

離散数理工学 第 10 回
離散確率論：乱択データ構造とアルゴリズム（発展）

岡本 吉央
okamotoy@uec.ac.jp

電気通信大学

2019 年 1 月 17 日

最終更新：2019 年 1 月 15 日 14:35

岡本 吉央 (電通大)

離散数理工学 (10)

2019 年 1 月 17 日 1 / 30

スケジュール 前半

- | | |
|------------------------|---------|
| ■ 数え上げの基礎：二項係数と二項定理 | (10/4) |
| ★ 休講 | (10/11) |
| ■ 数え上げの基礎：漸化式の立て方 | (10/18) |
| ■ 数え上げの基礎：漸化式の解き方 (基礎) | (10/25) |
| ★ 休講 | (11/1) |
| ■ 数え上げの基礎：漸化式の解き方 (発展) | (11/8) |
| ■ 離散代数：対称群と置換群 | (11/15) |
| ★ 休講 | (11/22) |
| ■ 離散代数：有限群 | (11/29) |
| ■ 離散代数：有限群の応用 | (12/6) |

スケジュール 後半 (予定)

- | | |
|--------------------------------|---------|
| ■ 8 離散確率論：確率的離散システムの解析 | (12/13) |
| ★ 中間試験 | (12/20) |
| ■ 9 離散確率論：乱択データ構造とアルゴリズム (基礎) | (1/10) |
| ■ 10 離散確率論：乱択データ構造とアルゴリズム (発展) | (1/17) |
| ■ 11 離散確率論：マルコフ連鎖 (基礎) | (1/24) |
| ■ 12 離散確率論：マルコフ連鎖 (発展) | (1/31) |
| ■ 13 離散確率論：エンタロピ | (2/7?) |
| ★ 期末試験 | (2/14?) |

注意：予定の変更もありうる

岡本 吉央 (電通大)

離散数理工学 (10)

2019 年 1 月 17 日 3 / 30

岡本 吉央 (電通大)

離散数理工学 (10)

2019 年 1 月 17 日 2 / 30

今日の目標

- 今日の目標
- 典型的な乱択アルゴリズムの設計と解析ができるようになる
- ▶ 確率の推定 (中央値トリック)

目次

① 確率の推定：単純なアルゴリズム

② 確率の推定：中央値トリック

③ 今日のまとめ

岡本 吉央 (電通大)

離散数理工学 (10)

2019 年 1 月 17 日 5 / 30

不公平な硬貨

設定

- ▶ 考えている硬貨について

$$\Pr(\text{表}) = p$$

ただし, $0 \leq p \leq 1$

- ▶ **目標** : p を知りたい

岡本 吉央 (電通大)

離散数理工学 (10)

2019 年 1 月 17 日 7 / 30

岡本 吉央 (電通大)

離散数理工学 (10)

2019 年 1 月 17 日 4 / 30

確率の推定：単純なアルゴリズム

不公平な硬貨

設定

- ▶ 硬貨が 1 つある
- ▶ 投げたとき, 表が出る確率はいつも変わらない
- ▶ その確率が分からぬ
- ▶ **目標** : 表が出る確率を知りたい
- ▶ 可能な操作 : 硬貨を投げる (ことのみ)

岡本 吉央 (電通大)

離散数理工学 (10)

2019 年 1 月 17 日 6 / 30

確率の推定：単純なアルゴリズム

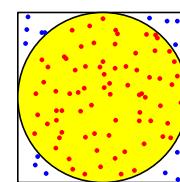
応用例：モンテカルロ法

モンテカルロ法：(実際には用いられない) 例

円周率の計算のために次を行なう

- | | |
|--|--|
| ■ [−1, 1] ² 内の点 (x, y) を一様分布に従って発生させる | |
| ■ $x^2 + y^2 \leq 1$ ならば, 1 を出力, そうでなければ 0 を出力 | |
| このとき, この方法が 1 を出力する確率 = $\pi/4$ | |

つまり, $p = \pi/4$ とした不公平な硬貨を考えていることになる



モンテカルロ法は、次の「単純なアルゴリズム」を実行する

岡本 吉央 (電通大)

離散数理工学 (10)

2019 年 1 月 17 日 8 / 30

単純なアルゴリズム

単純なアルゴリズム

硬貨を何度も投げてみる

- ▶ n 回投げるとする (独立な試行)
- ▶ 確率変数 X_i を次で定義 ($i \in \{1, \dots, n\}$)

$$X_i = \begin{cases} 0 & (i \text{回目に投げたとき裏が出る}) \\ 1 & (i \text{回目に投げたとき表が出る}) \end{cases}$$

- ▶ 次の量を出力

$$\frac{X_1 + X_2 + \dots + X_n}{n}$$

岡本 吉央 (電通大)

離散数理工学 (10)

2019年1月17日

9 / 30

誤差の解析 (1)

以後, $X = X_1 + X_2 + \dots + X_n$ とする

- ▶ 真の値 p から出力 $\frac{X}{n}$ がどれだけずれるか?
- ▶ そのずれが ε 未満である確率を知りたい
- ▶ その確率は次のように書ける

$$\Pr\left(\left|\frac{X}{n} - p\right| < \varepsilon\right)$$

- ▶ 計算

$$\begin{aligned} \Pr\left(\left|\frac{X}{n} - p\right| < \varepsilon\right) &= 1 - \Pr\left(\left|\frac{X}{n} - p\right| \geq \varepsilon\right), \\ \Pr\left(\left|\frac{X}{n} - p\right| \geq \varepsilon\right) &\leq \frac{\mathbb{E}[|X/n - p|]}{\varepsilon} \quad (\text{マルコフの不等式}) \end{aligned}$$

しかし, $\mathbb{E}[|X/n - p|]$ はどう計算したらいいか分からない

岡本 吉央 (電通大)

離散数理工学 (10)

2019年1月17日

11 / 30

誤差の解析 (3)

$$\mathbb{E}\left[\left|\frac{X}{n} - p\right|^2\right] = \mathbb{E}\left[\left(\frac{X}{n}\right)^2 - 2p\frac{X}{n} + p^2\right] = \frac{1}{n^2}\mathbb{E}[X^2] - \frac{2p}{n}\mathbb{E}[X] + p^2$$

ここで,

$$\begin{aligned} \mathbb{E}[X] &= \mathbb{E}[X_1 + \dots + X_n] = \sum_{i=1}^n \mathbb{E}[X_i] = pn \\ \mathbb{E}[X^2] &= \mathbb{E}[(X_1 + \dots + X_n)^2] = \sum_{i=1}^n \mathbb{E}[X_i^2] + \sum_{i=1}^n \sum_{j=1, (i \neq j)}^n \mathbb{E}[X_i X_j] \end{aligned}$$

岡本 吉央 (電通大)

離散数理工学 (10)

2019年1月17日

13 / 30

誤差の解析 (5)

ここまで、まとめると

$$\begin{aligned} \mathbb{E}\left[\left|\frac{X}{n} - p\right|^2\right] &= \frac{1}{n^2}\mathbb{E}[X^2] - \frac{2p}{n}\mathbb{E}[X] + p^2 \\ &= \frac{1}{n^2}(pn + p^2n(n-1)) - \frac{2p}{n}pn + p^2 \\ &= \frac{p}{n} + \frac{p^2(n-1)}{n} - p^2 \\ &= \frac{p - p^2}{n} \\ &= \frac{p(1-p)}{n} \end{aligned}$$

岡本 吉央 (電通大)

離散数理工学 (10)

2019年1月17日

15 / 30

期待値の解析

- ▶ 出力の期待値は

$$\begin{aligned} \mathbb{E}\left[\frac{X_1 + X_2 + \dots + X_n}{n}\right] &= \frac{1}{n} \sum_{i=1}^n \mathbb{E}[X_i] \\ &= \frac{1}{n} \sum_{i=1}^n (0 \cdot (1-p) + 1 \cdot p) \\ &= p \end{aligned}$$

期待値は正しい「推測」になっている

問題点

必ず「 p 」を出力するわけではない ⇔ 誤差が出る n を大きくすれば、誤差は小さくなりそう

岡本 吉央 (電通大)

離散数理工学 (10)

2019年1月17日

10 / 30

誤差の解析 (2)

$$\begin{aligned} \Pr\left(\left|\frac{X}{n} - p\right| \geq \varepsilon\right) &= \Pr\left(\left|\frac{X}{n} - p\right|^2 \geq \varepsilon^2\right) \leq \frac{\mathbb{E}\left[\left|\frac{X}{n} - p\right|^2\right]}{\varepsilon^2} \\ \mathbb{E}\left[\left|\frac{X}{n} - p\right|^2\right] &\text{を計算してみる} \end{aligned}$$

岡本 吉央 (電通大)

離散数理工学 (10)

2019年1月17日

12 / 30

誤差の解析 (4)

任意の $i \in \{1, \dots, n\}$ に対して

$$\mathbb{E}[X_i^2] = (1-p) \cdot 0^2 + p \cdot 1^2 = p$$

したがって,

$$\sum_{i=1}^n \mathbb{E}[X_i^2] = pn$$

任意の異なる $i, j \in \{1, \dots, n\}$ に対して, X_i と X_j は独立なので,

$$\mathbb{E}[X_i X_j] = \mathbb{E}[X_i] \mathbb{E}[X_j] = p \cdot p = p^2$$

したがって,

$$\sum_{i=1}^n \sum_{j=1, (i \neq j)}^n \mathbb{E}[X_i X_j] = p^2 n(n-1)$$

岡本 吉央 (電通大)

離散数理工学 (10)

2019年1月17日

14 / 30

誤差の解析 (6)

すなわち,

$$\Pr\left(\left|\frac{X}{n} - p\right| \geq \varepsilon\right) \leq \frac{\mathbb{E}\left[\left|\frac{X}{n} - p\right|^2\right]}{\varepsilon^2} = \frac{1}{\varepsilon^2} \frac{p(1-p)}{n}$$

- ▶ この不等式は「**チェビシェフの不等式**」と呼ばれる (ものの特殊な場合)

- ▶ この右辺を δ 以下にするには, $n \geq \frac{1}{\delta^2} \frac{p(1-p)}{\delta}$ とすればよい

結論

誤差が ε 以上になる確率を δ 以下とするためには,

$$n \geq \frac{1}{\varepsilon^2} \frac{p(1-p)}{\delta}$$

とすればよい

岡本 吉央 (電通大)

離散数理工学 (10)

2019年1月17日

16 / 30

疑問

単純なアルゴリズム

硬貨を何度も投げてみる

- ▶ n 回投げるとする (独立な試行)
- ▶ 確率変数 X_i を次で定義 ($i \in \{1, \dots, n\}$) (標示確率変数)

$$X_i = \begin{cases} 1 & (i \text{回目に投げたとき表が出る}) \\ 0 & (i \text{回目に投げたとき裏が出る}) \end{cases}$$

- ▶ 次の量を出力

$$\frac{X_1 + X_2 + \dots + X_n}{n}$$

疑問

この「単純なアルゴリズム」よりもよいアルゴリズムは無いのか？

岡本 吉央 (電通大)

離散数理工学 (10)

2019年1月17日 17 / 30

中央値トリック (median trick)

中央値トリック

- ▶ n 回投げるとする (独立な試行)
- ▶ $n = (2k - 1)t$ とする (k, t は自然数)
- ▶ 確率変数 X_i を次で定義 ($i \in \{1, \dots, n\}$)

$$X_i = \begin{cases} 0 & (i \text{回目に投げたとき裏が出る}) \\ 1 & (i \text{回目に投げたとき表が出る}) \end{cases}$$

- ▶ 確率変数 Y_j を次で定義 ($j \in \{1, \dots, 2k - 1\}$)

$$Y_j = \frac{X_{(j-1)t+1} + \dots + X_{(j-1)t+t}}{t}$$

- ▶ 次の量を出力

$$Y = \text{med}\{Y_1, \dots, Y_{2k-1}\}$$

med は中央値 : $\text{med}\{5, 1, 6, 2, 4\} = 4$

岡本 吉央 (電通大)

離散数理工学 (10)

2019年1月17日 19 / 30

中央値トリック：誤差の解析 (2)

- ▶ このとき、合併上界から

$$\Pr(k \text{ 個の } j \text{ に対して}, |Y_j - p| \geq \varepsilon) \leq \binom{2k-1}{k} \left(\frac{1}{8}\right)^k$$

- ▶ 二項係数に対する上界を用いて、右辺を整理すると

$$\binom{2k-1}{k} \left(\frac{1}{8}\right)^k \leq \left(\frac{e(2k-1)}{k}\right)^k \left(\frac{1}{8}\right)^k < \left(\frac{2e}{8}\right)^k < \left(\frac{3}{4}\right)^k$$

二項係数：簡単な評価

(第1回講義より)

任意の自然数 $a \geq 1$ と任意の自然数 $b \geq 1$ に対して、 $a \geq b$ であるとき、

$$\left(\frac{a}{b}\right)^b \leq \binom{a}{b} \leq \left(\frac{ea}{b}\right)^b$$

岡本 吉央 (電通大)

離散数理工学 (10)

2019年1月17日 21 / 30

中央値トリック：誤差の解析 (まとめ)

中央値トリック：誤差の解析 (まとめ)

- ▶ $t \geq \frac{8p(1-p)}{\varepsilon^2}$, $k \geq \log_{3/4} \delta$ とすると
誤差が ε 以上になる確率を δ 以下にできる
- ▶ このとき、硬貨を投げる回数 n は

$$n = (2k-1)t \geq \Omega\left(\frac{p(1-p)}{\varepsilon^2} \log \frac{1}{\delta}\right)$$

補足：単純なアルゴリズムにて、硬貨を投げる回数 n は

$$n \geq \frac{p(1-p)}{\varepsilon^2} \frac{1}{\delta}$$

つまり、中央値トリックにより、硬貨を投げる回数が減った

目次

① 確率の推定：単純なアルゴリズム

② 確率の推定：中央値トリック

③ 今日のまとめ

岡本 吉央 (電通大)

離散数理工学 (10)

2019年1月17日 18 / 30

中央値トリック：誤差の解析 (1)

- ▶ 次が成り立つために n が満たす条件を見つけて

$$\Pr(|Y - p| \geq \varepsilon) \leq \delta$$

- ▶ 今までの議論から、任意の $j \in \{1, \dots, 2k-1\}$ に対して

$$\Pr(|Y_j - p| \geq \varepsilon) \leq \frac{1}{\varepsilon^2} \frac{p(1-p)}{t}$$

- ▶ $t \geq \frac{8p(1-p)}{\varepsilon^2}$ とすると、 $\frac{1}{t} \leq \frac{1}{8} \frac{\varepsilon^2}{p(1-p)}$ ので、

$$\Pr(|Y_j - p| \geq \varepsilon) \leq \frac{1}{8}$$

岡本 吉央 (電通大)

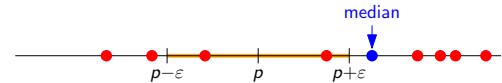
離散数理工学 (10)

2019年1月17日 20 / 30

中央値トリック：誤差の解析 (3)

- ▶ したがって (演習問題 11.3 参照),

$$\begin{aligned} \Pr(|Y - p| \geq \varepsilon) &\leq \Pr(k \text{ 個の } j \text{ に対して}, |Y_j - p| \geq \varepsilon) \\ &< \left(\frac{3}{4}\right)^k \end{aligned}$$



- ▶ $k \geq \log_{3/4} \delta$ とすると

$$\Pr(|Y - p| \geq \varepsilon) < \left(\frac{3}{4}\right)^k \leq \delta$$

岡本 吉央 (電通大)

離散数理工学 (10)

2019年1月17日 22 / 30

実験してみた

- ▶ パラメータ

$$p = 0.42$$

$$t = 100$$

k を変化させることで、 $n = (2k-1)t$ を変化させる

▶ Ruby 2.1.4 で実装

▶ n の増加に関する推定値の変化を図示してみた▶ 横軸が n , 縦軸が推定した p の値

注意：このパラメータ設定はとても恣意的なので、他のパラメータ設定で追試をしてみるとよい

岡本 吉央 (電通大)

離散数理工学 (10)

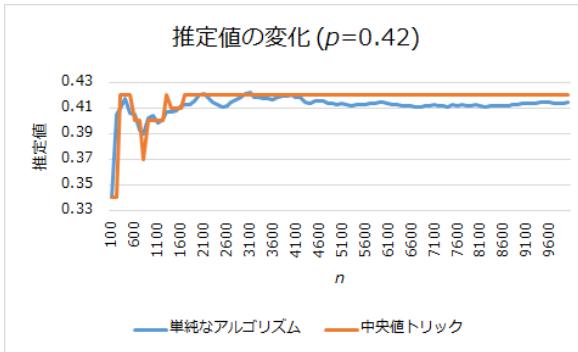
2019年1月17日 23 / 30

岡本 吉央 (電通大)

離散数理工学 (10)

2019年1月17日 24 / 30

実験してみた：結果



岡本 吉央 (電通大)

離散数理工学 (10)

2019年1月17日

25 / 30

中央値トリック：注意

注意

- ▶ 単純なアルゴリズムの出力 X/n に対して, $E[X/n] = p$ が成り立つ
- ▶ 中央値トリックの出力 Y に対して, $E[Y] = p$ が成り立つとは限らない

 X/n は不偏推定量であるが, Y はそうではない

どういうことか？

- ▶ n を大きくすると, X/n は p に近づいていく
- ▶ n を大きくすると, Y は p の近似値に近づいていく

岡本 吉央 (電通大)

離散数理工学 (10)

2019年1月17日

25 / 30

岡本 吉央 (電通大)

離散数理工学 (10)

2019年1月17日

26 / 30

今日のまとめ

目次

① 確率の推定：単純なアルゴリズム

② 確率の推定：中央値トリック

③ 今日のまとめ

今日のまとめ

今日の目標

今日の目標

典型的な乱択アルゴリズムの設計と解析ができるようになる

- ▶ 確率の推定 (中央値トリック)

岡本 吉央 (電通大)

離散数理工学 (10)

2019年1月17日

27 / 30

岡本 吉央 (電通大)

離散数理工学 (10)

2019年1月17日

28 / 30

今日のまとめ

残った時間の使い方

- ▶ 演習問題をやる
 - ▶ 相談推奨 (ひとりでやらない)
- ▶ 質問をする
 - ▶ 教員と TA は巡回
- ▶ 退室時, 小さな紙に感想などを書いて提出する ← 重要
 - ▶ 内容は何でも OK
 - ▶ 匿名で OK

岡本 吉央 (電通大)

離散数理工学 (10)

2019年1月17日

29 / 30