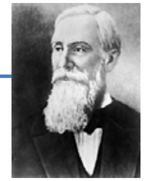
統計学入門 補助資料
～標本平均の性質(3つの定理)～

2022年度1学期:　　月曜2限
担当教員:　　　　石垣　司

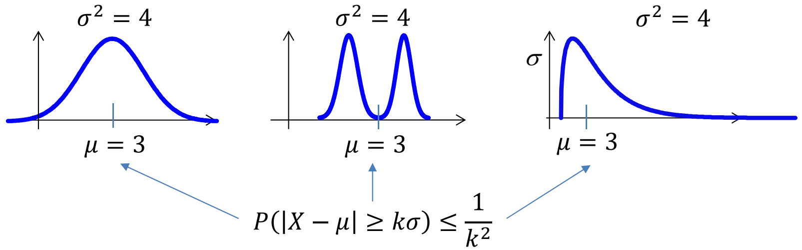
チェビシェフの不等式



- **<u>任意の確率分布</u>のある範囲内の確率の下限**

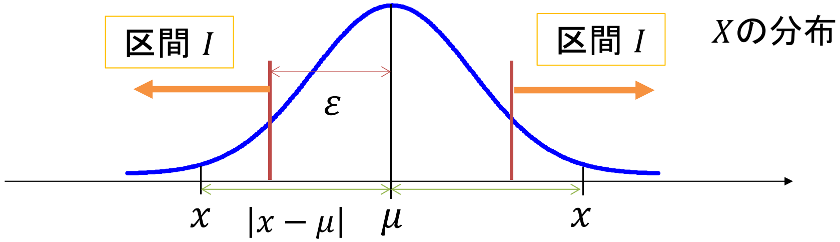
$$P(|X-\mu| \geq k\sigma) \leq \frac{1}{k^2}$$ check!

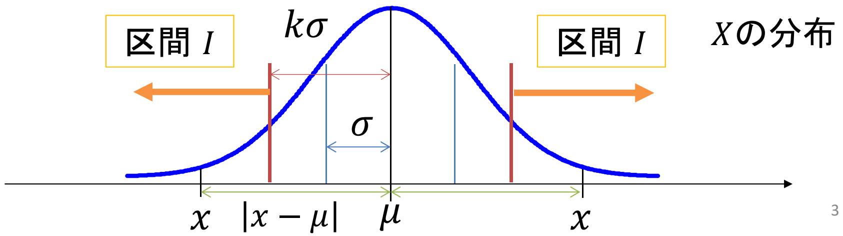
  - **平均 μ, 分散 σ^2 の確率分布(μ, σ^2 は有限)に従う確率変数 X が平均 μ から $k\sigma$ 以上ズレて出現する確率は $1/k^2$ 以下となる**
  - **分布の形が異なっていても上の不等式は成立**



チェビシェフの不等式の証明の補助図

- **区間 I での確率を考える**
- ε : **任意の正の定数, 区間** $I = \{|x-\mu| \geq \varepsilon\}$





チェビシェフの不等式の応用

- **例題**
  1. **平均 0, 分散 1 の任意の確率分布で, 確率変数が $-2 \leq X \leq 2$ となる確率の下限(最低値)を求めよ**
  2. **平均 0, 分散 1 の正規分布で, 確率変数が $-2 \leq X \leq 2$ となる確率を求めよ**
  3. **例題1と2の結果を比べ, その違いを議論せよ**

- **問題**
  - **ある新聞社が今までに発行した新聞内の各記事の文字数は平均が1000文字, 標準偏差200であった。このとき, ある一つの記事を抜き出したとき, 文字数が400文字以下, または1600文字以上となる確率の最大値を求めよ**

大数の法則

- **統計的推測の根幹となる定理**

$$\lim_{n\to\infty} P(|\bar{X}-\mu| \geq \varepsilon) = 0$$ check!

Jakob Bernoulli
(1654-1705)

  - **標本平均に対するチェビシェフの不等式**

– **サンプルサイズ n が無限大に近づくと, $\bar{X}$ は μ に限りなく近づくことを数学的に保証**

  - **(より正確な意味) 母平均と母分散が有限であれば, n を無限大に近づけていくことで, 標本平均 $\bar{X}$ と母平均 μ のズレの大きさが ε よりも大きくなる確率は 0 である**

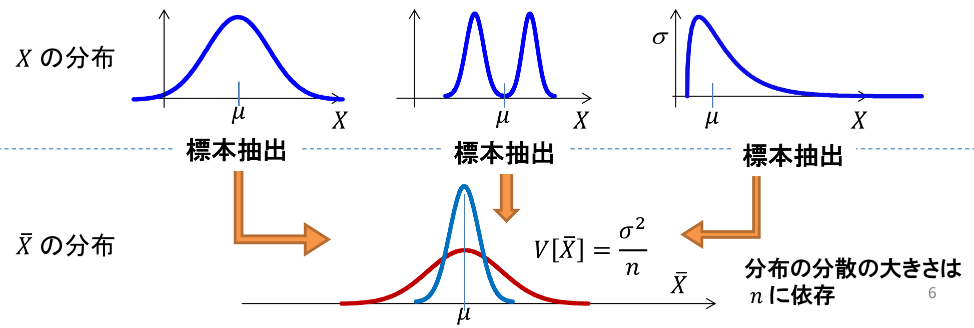
– **統計的確率と数学的確率の一致**

  - **例　サイコロを無限回投げ続ける(統計的確率)と 1 の目の出る割合は、理論値の 1/6 (数学的確率)に一致する**

中心極限定理

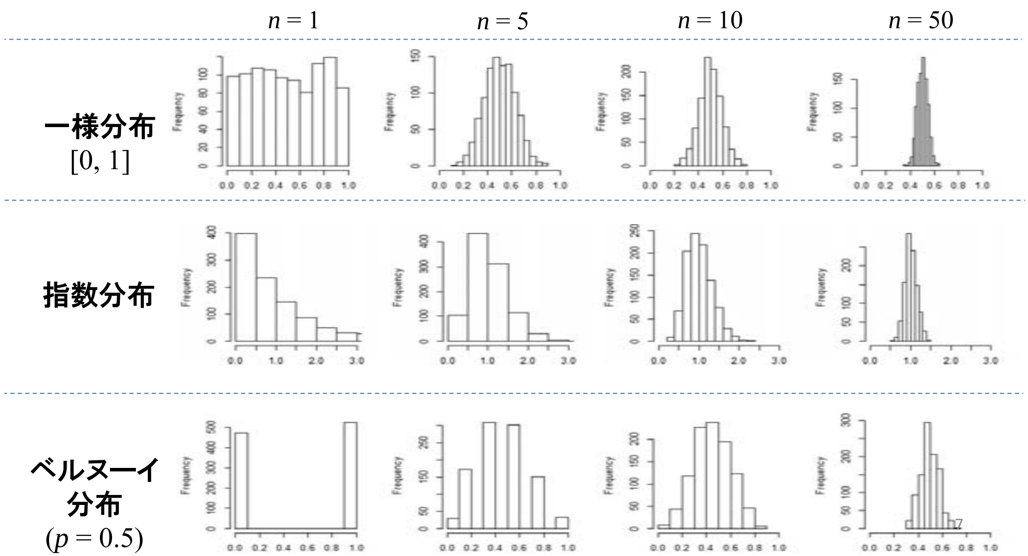
- **<u>任意の確率分布</u>(平均μ, 分散 σ^2 は有限)の標本平均の分布は, サイズ n が十分大きい時 $N\left(\mu, \frac{\sigma^2}{n}\right)$ に近似できる**

  - **統計学の定理の中で最も重要かつ非自明な定理の一つ**
  - **積率母関数(モーメント母関数)による証明** check!



中心極限定理のシミュレーション

– **確率分布(一様分布、指数分布、ベルヌーイ分布)から $n = \{1, 5, 10, 50\}$ の標本を1000回抽出。その標本平均値 $\bar{x}$ のヒストグラム**



中心極限定理の応用

- **問題**

– **サイコロを 18000 回投げ, 1 の目が 2850 回出た。その数は理論値よりも 150 回小さい。その誤差は許容できるか？**

  - **1の目の出る回数 $X = X_1 +, \ldots, +X_n$ は二項分布でモデル化できる**
  - $P(X \leq 2850) = \sum_{x=1}^{2850} {}_{18000}C_x \left(\frac{1}{6}\right)^x \left(\frac{5}{6}\right)^{18000-x}$
  - $X_i \sim Bel(p = 1/6), E[X_i] = p, V[X_i] = p(1-p)$

– **$\bar{X}$ をベルヌーイ試行の標本平均($\bar{X} = \frac{1}{n}(X_1 +, \ldots, +X_n)$)と考え, 中心極限定理より正規分布で近似する。このとき, $P(X \leq 2850)$ となる確率を求めよ**

  - **標準正規分布表より $P(Z \geq 3) = 0.00135$を用いる**

補足：積率母関数（モーメント母関数）#1

- **原点周りの m 次の積率を生成する関数**

$$M_X(t) = E[e^{tX}]$$

  - **積率（モーメント）**(期待値が存在する場合)
    - 1次の積率：平均, 2次の積率：分散, 3次の積率：(標準化)歪度, ・・・
    - 全ての次数の積率がわかれば, 確率分布の形もわかる
  - **確率分布と1対1の関係**
    - 例　正規分布 $N(\mu, \sigma^2) \Leftrightarrow M_X(t) = exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right)$
      - ただし, 積率母関数が存在しない分布もある。例：コーシー分布
  - **積率母関数のマクローリン展開**

$$M_X(t) = 1 + tE[X] + \frac{t^2}{2!}E[X^2] + \frac{t^3}{3!}E[X^3] + \cdots$$

補足：積率母関数（モーメント母関数）#2

- **中心極限定理の証明**
  - 標準化した標本平均 $\bar{Z} = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ の積率母関数 $M_{\bar{Z}}(t)$

$$\lim_{n\to\infty} M_{\bar{Z}}(t) = \exp\left(\frac{t^2}{2}\right)$$ check!

    - $n \to \infty$のとき, $M_{\bar{Z}}(t)$が標準正規分布の積率母関数に一致
- **正規分布の和の再生性の証明**
  - $X_1 \sim N(\mu_1, \sigma_1^2)$, $X_2 \sim N(\mu_2, \sigma_2^2)$で X_1 と X_2 は独立のとき, $Y = X_1 + X_2$の積率母関数 $M_Y(t)$

$$M_Y(t) = \exp\left\{(\mu_1 + \mu_2)t + (\sigma_1^2 + \sigma_2^2)\frac{t^2}{2}\right\}$$ check!

    - $M_Y(t)$が正規分布$N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$の積率母関数に一致

演習問題

- **コインを 10000 回投げたとき表の出る回数 X が5150回以上となる確率 $P(X \geq 5150)$ を中心極限定理による近似を用いて求めよ**