

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/>

- 同じ分布から独立に n 個の標本

$$X_1, X_2, \dots, X_n$$

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

- $X_1, X_2, \dots, X_n$ の同時確率密度関数は

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

i.i.d確率変数の標本平均

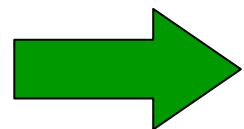
233

■ $X_1, X_2, \dots, X_n$: 期待値 μ , 分散 σ^2 のi.i.d.標本

■ 標本平均 $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ の性質 (証明は演習) :

● 期待値は変わらない: $E(\bar{X}_n) = \mu$

● 分散は $1/n$ になる: $V(\bar{X}_n) = \frac{\sigma^2}{n}$

 標本平均を取れば値が安定する！

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

$$E[Z] = \int \int \cdots \int Z g(X_1) g(X_2) \cdots g(X_n) dX_1 dX_2 \cdots dX_n$$

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

$$P(|\bar{X}_n - \mu| < \varepsilon) \rightarrow 1$$

- 確率論の用語ではこれを, $\bar{X}_n$ 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!

- 各標本の期待値を μ , 分散を σ^2 とすれば,

$$E(\bar{X}_n) = \mu$$

$$V(\bar{X}_n) = \frac{\sigma^2}{n}$$

- ここで、チェビシェフの不等式を使うと

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

- $n \rightarrow \infty$ のとき、右辺は常にゼロに収束する

(Q.E.D.)

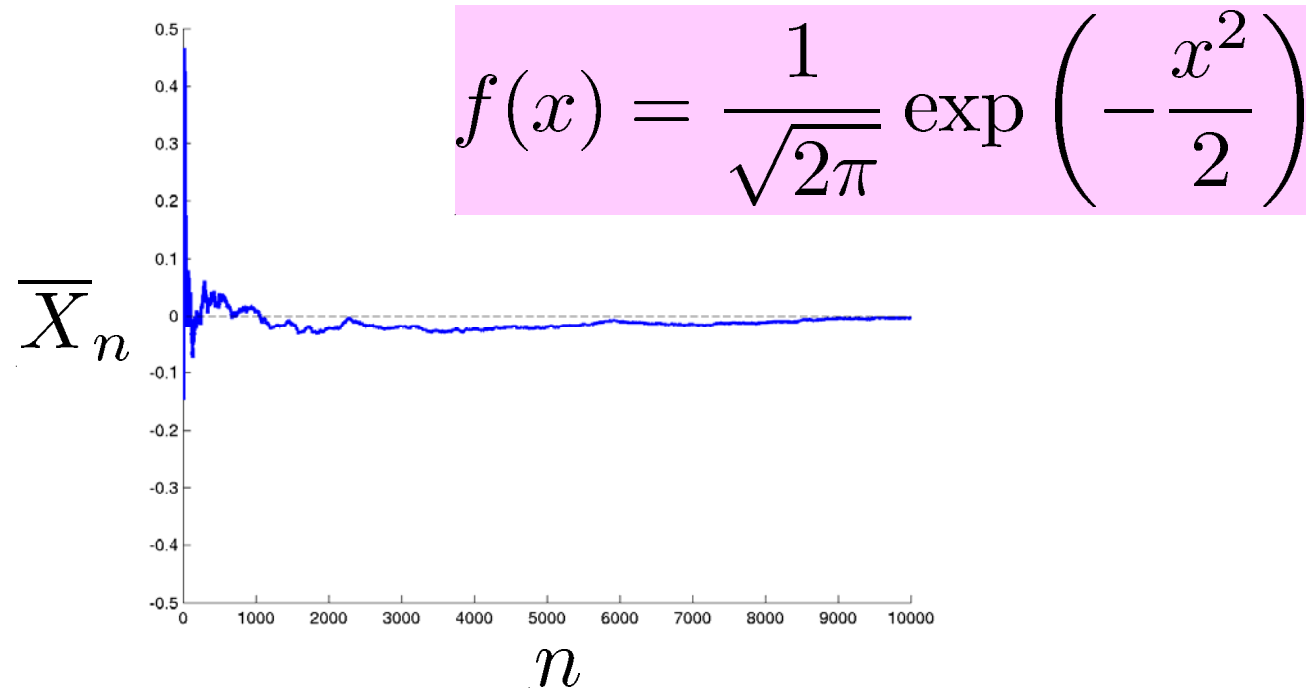
チェビシェフの不等式

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

大数の法則の例 (1)

236

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



標本平均は確かに母平均0に収束していく

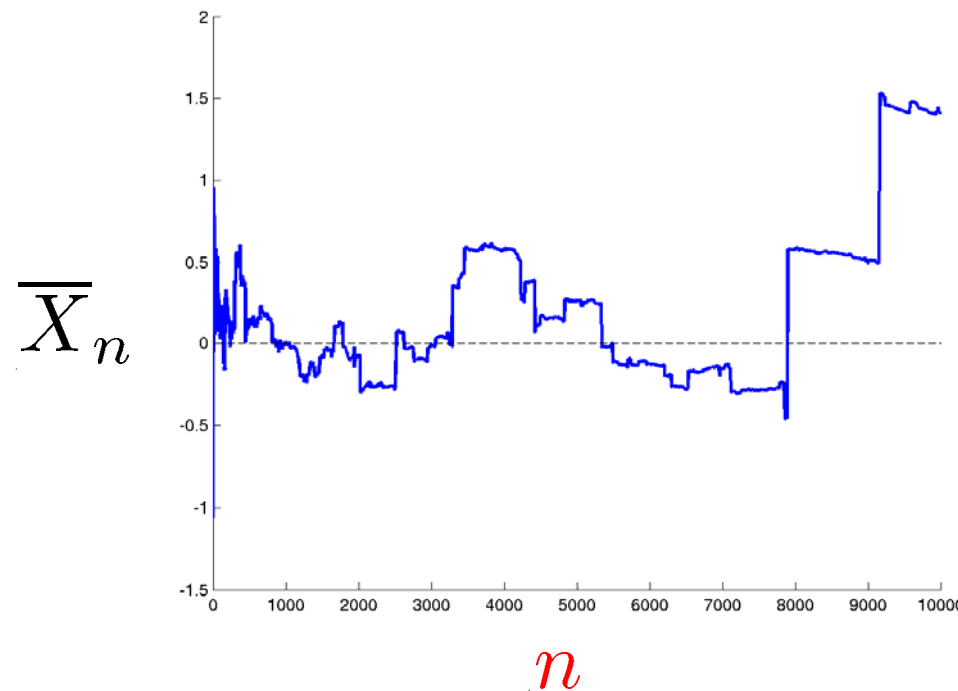
$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$$

大数の法則の例 (2)

237

- $\{X_i\}_{i=1}^n$ 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



コーシー分布の期待値は存在しない

標本平均の分布

238

- 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しょうか？

中心極限定理

239

- 標本平均を標準化する:

$$Z_n = \frac{\bar{X}_n - \mu}{\sigma / \sqrt{n}}$$

$$E[Z_n] = 0, \quad V[Z_n] = 1$$

- 中心極限定理(central limit theorem):

$n \rightarrow \infty$ のとき

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

中心極限定理 (続き)

240

- 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もいう.

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 $\bar{X}_n$ は,

平均 μ , 分散 σ^2/n の正規分布に大体従う

中心極限定理の証明

241

- 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

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

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

- Y_i は平均0, 分散1なので、積率母関数は

$$M_{Y_i}(t) = 1 + \mu_1 t + \frac{\mu_2}{2!} t^2 + \frac{\mu_3}{3!} t^3 + \dots$$
$$= 1 + \frac{1}{2} t^2 + \frac{\mu_3}{3!} t^3 + \dots$$
$$\mu_1 = 0$$
$$\mu_2 = 1$$

中心極限定理の証明 (続き)

242

■ Z_n の積率母関数は

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

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, $\lim_{n \rightarrow \infty} M_{Z_n}(t) = e^{t^2/2}$ の両辺の対数を取り

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

を示す

中心極限定理の証明 (続き)

243

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

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

- 従って $\log M_{Z_n}(t) - t^2/2 = n \log(1 + u) - t^2/2$
 $= n(u - u^2/2 + u^3/3 - \dots) - t^2/2$

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

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

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

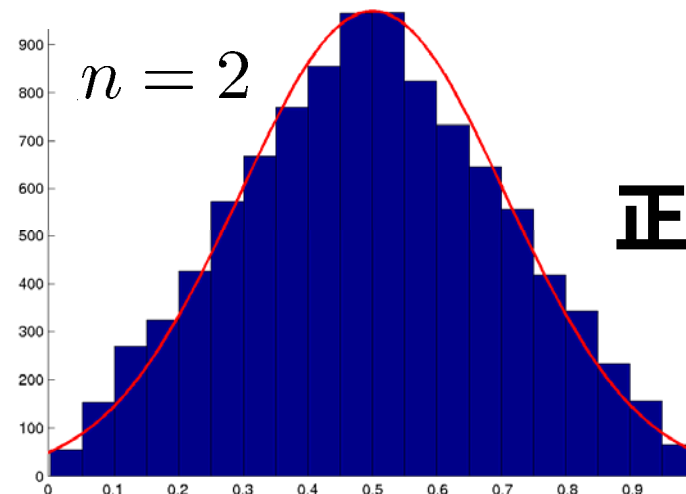
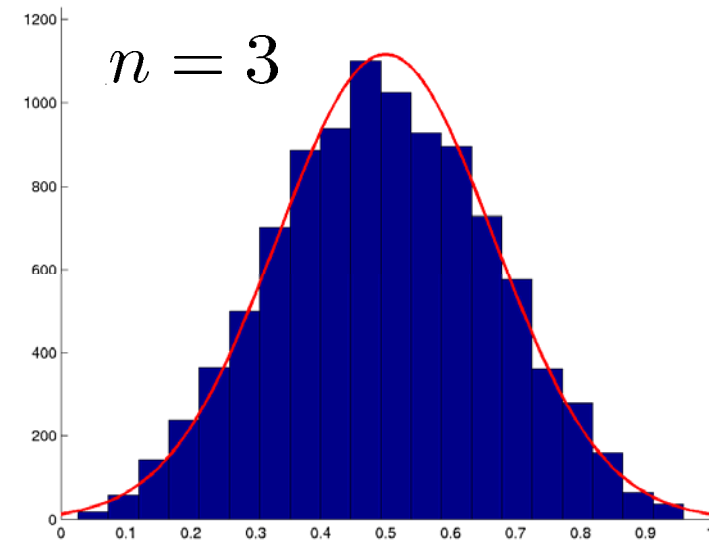
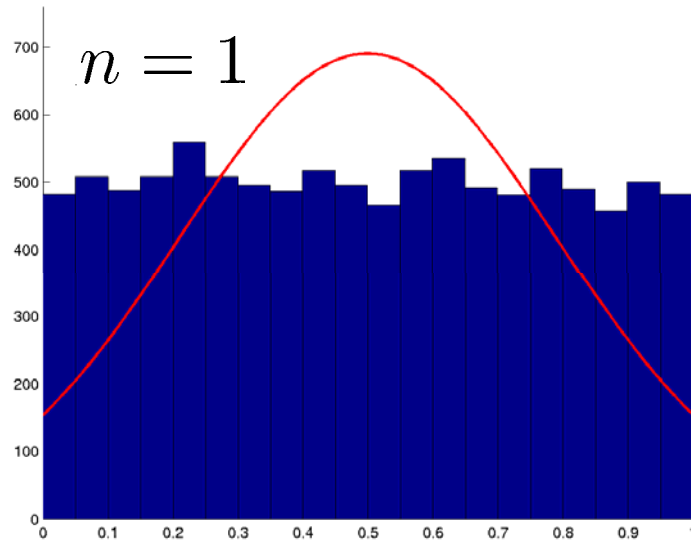
- 従って $\lim_{n \rightarrow \infty} (\log M_{Z_n}(t) - t^2/2) = 0$

(Q.E.D.)

中心極限定理の例 (1)

244

■ $(0, 1)$ 上の一様分布



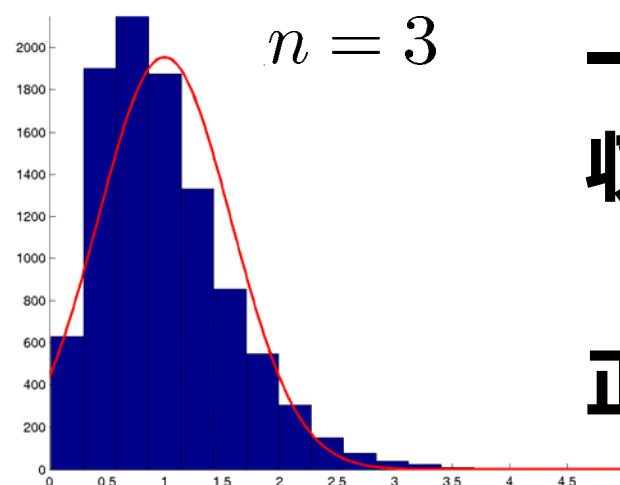
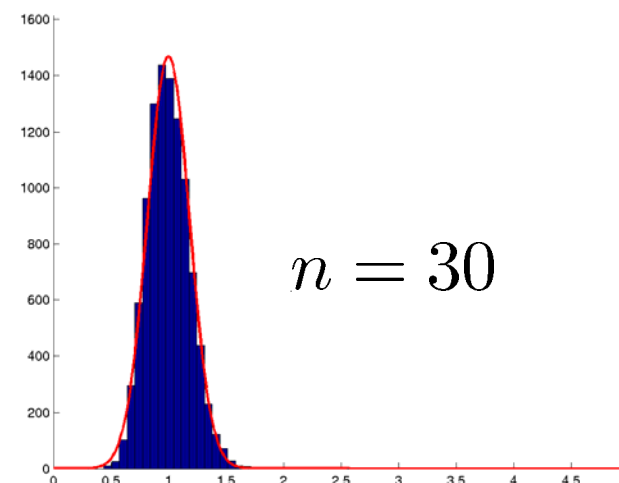
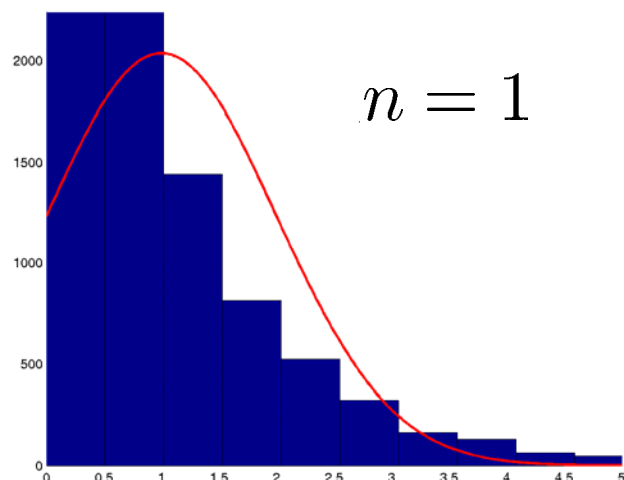
$n = 2, 3$ 位で
正規分布に似てくる

中心極限定理の例 (2)

245

■ 指数分布

$$f(x) = e^{-x}$$

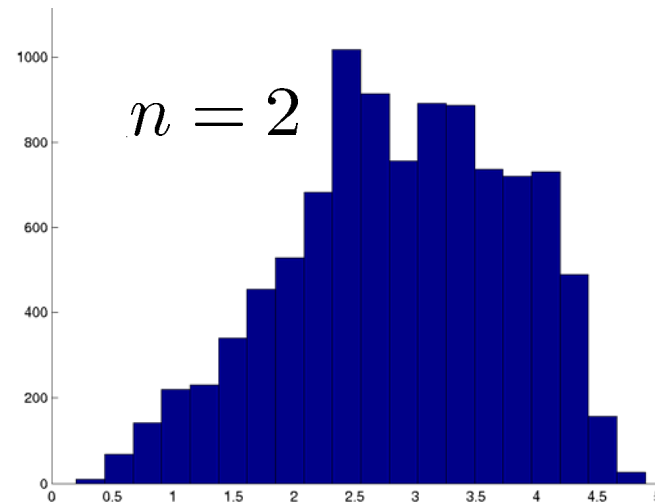
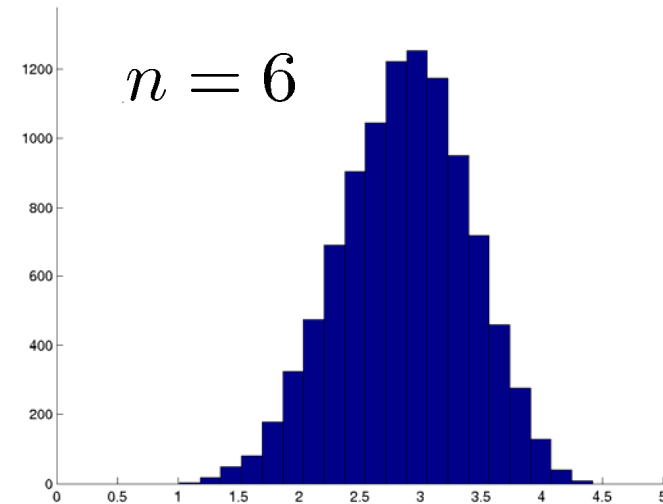
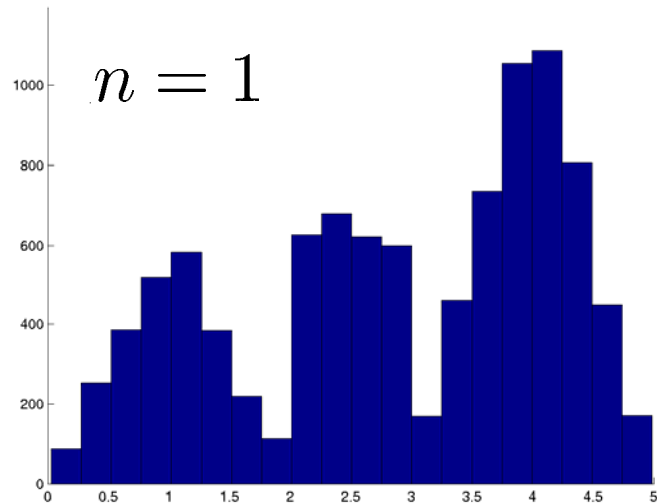


一様分布の場合より
収束が若干遅いが、
 $n = 30$ 位で
正規分布に似てくる

中心極限定理の例 (3)

246

■ 適当な分布



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。

宿題 (続き)

248

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**成り立たないこと**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.

試験について

249

- 日時: 7月27日 (金) 3,4限
- 場所: S125
- 試験内容:
 - 専門用語の英語名
 - 記述問題 (選択式2問)
- 教科書, ノートは持ち込み不可!