統計学の基礎18 (確率分布9/10)

■大数の法則と中心極限定理

確率変数の平均 $\overline{X} = \dfrac{1}{n}\sum_{i=1}^{n} X_i$ を標本平均という.

(大数の法則)

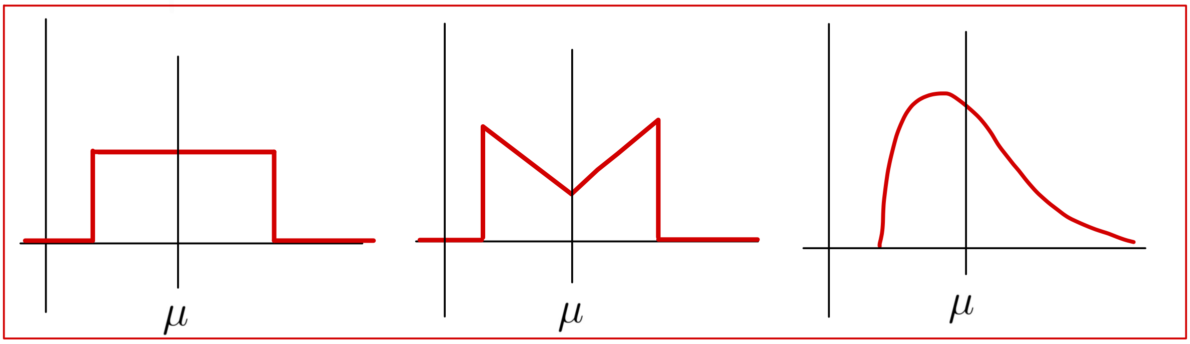
$X_1, X_2, \cdots, X_n$ は互いに独立で，平均μ 、分散σ^2の同一の確率分布に従うとする．このとき, 任意の $\varepsilon > 0$ に対して, $\lim_{n\to\infty} P(|\overline{X} - \mu| < \varepsilon) = 1$ となることが知られている.

この法則により、標本平均は n を大きくするとき，実現値が期待値μ に近い値をとることが期待できる.

(中心極限定理)

$X_1, X_2, \cdots, X_n$ は互いに独立で，平均μ，分散σ^2 の同一の確率分布に従うとする.

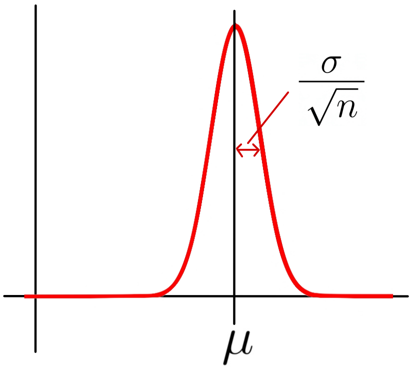
もとの分布(平均・分散が存在さえすればどんな分布でもよい)



$\Downarrow$

もとがどんな分布であっても，標本平均 $\overline{X} = \dfrac{1}{n}\sum_{i=1}^{n} X_i$ はn が十分大きいとき，近似的に以下正規分布に従うことが知られている.

←「n が十分大きい」の1つの目安は$n \geqq 30$ (村上・安田著『統計学演習』)



$$\overline{X} \sim \mathrm{N}\left(\mu, \frac{\sigma^2}{n}\right)$$

標準化すれば

$$\frac{\overline{X} - \mu}{\sqrt{\sigma^2/n}} \sim \mathrm{N}(0, 1)$$

←標本平均の平均は母集団の平均と等しく，標本平均の分散は母集団の分散を標本の大きさnで割ったものに等しいことは計算ですぐ分かるが(問題13-1参照)，これに加えて(もとがどんな分布でも)近似的に正規分布に従うことを主張している.

■二項分布の正規分布近似

$X \sim \mathrm{B}(n, p)$ のとき,

$X = X_1 + X_2 + \cdots + X_n$ （$X_1, X_2, \cdots, X_n$ は互いに独立で $\mathrm{Be}(p)$ に従う）

と表されるから, $\dfrac{X}{n} = \dfrac{X_1 + X_2 + \cdots + X_n}{n}$ は $\mathrm{Be}(p)$ に従う n 個の確率変数の標本平均とみなせる．よって中心極限定理により，近似的に

$$\frac{X}{n} \sim \mathrm{N}\left(p, \frac{p(1-p)}{n}\right)$$

← $\mathrm{Be}(p)$ の平均はp，分散は$p(1-p)$ である.

すなわち

$$X \sim \mathrm{N}\left(np, np(1-p)\right)$$

←確率変数をn倍したので，平均は n 倍，分散は n^2 倍になっている

(18-1)サイコロを100個投げるとき、その平均が4以上となる確率を求めよ.

(18-2)

ある試験は問題が全部で50問あり，各問題は○か×の2つの選択肢からなる．ある受験生が各問題ごとにランダムに解答するとき，30問以上正答する確率を求めよ．