

I482F 実践的アルゴリズム特論 (13) 乱択アルゴリズム：乱択による高速化

岡本 吉央
okamotoy@jaist.ac.jp

北陸先端科学技術大学院大学

2011年7月2日

”最終更新：2011/07/02 12:01”

① 乱択の重要性

② 確率の復習

③ 乱択による高速化：前進問題

④ 今日のまとめ

乱択アルゴリズムとは？

乱択アルゴリズム (randomized algorithm) とは？

乱数を使用するアルゴリズム

乱数を使用することのメリット

- ▶ アルゴリズム設計と実装が単純になる場合がある
- ▶ アルゴリズムの高速化が可能になる場合がある
- ▶ アルゴリズムの近似性能が上がる場合がある

乱数を使用することのデメリットとその解消

- ▶ アルゴリズムに望む性質が常に保証されるわけではなくなる
 - ▶ 例：停止しなくなる場合があるかもしれない
 - ▶ 解消法：「高い確率」で性質が保証されることを目指す
- ▶ アルゴリズムが乱数を作る必要が出てくる
 - ▶ 参考：擬似乱数生成 (pseudo-random number generation)
 - ▶ 解消法 (?)：その部分はブラックボックスとして扱う

① 乱択の重要性

② 確率の復習

③ 乱択による高速化：前進問題

④ 今日のまとめ

排反事象と独立事象

この講義で扱う確率はすべて離散確率 (確率空間が有限か可算無限)

- ▶ 2つの事象 A と B が**排反** (disjoint) であるとは

$$\Pr(A \text{ かつ } B) = 0$$

であること (注: 離散確率の場合の定義)

- ▶ 2つの事象 A と B が**独立** (independent) であるとは

$$\Pr(A \text{ かつ } B) = \Pr(A) \cdot \Pr(B)$$

であること

条件付き確率

事象 A のもとでの B の条件付き確率 (conditional probability) とは

$$\Pr(B|A) = \frac{\Pr(A \text{ かつ } B)}{\Pr(A)}$$

例: 公平なサイコロを1つ振る

偶数が出たという条件のもとで2が出る確率は

$$\frac{\Pr(\text{偶数が出て, かつ, 2が出る})}{\Pr(\text{偶数が出る})} = \frac{\Pr(2 \text{ が出る})}{\Pr(\text{偶数が出る})} = \frac{1/6}{1/2} = \frac{1}{3}$$

偶数が出たという条件のもとで3が出る確率は

$$\frac{\Pr(\text{偶数が出て, かつ, 3が出る})}{\Pr(\text{偶数が出る})} = \frac{\Pr(\emptyset)}{\Pr(\text{偶数が出る})} = 0$$

期待値と条件付き期待値

$\mathbb{N} = \{0, 1, 2, \dots\}$ 上の実数値確率変数 X の期待値 (expectation) とは

$$E[X] = \sum_{i=0}^{\infty} i \cdot \Pr(X = i)$$

事象 A のもとでの X の条件付き期待値 (conditional expectation) とは

$$E[X|A] = \sum_{i=0}^{\infty} i \cdot \Pr(X = i|A)$$

性質

全体事象が A と B に分割されるとき

$$E[X] = E[X|A] \Pr(A) + E[X|B] \Pr(B)$$

3つ以上の事象に分割される時も同様

期待値の線形性

性質 (期待値の線形性, linearity of expectation)

2つの実数値確率変数 X, Y に対して

$$E[X + Y] = E[X] + E[Y]$$

例: サイコロを2回振ったとき, 1回目の出目を X , 2回目の出目を Y とする

$$E[X + Y] = E[X] + E[Y] = \frac{7}{2} + \frac{7}{2} = 7$$

乱択アルゴリズムの解析でよく用いられる不等式

マルコフの不等式 (Markov's inequality)

非負値確率変数 $X \geq 0$ と正実数 $t > 0$ に対して

$$\Pr(X \geq t) \leq \frac{E[X]}{t}$$

証明：離散確率のときのみ証明 (連続確率の場合も同様)

$$\begin{aligned} E[X] &= \sum_{i=0}^{\infty} i \cdot \Pr(X = i) \\ &= \sum_{i=0}^{t-1} \underbrace{i}_{\geq 0} \cdot \Pr(X = i) + \sum_{i=t}^{\infty} \underbrace{i}_{\geq t} \cdot \Pr(X = i) \\ &\geq \sum_{i=t}^{\infty} t \cdot \Pr(X = i) = t \sum_{i=t}^{\infty} \Pr(X = i) = t \cdot \Pr(X \geq t) \end{aligned}$$

□

① 乱択の重要性

② 確率の復習

③ 乱択による高速化：前進問題

④ 今日のまとめ

前進問題

前進問題 (Going-forward problem)

設定

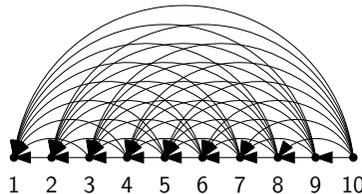
- ▶ 頂点集合を $\{1, 2, \dots, n\}$ とする有向グラフ
- ▶ 大きな数から小さな数へ向かう辺が必ず存在

行うこと

- ▶ 頂点 n から始めて、辺をたどることで頂点 1 に到達

問題

- ▶ 辺をいくつたどれば頂点 1 に到達できるか？



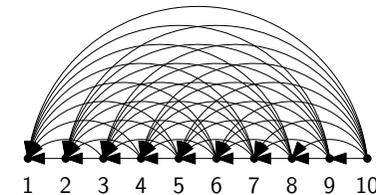
単純なアルゴリズム 1 (乱数を使わない)

- ① たどる辺を任意に選び、辺の先に移動する
- ② 移動先から出る辺がなければ終了、そうでなければ 1 に戻る

たどる辺の数

- ▶ 最悪の場合： $n - 1$ 個
- ▶ (最善の場合：1 個)

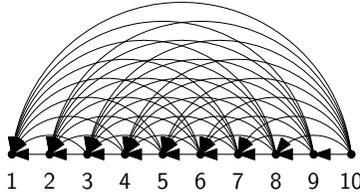
最悪計算量の意味では、よくないアルゴリズム



単純なアルゴリズム 2 (乱数を使う)

- ① たどる辺を一様分布に従って選び、辺の先に移動する
- ② 移動先から出る辺がなければ終了、そうでなければ 1 に戻る

一様分布に従って選ぶ \equiv 出る辺が k 個ある場合、それぞれを確率 $1/k$ で選ぶ



たどる辺の数の期待値

定理 1

単純なアルゴリズム 2 がたどる辺の数の期待値は

$$1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n-1}$$

注：次の H_n は n 次調和数 (n -th harmonic number) と呼ばれる

$$H_n = 1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n}$$

事実として、 $H_n = \ln n + O(1)$ が成り立つ

定理 1 の言い換え

単純なアルゴリズム 2 がたどる辺の数の期待値は $H_{n-1} = O(\log n)$

定理 1 の証明 (1)

- ▶ 任意の $k = 1, 2, \dots, n$ に対して

R_k = 頂点 k から始めて、頂点 1 への到達までにたどる辺数とする

- ▶ このとき、 $E[R_1] = 0$ で、 $k \geq 1$ のとき、

$$\begin{aligned} E[R_k] &= \sum_{i=1}^{k-1} E[R_k | \text{頂点 } k \text{ から頂点 } i \text{ に向かう辺を選ぶ}] \\ &\quad \cdot \Pr(\text{頂点 } k \text{ から頂点 } i \text{ に向かう辺を選ぶ}) \\ &= \sum_{i=1}^{k-1} (E[1 + R_i]) \cdot \frac{1}{k-1} = \sum_{i=1}^{k-1} (1 + E[R_i]) \cdot \frac{1}{k-1} \\ &= 1 + \frac{1}{k-1} \sum_{i=1}^{k-1} E[R_i] \end{aligned}$$

- ▶ この再帰式を解きたい

定理 1 の証明 (2)

- ▶ 両辺を $k-1$ 倍すると、 $k \geq 1$ のとき

$$(k-1)E[R_k] = k-1 + \sum_{i=1}^{k-1} E[R_i]$$

- ▶ よって、 $k \geq 2$ のとき

$$(k-2)E[R_{k-1}] = k-2 + \sum_{i=1}^{k-2} E[R_i]$$

- ▶ 上の式から下の式を引くと、 $k \geq 2$ のとき

$$(k-1)E[R_k] - (k-2)E[R_{k-1}] = 1 + E[R_{k-1}]$$

$$(k-1)E[R_k] - (k-1)E[R_{k-1}] = 1$$

$$E[R_k] = \frac{1}{k-1} + E[R_{k-1}]$$

定理 1 の証明 (3)

▶ したがって,

$$\begin{aligned}
 E[R_k] &= \frac{1}{k-1} + E[R_{k-1}] \\
 &= \frac{1}{k-1} + \frac{1}{k-2} + E[R_{k-2}] \\
 &= \frac{1}{k-1} + \frac{1}{k-2} + \frac{1}{k-3} + E[R_{k-3}] \\
 &= \dots \\
 &= \frac{1}{k-1} + \frac{1}{k-2} + \frac{1}{k-3} + \dots + \frac{1}{1} + \underbrace{E[R_1]}_{=0} \\
 &= H_{k-1}
 \end{aligned}$$

▶ 特に

$$E[R_n] = H_{n-1}$$

□

期待値が分かるとなぜよいか？

- ▶ たどる辺数の期待値が分かったからといって、アルゴリズムがそれだけの辺数しかたどらないとは限らない (乱数を使っているから)
- ▶ しかし,

$$\begin{aligned}
 \Pr(R_n \geq 2H_{n-1}) &\leq \frac{E[R_n]}{2H_{n-1}} \quad (\text{Markov の不等式}) \\
 &= \frac{H_{n-1}}{2H_{n-1}} = \frac{1}{2}
 \end{aligned}$$

- ▶ つまり, $\Pr(R_n < 2H_{n-1}) \geq 1 - \frac{1}{2} = \frac{1}{2}$
- ▶ $\frac{1}{2}$ 以上の確率で、たどる辺数は少ない ($2H_{n-1}$ 未満)

もう少し頑張ってみる：確率変数の指数関数

▶ R_n の代わりに 2^{R_n} を考えてみる

定理 2

任意の $k = 1, 2, \dots, n$ に対して

$$E[2^{R_k}] = k$$

▶ すなわち,

$$\begin{aligned}
 \Pr(R_n \geq 2 \log_2 n) &= \Pr(2^{R_n} \geq 2^{2 \log_2 n}) \\
 &= \Pr(2^{R_n} \geq n^2) \\
 &\leq \frac{E[2^{R_n}]}{n^2} = \frac{1}{n}
 \end{aligned}$$

▶ つまり, $1 - \frac{1}{n}$ 以上の確率でたどる辺数は少ない ($2 \log_2 n$ 未満)

定理 2 の証明 (1)

定理 1 と同様に進める

▶ $E[2^{R_k}] = 2^0 = 1$ で, $k \geq 1$ のとき,

$$\begin{aligned}
 E[2^{R_k}] &= \sum_{i=1}^{k-1} E[2^{R_k} | \text{頂点 } k \text{ から頂点 } i \text{ に向かう辺を選ぶ}] \\
 &\quad \cdot \Pr(\text{頂点 } k \text{ から頂点 } i \text{ に向かう辺を選ぶ}) \\
 &= \sum_{i=1}^{k-1} E[2^{1+R_i}] \cdot \frac{1}{k-1} = \sum_{i=1}^{k-1} E[2 \cdot 2^{R_i}] \cdot \frac{1}{k-1} \\
 &= \sum_{i=1}^{k-1} 2 E[2^{R_i}] \cdot \frac{1}{k-1} = \frac{2}{k-1} \sum_{i=1}^{k-1} E[2^{R_i}]
 \end{aligned}$$

▶ この再帰式を解きたい

定理 2 の証明 (2)

- ▶ 両辺を $k-1$ 倍すると, $k \geq 1$ のとき

$$(k-1)E[2^{R_k}] = 2 \sum_{i=1}^{k-1} E[2^{R_i}]$$

- ▶ よって, $k \geq 2$ のとき

$$(k-2)E[2^{R_{k-1}}] = 2 \sum_{i=1}^{k-2} E[2^{R_i}]$$

- ▶ 上の式から下の式を引くと, $k \geq 2$ のとき

$$\begin{aligned} (k-1)E[2^{R_k}] - (k-2)E[2^{R_{k-1}}] &= 2E[2^{R_{k-1}}] \\ E[2^{R_k}] &= \frac{k}{k-1} E[2^{R_{k-1}}] \end{aligned}$$

定理 2 の証明 (3)

- ▶ したがって,

$$\begin{aligned} E[2^{R_k}] &= \frac{k}{k-1} E[2^{R_{k-1}}] \\ &= \frac{k}{k-1} \cdot \frac{k-1}{k-2} E[2^{R_{k-2}}] \\ &= \dots \\ &= \frac{k}{k-1} \cdot \frac{k-1}{k-2} \cdot \dots \cdot \frac{2}{1} E[2^{R_1}] \\ &= k \end{aligned}$$

□

前進問題：まとめ

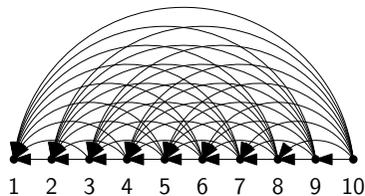
乱数を使わないアルゴリズム

- ▶ 最悪時：たどる辺数 = $n-1$

乱数を使うアルゴリズム

- ▶ 期待値：たどる辺数 = $H_{n-1} (= \ln n + O(1))$
- ▶ 高確率：たどる辺数 = $O(\log n)$

∴ 乱数を使うことで, 問題を高速に解けた



- ① 乱択の重要性
- ② 確率の復習
- ③ 乱択による高速化：前進問題
- ④ 今日のまとめ

今日のまとめ

乱択アルゴリズム

乱数を使用するアルゴリズム

- ▶ アルゴリズム設計と実装が単純になる場合がある
- ▶ アルゴリズムの高速化が可能になる場合がある

乱択による高速化

例：前進問題 ($O(n) \rightarrow O(\log n)$)

解析の方法論

- ▶ 確率変数の指数関数
- ▶ 期待値の計算
- ▶ 「Markov の不等式」の利用

参考文献

▶ 乱択アルゴリズムについて

- ▶ 玉木久夫．乱択アルゴリズム．共立出版，2008．
- ▶ 結城浩．数学ガール/乱択アルゴリズム．ソフトバンククリエイティブ，2011．
- ▶ R. Motwani, P. Raghavan. Randomized Algorithms. Cambridge University Press, 1998.
- ▶ M. Mitzenmacher, E. Upfal. Probability and Computing. Cambridge University Press, 2005. (日本語訳：小柴健史，河内亮周訳．確率と計算．共立出版，2009．)