t^2 + bt^3 + \cdots) \tag{E.6.1}$$

ここで、b は定数。
また、積率母関数には、独立な確率変数の和について、以下の性質も成り立つ。

$$M_{X_1+X_2+\cdots+X_n}(t) = M_{X_1}(t)M_{X_2}(t)\cdots M_{X_n}(t) \tag{E.6.2}$$

よって、$S_n = X_1 + X_2 + \cdots + X_n$ の積率母関数は、次になる。

$$M_{S_n}(t) = \left\{M_{X_1}(t)\right\}^n = \exp(n\mu t + \tfrac{1}{2}n\sigma^2 t^2 + nbt^3 + \cdots) \tag{E.6.3}$$

更に、積率母関数は、定義より、以下のような性質を持つ。

$$M_{cX+d}(t) = E(e^{t(cX+d)}) = e^{dt}E(e^{(ct)X}) = e^{dt}M_X(ct) \tag{E.6.4}$$

これらの３つの関係式と、S_n の母平均が $n\mu$ で母分散が $n\sigma^2$ であることから、
S_n を標準化した $Y = \frac{S_n - n\mu}{\sqrt{n}\sigma}$ の積率母関数は、$c = \frac{1}{\sqrt{n}\sigma}, d = -\frac{\sqrt{n}\mu}{\sigma}$ とおけば以下を得る。

$$\begin{aligned} M_Y(t) &= \exp(-dt)\exp(n\mu(ct) + \tfrac{1}{2}n\sigma^2(ct)^2 + nb(ct)^3 + \cdots) \\ &= \exp(-\tfrac{\sqrt{n}\mu}{\sigma}t)\exp\{n\mu(\tfrac{1}{\sqrt{n}\sigma}t) + \tfrac{1}{2}n\sigma^2(\tfrac{1}{\sqrt{n}\sigma}t)^2 + nb(\tfrac{1}{\sqrt{n}\sigma}t)^3 + \cdots\} \\ &= \exp(\tfrac{1}{2}t^2 + b\tfrac{1}{\sqrt{n}\sigma^3}t^3 + \cdots) \end{aligned} \tag{E.6.5}$$

したがって、$n \to \infty$ のとき、Y の極限分布が、標準正規分布 $N(0,1)$ に近づくことがわかる。
（略証終）

系 D.4. ラプラスの定理

$$X \sim B(n,p) \Rightarrow \lim_{n\to\infty} P\left(\frac{X - np}{\sqrt{np(1-p)}} \leqq a\right) = \Phi(a)$$

（略証）$E(X) = np, V(X) = np(1-p)$ であるから、X を標準化したものの極限分布は、中心極限定理によって、標準正規分布 $N(0,1)$ になる。（略証終）