標本と母集団

母集団 (population)

n ：標本の大きさ (sample size)

θ

抜き取り，抽出 (sampling)

標本，試料 (sample)

$x_1, x_2, x_3, \cdots, x_n$

母数

母平均 (μ, m)

母分散 (σ^2)

母標準偏差 (σ)

母比率 (p)

母集団の範囲 (R)

母集団の相関係数 (ρ)

$\hat{\theta}$

母数の推定

標本の性質を示す数値

標本平均 ($\bar{x}$)

標本分散 (s^2)

標本の標準偏差 (s)

標本比率 ($\hat{p}$)

標本の範囲 (R)

標本の相関係数 (ρ) など

N：母集団の大きさ (population size)

標本分布 (sampling dist.)

標本平均($\bar{x}$), 標本分散(s^2), 不偏標本分散(u^2) 等の統計量は確率変数だから確率分布を持つ.

統計量の確率分布は標本分布と呼ばれる.

標本の分布を知ることは, 元となる母集団の分布を知る上で重要である.

標本の分布と母集団の分布の関係が重要

たとえば, 標本平均から母平均を知る.

中心極限定理 (central limit theorem)

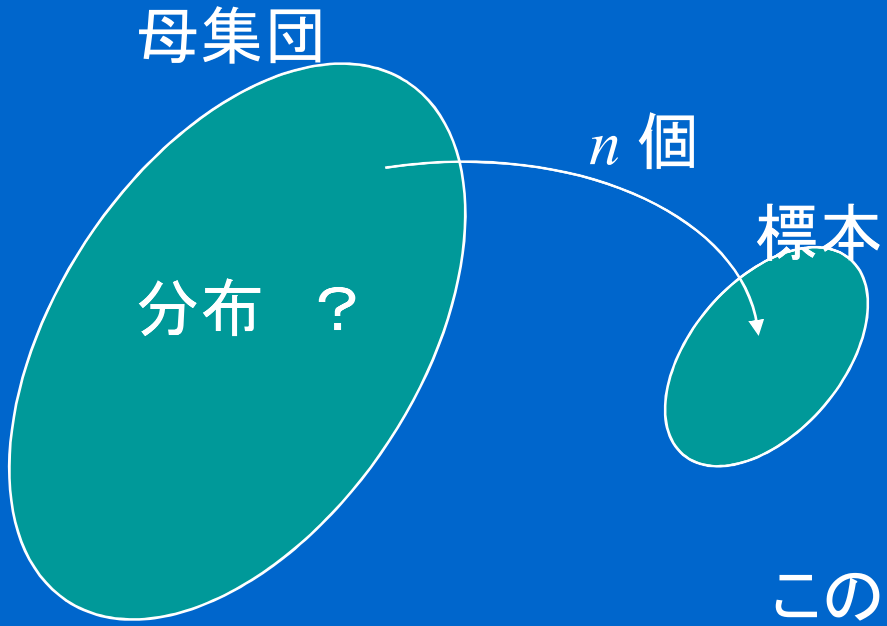
母集団の(分布の)型に関係なく,

標本平均 ($\bar{x}$) の分布は,

標本数 (n) が大きければ,

正規分布 $N\left(\mu,\ \dfrac{\sigma^2}{n}\right)$ に従う.

中心極限定理



$x_1, x_2, x_3, \cdots, x_n$　：　$\bar{x}$

この操作を繰り返す

$\bar{x}_1, \bar{x}_2, \bar{x}_3, \cdots$

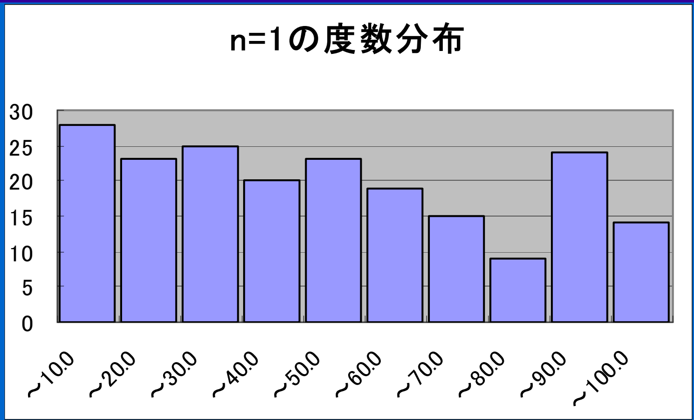
この $\bar{x}$ の分布は n が大きくなると

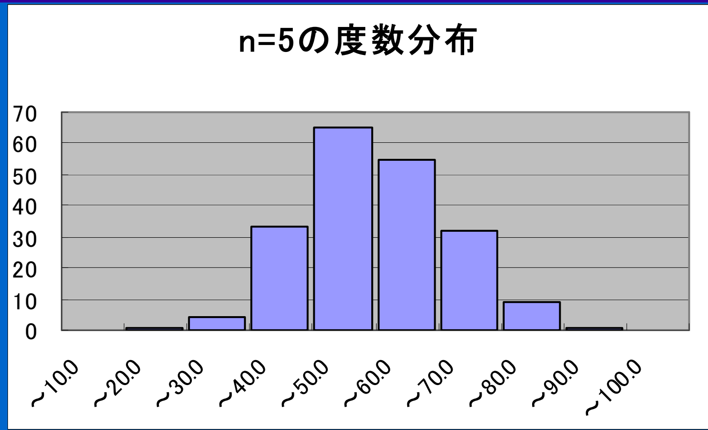
$\left(\mu, \dfrac{\sigma^2}{n}\right)$ に従う

n が大きくなるほど

その精度は高くなる

n個の平均値の分布





平　　均　44.32

標準偏差　29.46

平　　均　50.22

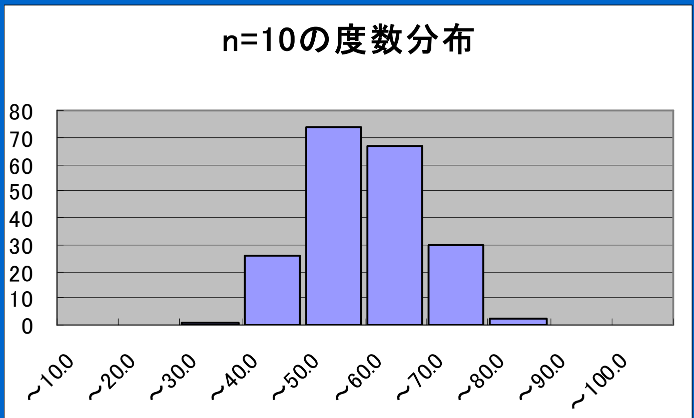
標準偏差　8.75

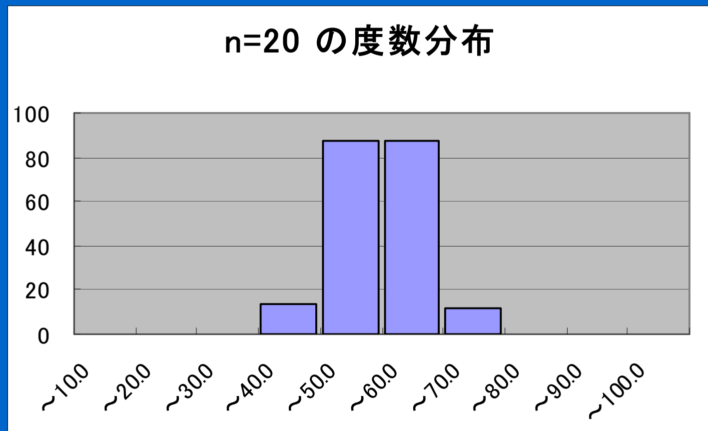
平　　均　50.32

標準偏差　11.72

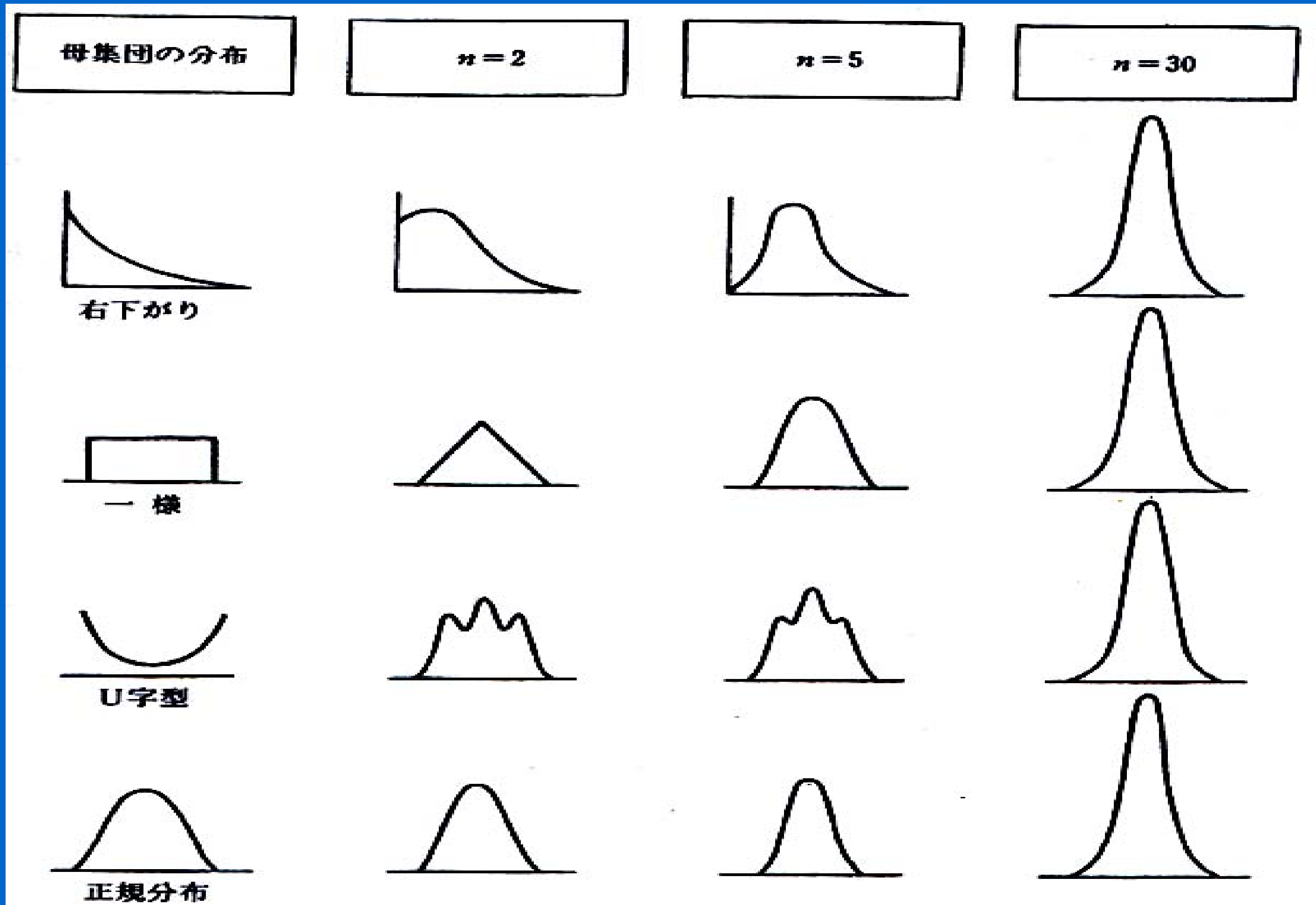
平　　均　50.01

標準偏差　6.65





中心極限定理のイメージ



中心極限定理の応用

ランダム標本 $X = (X_1, X_2, \cdots, X_n)$ について,
各 $\overline{X}_i$ の平均値が$\mu(\infty < \mu < \infty)$, 分散が$\sigma^2(0 < \sigma^2 < \infty)$であるとき,
標本平均 X の標本分布は, 標本の大きさ n が十分大きければ,
正規分布 $N(\mu, \sigma^2/n)$ に近似する.
すなわち,

$$\overline{X} \sim N(\mu, \sigma^2/n) \quad (n \to \infty)$$

二項分布の正規近似

$B(np, npq)$ において $$P(a \leqq S_n \leqq b) \fallingdotseq \int_{\frac{a-np}{\sqrt{npq}}}^{\frac{b-np}{\sqrt{npq}}} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$$

(a, bは定数)

点推定 (point estimation)

少数の標本から，母数をずばり
1つ数値で言い当てる.

どの推定量を利用するかの基準

- 不偏性 (unbiasedness)
- 一致性 (consistent)
- 有効性 (efficiency)

不偏性

未知母数をθ，その推定量を$\hat{\theta}$としたとき，

$$E(\hat{\theta}) = \theta$$

となるような$\hat{\theta}$をθの不偏推定量と呼ぶ.

この推定量$\hat{\theta}$は不偏性をもつという.

たとえば

$$E(\bar{x}) = \{E(x_1) + E(x_2) + \cdots + E(x_n)\} / n$$

$$= (\mu + \mu + \cdots + \mu) / n$$

$$= \mu$$

すなわち, $\bar{x}$ は母平均μの不偏推定量である.

$E(\tilde{x}) = \mu$ となり,

$\tilde{x}$ も母平均μの不偏推定量である.

一致性

未知母数をθ，その推定量を$\hat{\theta}$としたとき，

$$\hat{\theta} \to \theta \qquad (n \to \infty)$$

となるような$\hat{\theta}$をθの一致推定量と呼ぶ．

この推定量$\hat{\theta}$は一致性をもつという．

有効性

2つの分散の異なる不偏推定量$\hat{\theta}$, $\bar{\theta}$
があるとき,

$$V(\hat{\theta}) \leqq V(\bar{\theta})$$

ならば, 分散の小さい推定量$\hat{\theta}$が望ましい

このとき, $\hat{\theta}$は$\bar{\theta}$よりも有効性をもつという.