数理解析研究所講究録 1351 巻 2004 年 134-142

計数問題における複合モンテカルロ法について

中村 将人 (金沢大学大学院 自然科学研究科)
小川 重義 (立命館大学理工学部 数理科学科)

1 はじめに

これから述べる複合モンテカルロ法は、バックトラック法とモンテカルロ法を組み合わせたアルゴリズムで、計数問題の近似解を得るために [1] により考案された。その手法について、[1],[2] において、精度と計算時間の両面において有効な探索が行えることが実験的に確かめられている。しかし、なぜそのような探索が可能になるかについての理論的考察は行われていない。

そこで、本稿ではバックトラック法による全探索法、モンテカルロ法、そして複合モンテカルロ法の計算量を比較することで、より定量的な考察を行うことを目的とする。

キーワード : 計数問題、バックトラック法、モンテカルロ法

2 計数問題

ある配置の集合の中から、解となる要素をすべて数え上げる問題である。

計数問題の定式化

$$
\begin{aligned}
\Omega &= \{0,1\}^N \quad \text{配置の集合} \\
S &\subset \Omega \quad \text{解となる配置の集合} \\
x_i &\in \{0,1\} \quad i=1,2,\cdots,N \\
\boldsymbol{x} &= (x_1,x_2,\cdots,x_N) \quad (\in \Omega) \\
\mathbf{1}_S(\boldsymbol{x}) &= \begin{cases} 0 & , \ \boldsymbol{x} \notin S \\ 1 & , \ \boldsymbol{x} \in S \end{cases}
\end{aligned}
$$

このとき

$$
|S| = \sum_{\boldsymbol{x}\in\Omega} \mathbf{1}_S(\boldsymbol{x})
$$

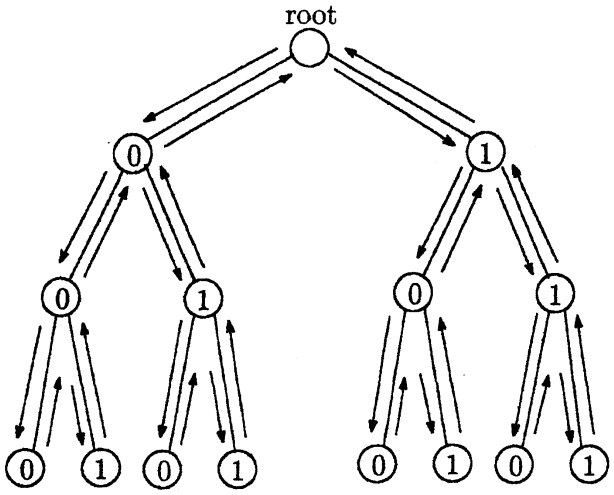
を推定したい。

したがって、正確な解の総数を求めるには、深さNの2分木の全探索が必要となる。このような木を探索するアルゴリズムに**バックトラック法**がある。

(注) ここで、配置の集合ΩとしてN個の0と1からなる集合としているのは、あらゆる計数問題に対して適用できる手法を考えたいがためである。

3 バックトラック法

バックトラック法を用いると、効率よく配置を作ることができる。
$N=3$の場合におけるバックトラック法で配置の作り方を図示したものを以下に示す。



今、どの枝も移動するのに要する時間を一定とすると、バックトラック法の計算量は木の枝の本数の2倍であるといえる。

$$2\cdot\sum_{i=1}^{N}2^i=4(2^N-1) \tag{1}$$

この式が示す通り、Nが大きくなることで計算量が増大し、実行不可能となる。そこで、ある程度の誤差を許容し、解の密度を推定することを考える。

4 近似値の精度について

単純に誤差$|\hat{p}-p|$をとったのでは、解の密度が小さい場合、ほぼ0になってしまい都合が悪い。そこで、相対誤差がϵ以下になる確率が$1-\alpha$以上となる解の密度の推定値を得ることを目標と定める。

$$Prob\{\frac{|\hat{p}-p|}{p}<\epsilon\}>1-\alpha \tag{2}$$

$$p=\frac{|S|}{|\Omega|}\ \text{解の密度},\quad \hat{p}\ :\ \text{解の密度の推定量}$$

こうすることで、配置の総数$|\Omega|(=m$とする)よりも小さいサンプル数nで、計算が実行できることが期待できる。

(※) 中心極限定理から次のことがわかっている。

$$Prob\{\frac{|\hat{p}-p|}{\sqrt{Var(\hat{p})}}<z_{\frac{\alpha}{2}}\}\ \overset{n\to\infty}{\approx}\ 1-\alpha \tag{3}$$

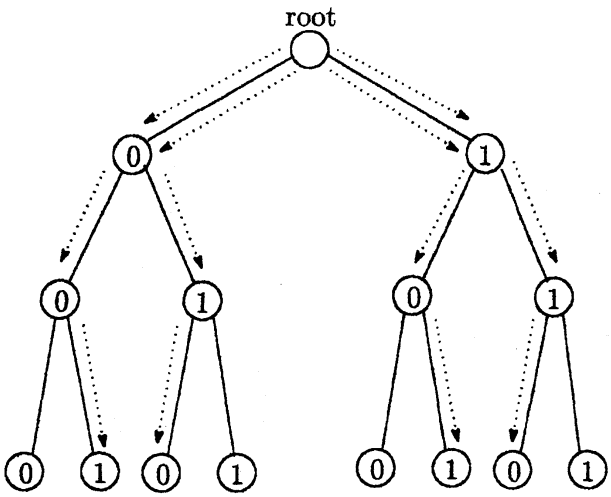
$$\text{ここで、}z_{\frac{\alpha}{2}}\text{は}N(0,1)\text{の}\frac{\alpha}{2}\text{点}$$

したがって、式(2)を満たすには密度の推定量の分散に課す条件は次式で表せる。

$$z_{\frac{\alpha}{2}}\sqrt{Var(\hat{p})}<\epsilon p \tag{4}$$

5　モンテカルロ法

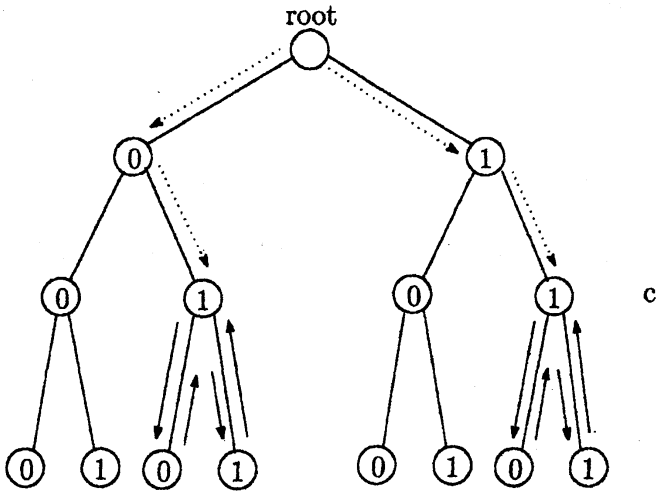
具体的な推定量としてすぐに思いつくのがモンテカルロ法である。ここでも、$N = 3$ の場合の配置の作り方を以下に示す。



どの葉を調べる確率も等しくなるように探索を行い、見つかった解の頻度から推定する。しかしこの方法は、常に根からスタートするため、探索の効率においてバックトラック法に劣るという欠点がある。
そこで、バックトラック法とモンテカルロ法の長所を組み合わせたアルゴリズムである**複合モンテカルロ法**が考案された。

6　複合モンテカルロ法

$N = 3$ の場合、複合モンテカルロ法による配置の作り方を図示すると以下のようになる。



上図において、モンテカルロ法で降下する枝は**点線の矢印**、バックトラック法で移動する枝は**実線の矢印**で表現されている。こうして、各グループごとの解の頻度を調べ、それらの平均値を推定量とする。

6.1 カットオフレベルの調節

複合モンテカルロ法において最も重要なパラメータであるカットオフレベルcについて説明する。カットオフレベルとはモンテカルロ法で根からランダムに降下する枝の本数のことである。上の図で言えばカットオフレベルは2である。なぜ重要かといえば、カットオフレベルにより推定量の分散が変化するからである。例えば、$c = N$とすればモンテカルロ法に完全に一致する。
また、カットオフレベルを決めることは集合Ωを交わりの無い部分集合(グループ)に分けることに相当する。

6.2 複合モンテカルロ法による推定量

これまでのことを定式化すると以下のようになる。

複合モンテカルロ法の定式化

$$
\begin{aligned}
a &:= |S| \quad \text{解の総数} \\
s &: \quad m \text{の約数} (= 2^{N-c}) \quad \text{グループ内の要素数} \\
b &:= \frac{m}{s} \quad \text{グループの数} \\
A(j) &: \quad j \text{番目のグループの解の個数}, \ 1 \le j \le b \\
X_k &= A(r_k), \quad (\text{ただし、} r_k \sim 1 \text{から} b \text{までの離散一様分布 } i.i.d. \ ) \qquad (5) \\
\hat{p} &= \frac{1}{s}\frac{1}{n'}\sum_{k=1}^{n'} X_k, \quad (\text{ただし、} sn' = n) \qquad (6) \\
E(\hat{p}) &= p, \quad Var(\hat{p}) = \frac{\sum_{j=1}^{b}\{A(j)\}^2 - \frac{a^2}{b}}{nm} \qquad (7)
\end{aligned}
$$

式(7)と、相対誤差に関する条件式(4)から必要なサンプル数が求まる。

$$
n > \frac{z_{\frac{\alpha}{2}}^2 (\sum_{j=1}^{b}\{A(j)\}^2 - \frac{a^2}{b})}{m\epsilon^2 p^2} \qquad (8)
$$

6.3 複合モンテカルロ法の計算量

複合モンテカルロ法の計算量も、これまで同様移動する木の枝の本数で見積もることにする。したがって、まずc本の枝をランダムに降下し、その後の枝をバックトラック法で移動する、という試行をn'回繰り返すので以下の式のようになる。

$$
\begin{aligned}
& (c + 2 \cdot \sum_{i=1}^{N-c} 2^i) n' \\
= \ & (c + 4(2^{N-c} - 1)) \frac{n}{2^{N-c}} \\
= \ & (c2^{c-N} + 4(1 - 2^{c-N})) \frac{z_{\frac{\alpha}{2}}^2 (\sum_{j=1}^{b}\{A(j)\}^2 - \frac{a^2}{b})}{m\epsilon^2 p^2} \\
= \ & \left((c-4)2^c + 2^{N+2}\right) \frac{z_{\frac{\alpha}{2}}^2 (\sum_{j=1}^{b}\{A(j)\}^2 - \frac{a^2}{b})}{\epsilon^2 a^2} \qquad (9)
\end{aligned}
$$

6.4 計算量比

ここで、全探索を行う場合と比較がしやすいように計算量比を定義しておく。

$$\text{計算量比}\,(c) \;:=\; \frac{\text{あるカットオフレベルにおける計算量 (式 (9))}}{\text{バックトラック法における計算量 (式 (1))}}$$

$$= \frac{\left((c-4)2^c + 2^{N+2}\right)\dfrac{z_{\frac{\alpha}{2}}^2\left(\sum_{j=1}^{b}\{A(j)\}^2 - \frac{a^2}{b}\right)}{\epsilon^2 a^2}}{4(2^N-1)}$$

こうすると、各カットオフレベルの計算量がバックトラック法による全探索を行う場合の計算量の何倍になるかが分かる。この式を最小にするようなcは、精度と計算量の両面で**最適なカットオフレベル**ということができる。
しかし、解の分布を仮定しない限り未知のパラメータがあるため、この計算量比を具体的に求めることができない。
そこで、次節からの例題で解の分布を仮定してみる。なお、近似値の精度に関するパラメータは、$\epsilon = 0.1, z_{\frac{\alpha}{2}} = 1.96$ としている。

7 例題

7.1 <例題1> 解が一様に散らばっている場合(バランスしたツリー構造)

$|\Omega| = 2^{100}$ 配置の総数　$|S| = 724$ 解の総数

● 解となる配置　　○ 解ではない配置

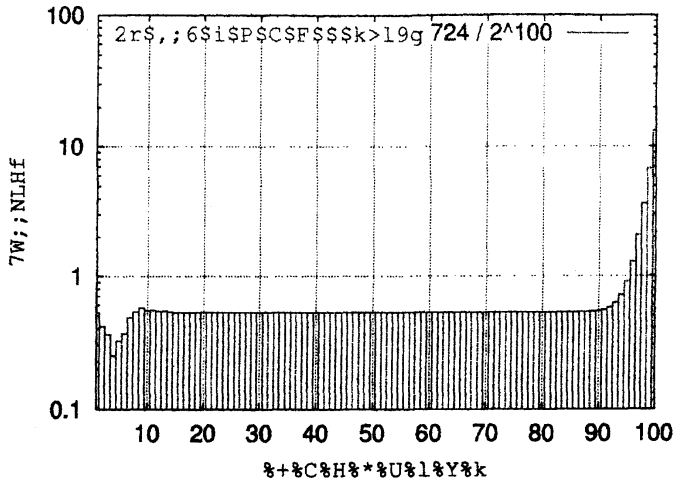
解のばら撒き方

1. Y_1 を Ω の要素から等しい確率で選び、その配置を解とする。
2. Y_2 を $\Omega - \{Y_1\}$ の要素から等しい確率で選び、その配置を解とする。
3. 同様に解となる配置を決めていく。
4. Y_a を $\Omega - \{Y_1, Y_2, \cdots, Y_{a-1}\}$ の要素から等しい確率で選び、その配置を解とする。

このように解をばら撒くことでどのような解の配置も等しい確率で起こる。
略証 a 個の1と $m-a$ 個の0からなる任意の配置が起こる確率は、

$$a! \cdot \frac{1}{m} \cdot \frac{1}{m-1} \cdots \frac{1}{m-a+1} = \frac{1}{\binom{m}{a}}$$

である。



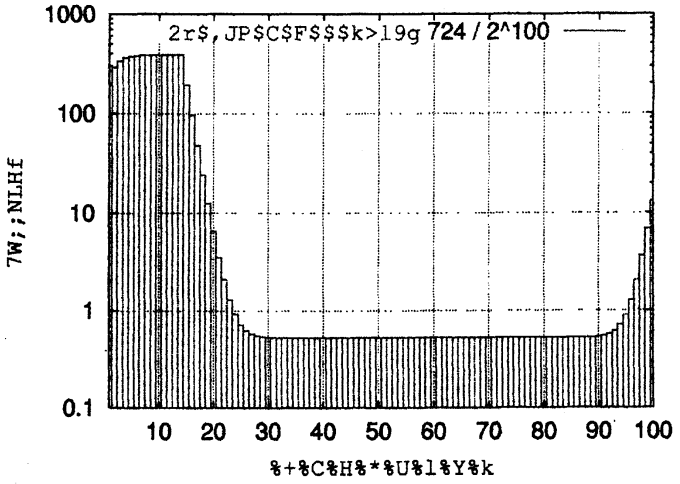
この場合、カットオフレベルを深くしすぎなければ、どのカットオフレベルでも1以下となっているので全探索をする場合に比べ、有効な推定が行える。
しかし、実際に解く問題ではこのように都合よく解が散らばっていてくれるとは限らない。

7.2 ＜例題2＞ 解が偏っている場合 (バランスしていないツリー構造)

$|\Omega| = 2^{100}$ 配置の総数 $|S| = 724$ 解の総数

●●○●○○●○●●○●○○●●○○○○○○○○○○○○

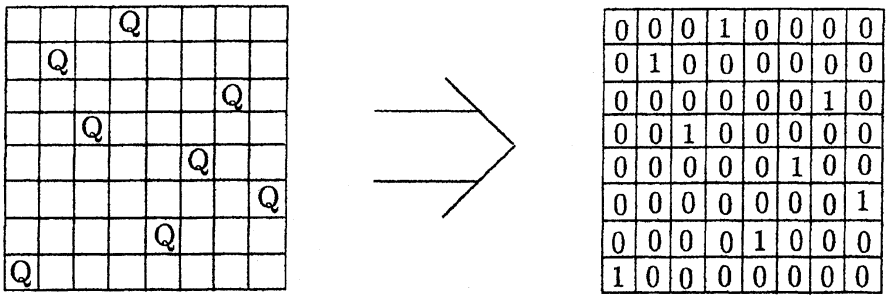
● 解となる配置　　○ 解ではない配置



このグラフから、解の分布がある程度偏っている場合においても、良いカットオフレベルを選べば、複合モンテカルロ法が有効に機能することがわかる。では、実際の計数問題に適用してみる。

7.3 ＜例題３＞ M-Queens Problem

Ｍ×Ｍのチェス盤上にQueenをＭ個置き、それらが互いに取り合うことの無い配置の総数を求める問題。



0	0	0	1	0	0	0	0
0	1	0	0	0	0	0	0
0	0	0	0	0	0	1	0
0	0	1	0	0	0	0	0
0	0	0	0	0	1	0	0
0	0	0	0	0	0	0	1
0	0	0	0	1	0	0	0
1	0	0	0	0	0	0	0

図 1: 8-Queens Problem の解の配置の例

このように配置を表現すると、配置の総数は

$$N = M \times M = 8 \times 8$$
$$|\Omega| = 2^N = 2^{64}$$

となる。

(注) M-Queens Problemの場合、どの行と列にも１個しかQueenを置けないことから、更に枝刈りが可能だがどのような計数問題に対しても適用できるようにしたいので、枝刈りなどの各問題固有の性質は考えないことにする。

10-Queens Problem

Queenの数が１０のときの計算量比は以下のようになった。

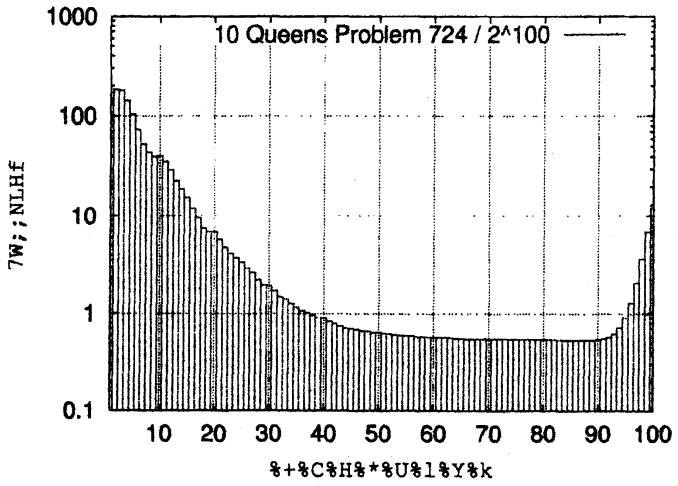


図 2: 10-Queens Problem $p = \frac{724}{2^{100}}$

以上の例題から、解の配置が偏っていると複合モンテカルロ法が有効に機能しないことが見てとれる。そこで、解の配置の偏り度を定義し、その偏り度が大きい値をとるような解の配置は非常に少ないことを次節で示す。

8 解の配置が偏る確率は少ないことの証明

まず、解の配置 Φ を以下のように表現する。

$$\begin{aligned}\Phi &= \{\boldsymbol{y}=(y_1,y_2,\cdots,y_m)\in\{0,1\}^m|\sum_{i=1}^{m}y_i=a\}\\ \forall\boldsymbol{z} &= (z_1,z_2,\ \cdots\ ,z_m)\ \in\ \Phi\\ z_i &= \begin{cases}0\ ,\ i\text{番目の配置が解でない}\\ 1\ ,\quad i\text{番目の配置が解}\end{cases}\end{aligned}$$

ただし、i 番目の配置とは $i-1$ の２進展開の配置に対応することにする。
ここで、計算の便宜上 $\frac{a}{m}:=p<\frac{1}{2}$ を仮定しておく。もし、$p>\frac{1}{2}$ であるような解の配置に対してこれから定義する偏り度を適用したいのであれば、0 と 1 を反転すればよい。

8.1 局所密度 $\delta_k(\boldsymbol{z})$ の定義

次式で**局所密度**を定義する。

$$\begin{aligned}\delta_k(\boldsymbol{z}) &:= \frac{1}{a}\sum_{i=k}^{k+a-1}z_i \qquad (10)\\ \text{ただし、}\quad & 1\le k\le m\ ,\ z_{m+i}=z_i\end{aligned}$$

(※) ここで、$z_{m+i}=z_i$ というようにループ状にしたのは、

$$\sum_{k=1}^{m}\delta_k(\boldsymbol{z}) = a \qquad (11)$$

としたいためである。

8.2 平均局所密度 $\bar{\delta}(\boldsymbol{z})$

この局所密度の平均を計算する。

$$\begin{aligned}\bar{\delta}(\boldsymbol{z}) &:= \frac{1}{m}\sum_{k=1}^{m}\delta_k(\boldsymbol{z})\\ &= p\ ,\quad (\because (11))\end{aligned}$$

8.3 $\bar{\delta}(\boldsymbol{z})$ まわりの２次モーメント $M_2(\boldsymbol{z})$

$$\begin{aligned}M_2(\boldsymbol{z}) &:= \frac{1}{m}\sum_{k=1}^{m}\left(\delta_k(\boldsymbol{z})-\bar{\delta}(\boldsymbol{z})\right)^2\\ &= \frac{1}{m}\sum_{k=1}^{m}(\delta_k(\boldsymbol{z}))^2-p^2 \qquad (12)\end{aligned}$$

この式 (12) を散らばり具合を表す指標としたのでは、0 と 1 の間の値にならない。そこで、(12) の最大値を求め、その値で割って規格化することにする。

8.4 $M_2(z)$ の最大値 M_2^{max}

$M_2(z)$ を最大にするような解の配置は $11\cdots100\cdots0$ となる場合で、容易に計算される。ただし、一番最初に仮定したとおり、1 の数のほうが 0 の数よりも少ないとする。

$$\begin{aligned} M_2^{max} &= \frac{1}{a^2m}(a^2+2\sum_{k=1}^{a-1}k^2)-p^2 \\ &= \frac{1}{a^2m}(a^2+2\frac{(a-1)a(2a-1)}{6})-\frac{a^2}{m^2} \\ &= \frac{m(2a^2+1)-3a^3}{3am^2} \end{aligned} \tag{13}$$

8.5 偏り度 $D(z)$ の定義

これらを用いて解の配置の偏り度を次式で定義する。

$$D(z) := \frac{M_2(z)}{M_2^{max}} \tag{14}$$

どのような解の配置も等確率と仮定したとき、ここで定義した偏り度の期待値は、

$$E(D(z)) = \frac{3am(1-p)(m-a)}{(m-1)(m(2a^2+1)-3a^3)} \tag{15}$$

と求まる。

8.6 マルコフの不等式

$\forall\alpha > 0$ に対して、

$$Prob\{D(z) \geq \alpha\} \leq \frac{E(D(z))}{\alpha} \tag{16}$$

がいえる。

(確認) もしレアイベント $(p \ll 1)$ ならば、

$$E(D(z)) \approx \frac{3am^2}{2a^2m^2} = \frac{3}{2a}$$

となり、解の数 a が大きければ確かに 0 に近い値となり、解が偏った配置となる確率は小さいといえる。したがって、複合モンテカルロ法が有効に機能しないような解の配置は非常に少ないことが言えた。 □

参考文献

[1] M.Terada. Estimating Number of Solutions with Hybrid Monte Carlo Method. (1997)

[2] 紺谷真紀子 計数問題における複合モンテカルロ法とその有効性 (2003)