確率と統計(0)
「大数の法則と中心極限定理(第8章)」

- 担当教員： 杉山　将 (計算工学専攻)
- 居室： W8E-505
- 電子メール: sugi@cs.titech.ac.jp

- 授業のウェブサイト：
  http://sugiyama-www.cs.titech.ac.jp/~sugi/

独立同一分布

- 同じ分布から独立に n 個の標本

$$X_1, X_2, \ldots, X_n$$

を取り出したとき， これらは**独立同一分布に従う**(independently and identically distributed, **略して**i.i.d.)という

- $X_1, X_2, \ldots, X_n$ の同時確率密度関数は

$$f(x_1, x_2, \ldots, x_n) = g(x_1)g(x_2)\cdots g(x_n)$$

- 以下, i.i.d.標本の標本平均 $\overline{X}_n$ の振る舞いを調べる.

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

i.i.d確率変数の標本平均

- $X_1, X_2, \ldots, X_n$ がi.i.d.で期待値 μ, 分散 σ^2のときの標本平均 $\overline{X}_n$ の性質:
  - 期待値はそのまま:　$E(\overline{X}_n) = \mu$
  - 分散は $1/n$:　$V(\overline{X}_n) = \dfrac{\sigma^2}{n}$

標本平均を取れば安定する！

注意:ここでの期待値 E は, n 個全ての確率変数に対する期待値である

$$E[Z] = \int\int\cdots\int Zg(X_1)g(X_2)\cdots g(X_n)dX_1dX_2\cdots dX_n$$

証明

- $$E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n} E[X_i] = \frac{1}{n}\sum_{i=1}^{n} \mu$$
  $$= \mu$$

- $$V\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n^2}\sum_{i=1}^{n} V[X_i] = \frac{1}{n^2}\sum_{i=1}^{n} \sigma^2$$
  $$= \frac{\sigma^2}{n}$$

大数の法則

- **大数の法則**(law of large numbers)：任意の正の定数 ε に対して

$$P(|\overline{X}_n - \mu| < \varepsilon) \to 1$$

- 確率論の用語ではこれを，$\overline{X}_n$ が μ に**確率収束**(convergence in probability)するという

- **解釈**：標本を十分たくさん取れば，標本平均を母平均とみなしてもよい！

証明

- 各標本の期待値を μ ，分散を σ^2とすれば,

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

- ここで、チェビシェフの不等式を使うと

$$P(|\overline{X}_n - \mu| > \varepsilon) \leq \frac{\sigma^2/n}{\varepsilon^2}$$

- $n \to \infty$ のとき、右辺は常にゼロに収束する

(Q.E.D.)

チェビシェフの不等式

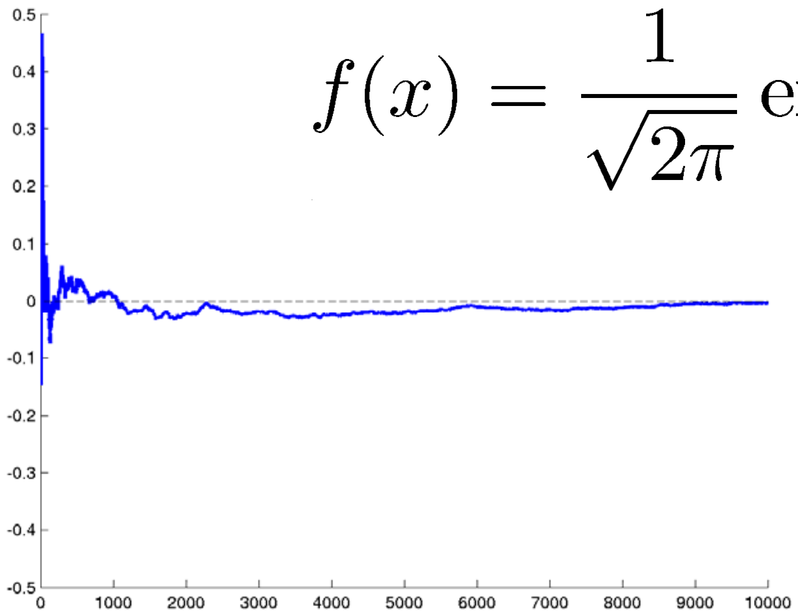
$$P(|X - E(X)| \geq k) \leq \frac{V(X)}{k^2}$$

大数の法則の例 (1)

- $\{X_i\}_{i=1}^{n}$ が、平均0、分散1の標準正規分布に独立に従うとき

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$$

$\overline{X}_n$



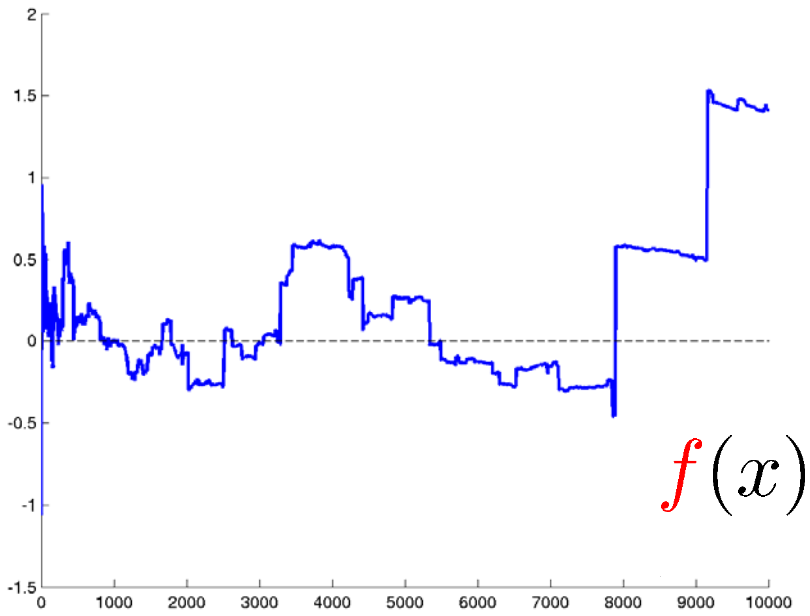
n

標本平均は確かに母平均0に収束していく

$$\overline{X}_n = (X_1 + X_2 + \cdots + X_n)/n$$

大数の法則の例 (2)

- $\{X_i\}_{i=1}^{n}$ が、“中心”0の
コーシー分布に独立に従うとき



$$f(x) = \frac{1}{\pi(1+x^2)}$$

標本平均は収束しない

⇒ **コーシー分布の平均は存在しない**

標本平均の分布

- 大数の法則から，標本平均が母平均に近づいていくことがわかった
- 大標本の極限の少し手前では，標本平均はどのように分布しているのであろうか？

中心極限定理

- 標本平均を標準化する：

$$Z_n = \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \qquad E[Z_n] = 0, \;\; V[Z_n] = 1$$

- 中心極限定理(central limit theorem)：

$n \to \infty$ のとき

$$P(a \leq Z_n \leq b) \to \int_a^b \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$$

中心極限定理

- 確率論の用語ではこれを, Z_n が標準正規分布に**法則収束**(convergence in law), または**分布収束**(convergence in distribution)するという

- また, Z_n は**漸近的**(asymptotically)に標準正規分布に従うという

- **中心極限定理の解釈**：大標本の極限では, もとの分布が何であろうと標本平均は大体, 平均 μ, 分散 σ^2/n の正規分布に従う

中心極限定理の証明

- 標準正規分布の積率母関数は $e^{t^2/2}$ なので，次式を示す

$$\lim_{n\to\infty} M_{Z_n}(t) = e^{t^2/2}$$

- $$Z_n = \frac{\sqrt{n}\,\overline{X}_n - \sqrt{n}\mu}{\sigma} = \frac{\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i - \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mu}{\sigma} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{X_i-\mu}{\sigma}$$

$$= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} Y_i$$

$$Y_i = \frac{X_i-\mu}{\sigma}$$

- Y_i は平均0，分散1なので，積率母関数は

$$M_{Y_i}(t) = 1 + \mu_1 t + \frac{\mu_2}{2!}t^2 + \frac{\mu_3}{3!}t^3 + \cdots$$

$$= 1 + \frac{1}{2}t^2 + \frac{\mu_3}{3!}t^3 + \cdots$$

$$\mu_1 = 0$$

$$\mu_2 = 1$$

中心極限定理の証明 (続き)

- Z_n の積率母関数は

$$
\begin{aligned}
M_{Z_n}(t) &= \left[M_{Y_i/\sqrt{n}}(t)\right]^n \\
&= \left[M_{Y_i}\left(\frac{t}{\sqrt{n}}\right)\right]^n \\
&= \left[1 + \frac{1}{2}\left(\frac{t}{\sqrt{n}}\right)^2 + \frac{\mu_3}{3!}\left(\frac{t}{\sqrt{n}}\right)^3 + \cdots\right]^n \\
&= (1+u)^n
\end{aligned}
$$

独立ならば

$$M_{Y_1+Y_2}(t) = M_{Y_1}(t)M_{Y_2}(t)$$

一般に

$$M_{aY}(t) = E(e^{taY}) = M_Y(at)$$

$$u = \frac{t^2}{2n} + \frac{\mu_3}{3!}\left(\frac{t}{\sqrt{n}}\right)^3 + \cdots$$

- 以下，$\lim_{n\to\infty} M_{Z_n}(t) = e^{t^2/2}$ の両辺の対数をとり

$$\lim_{n\to\infty}\left(\log M_{Z_n}(t) - t^2/2\right) = 0$$

を示す

中心極限定理の証明 (続き)

- $n \to \infty$ のとき, $|u| < 1$ なので次のようにテーラー展開できる

$$\log(1+u) = u - u^2/2 + u^3/3 - \cdots$$

- 従って $\log M_{Z_n}(t) - t^2/2 = n\log(1+u) - t^2/2$

$$= n(u - u^2/2 + u^3/3 - \cdots) - t^2/2$$

- $nu = \frac{t^2}{2} + \frac{\mu_3}{3!}\frac{t^3}{n^{1/2}} + \cdots$ より

$$\lim_{n\to\infty} nu = t^2/2$$

- また

$$\lim_{n\to\infty} nu^q = 0 \quad (q \geq 2)$$

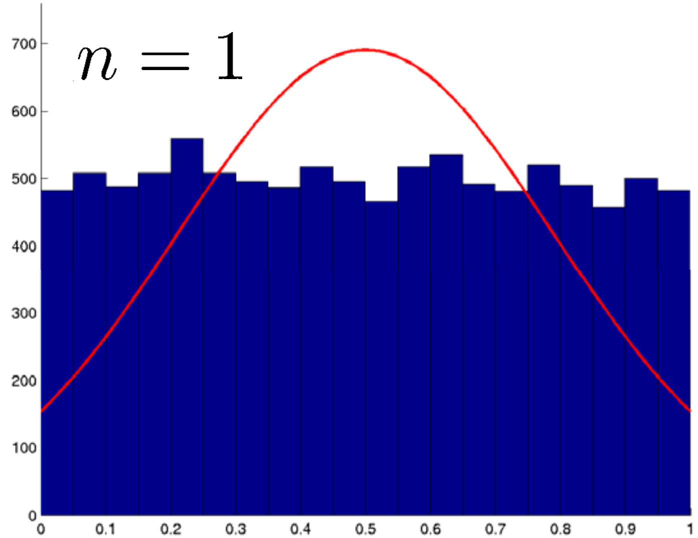
- 従って

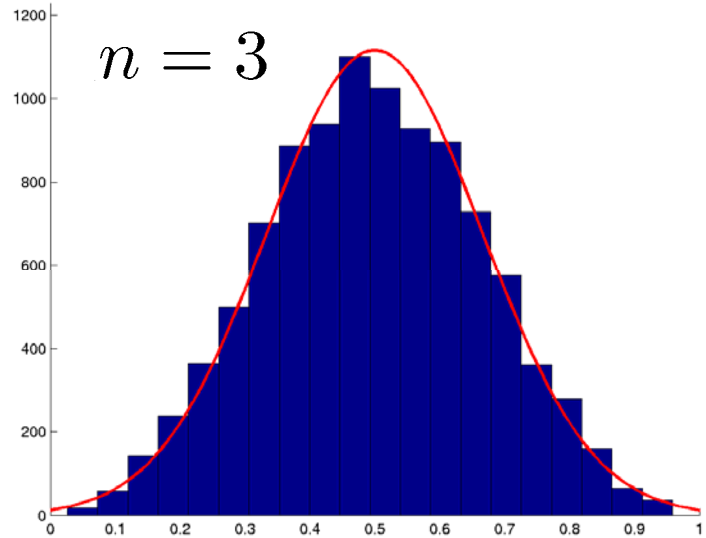
$$\lim_{n\to\infty} (\log M_{Z_n}(t) - t^2/2) = 0$$

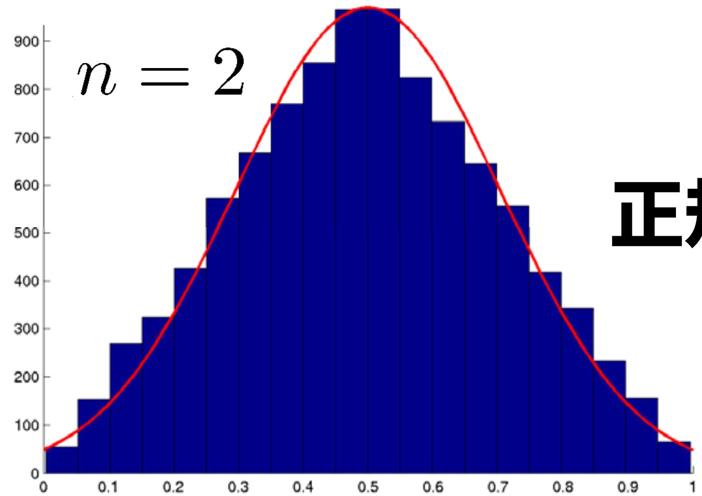
(Q.E.D.)

中心極限定理の例 (1)

- (0, 1) 上の一様分布



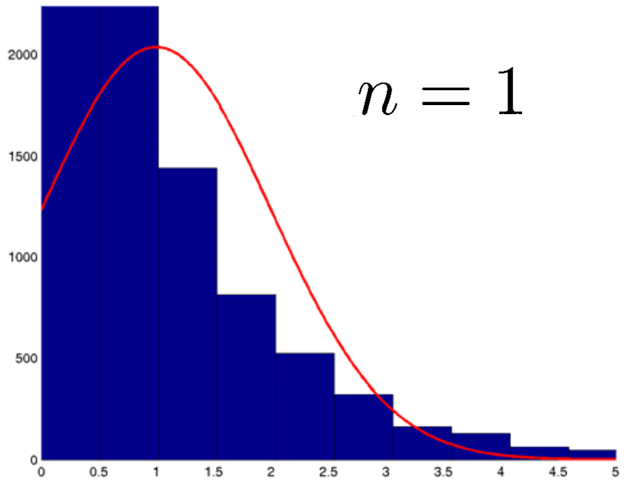


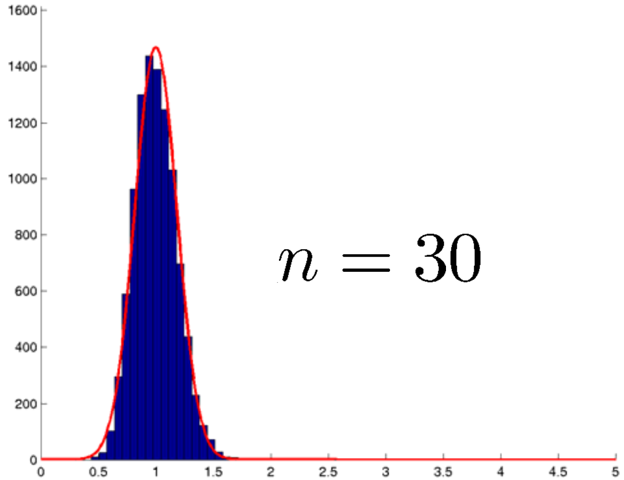


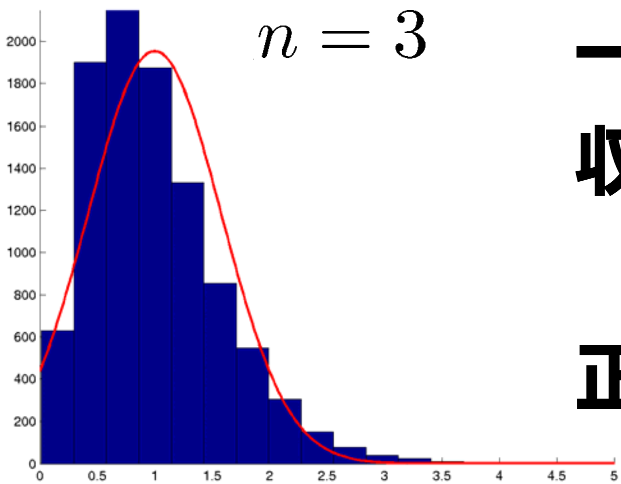
$n=2,3$ 位で
正規分布に似てくる

中心極限定理の例 (2)

■ 指数分布　$f(x) = e^{-x}$



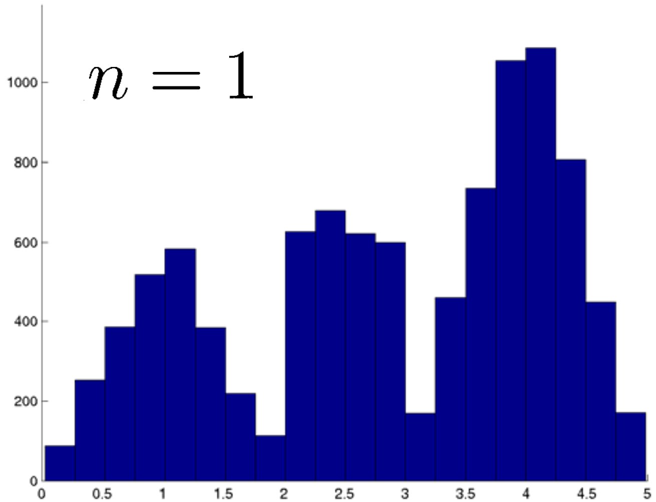


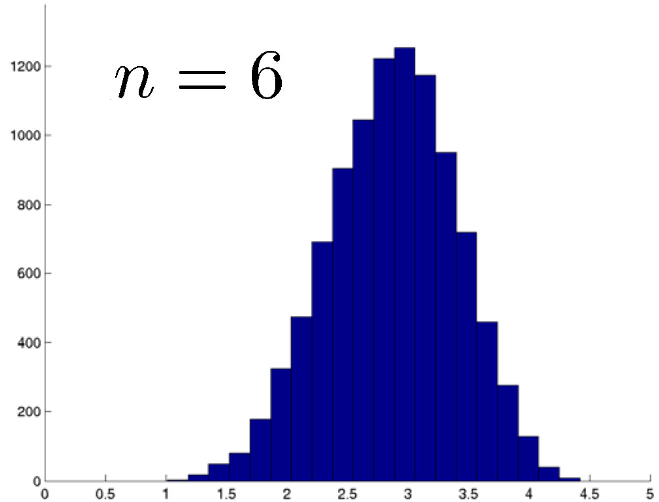


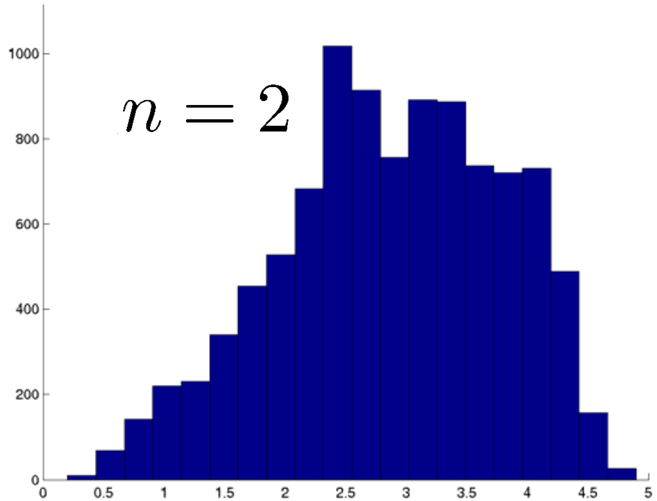
一様分布の場合より
収束が若干遅いが,
$n = 30$ 位で
正規分布に似てくる

中心極限定理の例 (3)

■ 適当な分布







小レポート

1. Octaveなどを用いて，**一様分布**に独立に従う確率変数を生成し，標本数を増やしていったときの，**標本平均のグラフ**を作成し，大数の法則が成り立つことを確認せよ．
2. 同様に，標本数を増やしていったときの**標本平均の分布ヒストグラム**（つまり，n 個の標本を生成し標本平均を求めるという過程を何度も繰り返し，その分布をプロットする）を作成し，中心極限定理が成り立つことを確認せよ．

小レポート(続き)

3. コーシー分布に従う確率変数に対して1，2と同様の実験を行い，大数の法則，および，中心極限定理が成りたたないことを確認せよ．

   ヒント：標準正規分布に独立に従う確率変数 X, Y の比 X/Y はコーシー分布

   $$f(x) = \frac{1}{\pi(1+x^2)}$$

   に従う

4. 以前の課題で自作した独自の確率分布に従う確率変数に対して，1，2と同様の実験を行ない，大数の法則，および，中心極限定理が成り立つかどうか確認せよ．