

I482F 実践的アルゴリズム特論
9回目：確率解析

上原隆平
(uehara@jaist.ac.jp)

困難な問題に対する挑戦

- ▶ (計算機で解きたい) 困難な問題は
 - ▶ 解答の妥当性は、ある程度わかる
 - ▶ 可能な選択肢が爆発的に大きくて手に負えない
 - ▶ 理論的なモデルでは
 - ▶ 最悪の場合の計算量の下界を与える
 - ▶ 正確な解を求める
 - ▶ 入力サイズが大きくなったときの漸近的なふるまいを示すことが多い

もう少しなんとかしたい！！



困難な問題に対する挑戦

- ▶ (計算機で解きたい) 困難な問題は
 - ▶ 理論的なモデルでは
 - ▶ 最悪の場合の計算量の下界を与える
 - ▶ 正確な解を求める
 - ▶ 入力サイズが大きくなったときの漸近的なふるまいを示すことが多い

もう少しなんとかしたい！！

- ▶ 最近のトレンド:
 - ▶ ランダム性を考えた平均的なふるまいを考える ← 乱択アルゴリズム
 - ▶ 正解からの誤差を許した計算を考える ← 近似アルゴリズム
 - ▶ (指数時間かけてでも、ある程度の規模の問題を解く)



困難な問題に対する挑戦

- ▶ (計算機で解きたい) 困難な問題は
 - ▶ 理論的なモデルでは
 - ▶ 最近のトレンド:
 - ▶ ランダム性を考えた平均的なふるまいを考える: 乱択アルゴリズム
 - “ランダム性”を入力に仮定する
 - ...この場合はアルゴリズムは決定性でもよい
 - 一般に与えられる入力の分布はわからないことが多い
 - アルゴリズムで“乱数”を使う
 - “良い”疑似乱数を生成する方法はけっこう難しい
 - 効率(実行時間/使用メモリ)の期待値/最悪値などを解析する必要がある
 - ▶ 正解からの誤差を許した計算を考える: 近似アルゴリズム
 - 近似率を理論的に保証することができる場合がある



確率解析...の前に

▶ アルゴリズムで“乱数”を使う

▶ “良い”疑似乱数を生成する方法はけっこう難しい

- 良い疑似乱数を使用することで、モンテカルロ法が1000倍速くなり、それまで2年かかっていた計算が1日で終わるようになった事例がある。

- 疑似乱数の「良さ」とは？

コンピュータ・サイエンスのバイブル

“The Art of Computer Programming”, Vol. 2, by D.E. Knuth

でも詳しく議論されているテーマ

▶ メルセンヌ・ツイスター(Mersenne twister)

- 松本眞、西村拓士による疑似乱数生成器
- 公式サイトからダウンロードして使用可能
- <http://www.math.sci.hiroshima-u.ac.jp/~m-mat/MT/mt.html>
- 速い/コードが単純/性能が理論保証されている



確率解析の例: クーポンコレクター問題

▶ クーポンコレクター問題

n 種類のクーポンが無限にある。毎回それぞれのクーポンを $1/n$ の確率で引き当てる。すべてのクーポンが集まるまでクーポンを引き続けるとすると、クーポンを何枚引かなければならないだろうか。

▶ 定理

n 種類のクーポンが集まるまでにクーポンを引く枚数を $C(n)$ とおくと、 $C(n)$ の期待値 $E(C(n))$ は次の式で表現される:

$$E(C(n)) = O(n \log n)$$



確率解析の例: クーポンコレクター問題

▶ 定理

n 種類のクーポンが集まるまでにクーポンを引く枚数を $C(n)$ とおくと、 $C(n)$ の期待値 $E(C(n))$ は次の式で表現される:

$$E(C(n)) = O(n \log n)$$

▶ おまけ: 上記の式は $E(C(n)) = n \log n + o(1)$ と書ける。

▶ 例:

- ▶ クーポンが10種類なら23強
- ▶ クーポンが100種類なら460強

余談: マクドナルドの101匹わんちゃんのコンプリートセットは定価5万円でした。

▶ 余談: オークションで7000円でした。

確率解析の例: クーポンコレクター問題

- ▶ 定理: n 種類のクーポンが集まるまでにクーポンを引く枚数を $C(n)$ とおくと、 $C(n)$ の期待値 $E(C(n))$ は次の式で表現される: $E(C(n)) = O(n \log n)$
- ▶ 証明
 - ▶ 現在 i 種類のクーポンを持っている状態を 状態 i とする。
 - ▶ 時刻 t に t 枚目のクーポンを引くとする。
 - ▶ 時刻0は状態0で、時刻1は状態1である。
 - ▶ 状態 i でクーポンを1枚引くと、
 - 新しいクーポンを引いて状態 $i+1$ に状態遷移する確率: $(n-i)/n$
 - すでに持っているクーポンを引いて状態が変わらない確率: i/n
 - ▶ 新しい状態に遷移することを 成功 と考えると、これは 幾何分散 に関するよく知られた定理が使える。

確率解析の例: クーポンコレクター問題

- ▶ 幾何分散に関するよく知られた定理: 1回の試行で成功する確率が p である事象を繰り返し行くと、1回成功するまでの繰り返しの回数の期待値 $E(S(p))$ は次の式で表現される: $E(S(p)) = 1/p$
- ▶ 例:
 - ▶ 成功確率が $1/2$ なら繰り返しの回数の期待値は2回
 - ▶ 成功確率が $1/6$ なら繰り返しの回数の期待値は6回
 - ▶ 成功確率が $1/100$ なら繰り返しの回数の期待値は100回
- ▶ 証明
 - ▶ 定義通り、
$$1 \cdot p + 2 \cdot (1-p) \cdot p + 3 \cdot (1-p)^2 p + \dots + i \cdot (1-p)^{i-1} p + \dots$$
を計算すればよい。
 - ▶ 注意: あくまで期待値の話であって、博打には役立ちません

確率解析の例: クーポンコレクター問題

- ▶ 定理: n 種類のクーポンが集まるまでにクーポンを引く枚数を $C(n)$ とおくと、 $C(n)$ の期待値 $E(C(n))$ は次の式で表現される: $E(C(n)) = O(n \log n)$
- ▶ 証明
 - ▶ 状態 i でクーポンを1枚引くと、
 - 新しいクーポンを引いて状態 $i+1$ に状態遷移する確率: $(n-i)/n$
 - すでに持っているクーポンを引いて状態が変わらない確率: i/n
 - ▶ 新しい状態に遷移することを成功と考えると、これは幾何分散に関するよく知られた定理が使える。
 - ▶ よって状態 i から状態 $i+1$ に遷移するまでの時間の期待値 = 状態 i の区間の長さの期待値 = $n/(n-i)$
 - ▶ 期待値の線形性より、これらを $i=0 \sim n-1$ まで足せばよい
(意外と重要: 期待値には線形性があり、そのまま足すことができる。)

確率解析の例: クーポンコレクター問題

- ▶ 定理: n 種類のクーポンが集まるまでにクーポンを引く枚数を $C(n)$ とおくと、 $C(n)$ の期待値 $E(C(n))$ は次の式で表現される: $E(C(n)) = O(n \log n)$

- ▶ 証明

- ▶ したがって求める期待値は

$$\sum_{i=0}^{n-1} \frac{n}{n-i} = n \sum_{i=0}^{n-1} \frac{1}{n-i} = n \sum_{i=1}^n \frac{1}{i} = nH(n)$$

となる

- ▶ ここで $\sum_{i=1}^n \frac{1}{i} = H(n)$ は調和数 (Harmonic number) と呼ばれる数列で

$$H(n) = \log n + \gamma + \frac{1}{n} - \frac{1}{12n^2} + \frac{1}{120n^4} - \varepsilon, \quad 0 < \varepsilon < \frac{1}{256n^6}, \quad \gamma = 0.5772156649\dots$$

が知られている。よって定理を得る。



確率解析の例: クーポンコレクター問題

- ▶ 調和数についての参考文献
 - ▶ 簡単に読めて楽しい(?)本
 - 『数学ガール』結城浩、ソフトバンククリエイティブ、2007.
ハーモニックナンバーに関する章だけなら以下のWeb版でも読めます:
 - テトラちゃん和ハーモニック・ナンバー
<http://www.hyuki.com/girl/harmonic.pdf>
結城浩、2006年4月
 - 『世界でもっとも奇妙な数学パズル』ジュリアン・ハヴィル著、松浦俊輔訳、青土社、2009.
第11章に載ってます。いろいろな「不思議」が載っている良書。

