モンテカルロ法，乱数，および疑似乱数

杉　田　　洋
sugita@math.sci.osaka-u.ac.jp

大阪大学大学院理学研究科

例題

- 硬貨を $2^7 = 128$ 回投げるとき，表が続けて7回以上出る確率 p を求めよ．

モンテカルロ法による解法

- 試行「硬貨投げを128回行う」を $2^{19} = 524{,}288$ 回繰り返す．
- そのうち事象「表が7回以上続けて出る」の起こる回数を S とする．
- 比 $\dfrac{S}{524288}$ を p の推定値とする．

根拠：大数の法則

$$\text{Probablity}\left(\left|\frac{S}{524288} - p\right| > \frac{1}{256}\right) < \frac{1}{32}.$$

実際の手続き

- 硬貨を投げる回数は

$$2^7 \times 2^{19} = 128 \times 524,288 = 67,108,864$$

- 本物の硬貨を投げるのは手間が掛かり過ぎて不可能．
- そこで「乱数」を用いてコンピュータ上でシミュレートしたい．
- しかしコンピュータは乱数を生成することができない．
- 乱数の代わりに，コンピュータで生成する「疑似乱数」を用いて計算する．

デモンストレーションをご覧下さい．

問題

- 乱数とは何か？
- なぜ乱数はコンピュータで生成できないのか？
- なぜモンテカルロ法には乱数が必要なのか？
- 疑似乱数による計算が正当化できるのか？

解答

- 乱数の定義：Kolmogorov，Chaitin，Martin-Löf (1960's)
- 疑似乱数の定義：Blum-Blum-Shub，Yao (1980's)

$\Longrightarrow$ これらの定義を基にモンテカルロ法を定式化する．

$\Longrightarrow$ **疑似乱数による計算が正当化できる場合がある！**

例題のための定式化(1)

- 硬貨の表を1，裏を0，で表す．
- $\{0,1\}^m$： 0と1からなる長さmの列の全体．
- 関数 $X : \{0,1\}^{128} \to \mathbf{R}$，
  各 $(\xi_1, \ldots, \xi_{128}) \in \{0,1\}^{128}$ に対して

$$X(\xi_1, \ldots, \xi_{128}) := \begin{cases} 1 & (\xi_1, \ldots, \xi_{128} \text{ の中で1が7個以上続くとき}) \\ 0 & (\text{そうでないとき}) \end{cases}$$

- 確率空間： $(\{0,1\}^{67108864}, P)$，$P :=$ 一様分布．

  すなわち，$\{0,1\}^{67108864}$ のすべての元に等確率 $2^{-67108864}$ を与える．

例題のための定式化(2)

- X の独立なコピーの列： $X_1, \ldots, X_{524288} : \{0,1\}^{67108864} \to \mathbf{R}$,
  i.e., 各 $\omega = (\omega_1, \ldots, \omega_{67108864}) \in \{0,1\}^{67108864}$ に対して

$$
\begin{aligned}
X_1(\omega) &:= X(\omega_1, \ldots, \omega_{128}) \\
X_2(\omega) &:= X(\omega_{129}, \ldots, \omega_{256}) \\
&\cdots \\
X_{524288}(\omega) &:= X(\omega_{67108737}, \omega_{6708738}, \ldots, \omega_{67108864})
\end{aligned}
$$

- 目的の確率変数： $S : \{0,1\}^{67108864} \to \mathbf{R}$,

$$
S(\omega) := \sum_{k=1}^{524288} X_k(\omega) = X_1(\omega) + X_2(\omega) + \cdots + X_{524288}(\omega)
$$

例題のための定式化(3)

- S の平均 $E[S(\omega)]$ と分散 $V[S(\omega)]$

$$E[S(\omega)] = 524288\, E[X_1(\omega)] = 524288\, p$$
$$V[S(\omega)] = 524288\, V[X_1(\omega)] = 524288\, p(1-p)$$
$$\leq 524288 \cdot \frac{1}{4}$$

- 大数の法則：チェビシェフの不等式

集合 $A \subset \{0,1\}^{67108864}$ を

$$A := \left\{ \omega \in \{0,1\}^{67108864} ;\ \left| \frac{S(\omega)}{524288} - p \right| > \frac{1}{256} \right\}$$

とするとき，その確率は

$$P(\omega \in A) < \frac{V[S(\omega)] \cdot 256^2}{524288^2} \leq \frac{1}{32}$$

「賭け」としてのモンテカルロ法

$$A = \left\{ \omega \in \{0,1\}^{67108864} ; \left| \frac{S(\omega)}{524288} - p \right| > \frac{1}{256} \right\}$$

$$P(\omega \in A) < \frac{1}{32}$$

- プレーヤー，アリスが $\omega \in \{0,1\}^{67108864}$ を自分の意思で選ぶ．
- もし $\omega \notin A$ ならば勝ち，$\omega \in A$ ならば負け，とする．
- p が未知なので，勝ち負けの結果は賭けの実行後も不明．
- 負ける確率は $1/32$ より小さい．

じつはアリスが勝つことは容易ではない．なぜか？

乱数の問題(1)

- アリスが**自分の意志**で選べる $\omega \in \{0,1\}^{67108864}$ は非常に少数．
  その理由は…
  - 各 $\omega \in \{0,1\}^{67108864}$ は8MBの長大なデータなのでキーボードからそのまま入力することは事実上不可能．
  - アリスが1,000ビットまでキーボードから入力できるとすると，選べる $\omega \in \{0,1\}^{67108864}$ は $2^{1001} - 2$ 個．

> k ビットの入力で選べる ω は 2^k 個．だから1,000ビット以下の入力で選べる ω の個数は
>
> $$2 + 2^2 + \cdots + 2^{1000} = 2^{1001} - 2.$$

  - $\{0,1\}^{67108864}$ の元の総数は $2^{67108864}$ 個，アリスの選べる元の個数は $2^{1001} - 2$ 個．

乱数の問題 (2)

- 乱数とは...
  - たとえアリスが 67,108,864−10 ビットまで入力できたとしても，アリスが選べる ω の個数は

$$2 + 2^2 + \cdots + 2^{67108864-10} = 2^{67108864-9} - 2$$

    これは全体 $2^{67108864}$ 個の 2^{-9} 倍，すなわち 1/512 に過ぎない．
  - $\{0,1\}^{67108864}$ の元のうち 511/512 は入力が 67,108,864−9 ビット以上必要である．

それ自身とほぼ同じ長さの入力が必要な ω を**乱数**という．

乱数の問題(3)

- アリスの負ける確率は一様分布 P に基づいている．それが意味を持つためには ω は圧倒的多数を占める乱数から選ばれるべきである．このことが「モンテカルロ法には乱数が必要である」と言われる理由である．
- 乱数はアリスの(人間の)意志では選べない．

(問題)

- モンテカルロ法には乱数が絶対に必要か？

否，疑似乱数で十分な場合もある！

疑似乱数生成器(1)

もう一度，デモンストレーションをご覧下さい．

- 疑似乱数生成器 $g : \{0,1\}^{292} \rightarrow \{0,1\}^{67108864}$
  - アリスが種 $\omega' \in \{0,1\}^{292}$ を選ぶ．
  - g が ω' を疑似乱数 $g(\omega') \in \{0,1\}^{67108864}$ に引き伸ばす．
  - $S(g(\omega'))$ を計算して p を推定する．
- 新しい「賭け」
  - $g(\omega') \notin A$ ならアリスの勝ち，$g(\omega') \in A$ ならアリスの負け．

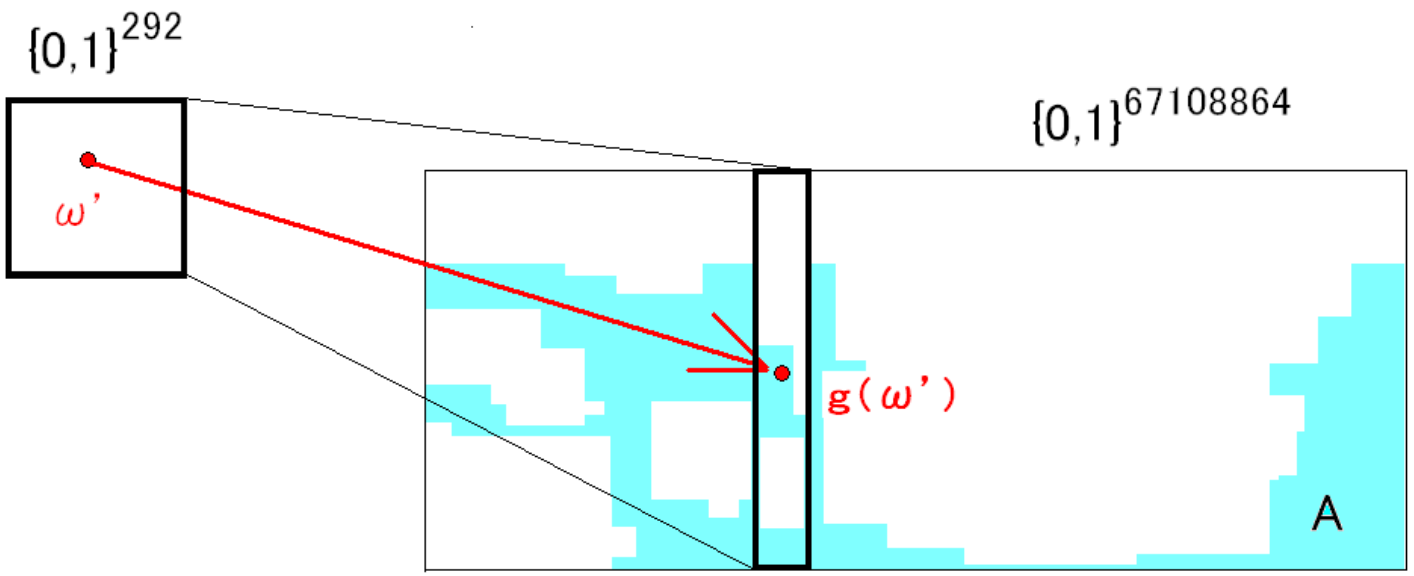
疑似乱数生成器(2)

(問題)

- 新しい「賭け」におけるアリスの負ける確率はいくらか？
  すなわち，P'：$\{0,1\}^{292}$ 上の一様分布，のとき

$$P'\left(\left\{\omega' \in \{0,1\}^{292};\ g(\omega') \in A\right\}\right)$$

  を評価せよ．



疑似乱数生成器(3)

- デモに用いた疑似乱数生成器 g は次の不等式を満たす.

$$P'\left(\left\{\omega' \in \{0,1\}^{292};\ g(\omega') \in A\right\}\right) < \frac{1}{32}$$

- g はアリスの**負ける確率**を大きくしないので「A に対して**安全**な疑似乱数生成器」と呼ばれる.
- このような g が存在するので例題を解くのに**乱数は必要でない**.

デモは疑似乱数生成器を用いているが数学的に正当なもの

疑似乱数生成器 $g : \{0,1\}^{292} \to \{0,1\}^{67108864}$ の定義

- 種 $\{0,1\}^{292} \ni \omega' = (\omega'_1, \ldots, \omega'_{146}, \omega'_{147}, \ldots, \omega'_{292}) =$

  1110110101 1011101101 0100000011 0110101001 0101000100 0101111101
  1010000000 1010100011 0100011001 1101111101 1101010011 1111001000
  1010010000 1101011101 0011001100 0001110111 0001000001 1010111001
  0000010010 0010001010 1011011110 1011100010 0100111000 0000110101
  0001100101 1100100101 1111100110 1100000101 1000011011 10

- $\tilde{x} := \sum_{i=1}^{146} 2^{-i} \omega'_i \qquad \tilde{\alpha} := \sum_{i=1}^{146} 2^{-i} \omega'_{146+i}$

  $\tilde{x} = 0.$ 1110110101 1011101101 0100000011 0110101001 0101000100
  0101111101 1010000000 1010100011 0100011001 1101111101
  1101010011 1111001000 1010010000 1101011101 001100

  $\tilde{\alpha} = 0.$ 1100000111 0111000100 0001101011 1001000001 0010001000
  1010101101 1110101110 0010010011 1000000011 0101000110
  0101110010 0101111110 0110110000 0100100001 101110

- $\tilde{z}_k := \tilde{x} + k\tilde{\alpha} \pmod 1, \quad k = 1, 2, 3, \ldots, 2^{19} = 524288$

$\tilde{z}_1 = 0.$ 1010111101 0010110001 0101101110 1111101010 0111001101
0000101011 1000101110 1100110110 1100011101 0011000100
0011000110 0101000111 0001000001 0010111110 111010

$\tilde{z}_2 = 0.$ 0111000100 1001110101 0111011010 1000101011 1001010101
1011011001 0111011100 1111001010 0100100000 1000001010
1000111000 1011000101 0111110001 1000100000 101000

- $\{0,1\}^{128} \ni g_1(\omega'), g_2(\omega'), \ldots$

$g_1(\omega') =$ 1010111101 0010110001 0101101110 1111101010 0111001101
0000101011 1000101110 1100110110 1100011101 0011000100
0011000110 0101000111 00010000

$g_2(\omega') =$ 0111000100 1001110101 0111011010 1000101011 1001010101
1011011001 0111011100 1111001010 0100100000 1000001010
1000111000 1011000101 01111100

- $g(\omega') := (g_1(\omega'), \ldots, g_{524288}(\omega')) \in \{0,1\}^{67108864}, \quad \omega' \in \{0,1\}^{292}.$

定理 1

E'，V' を $\{0,1\}^{292}$ 上の一様分布 P' に関する平均，分散とするとき

$$E'[S(g(\omega'))] = E[S(\omega)] = 524288\,p$$

$$V'[S(g(\omega'))] = V[S(\omega)] = 524288\,p(1-p)$$

が成り立つ．

定理 1 よりチェビシェフの不等式が $S(\omega)$ の場合と同じ形で成り立つ．

$$P'(g(\omega') \in A) < \frac{V'[S(g(\omega'))] \cdot 256^2}{524288^2} \leq \frac{1}{32}.$$

すなわち，g は A に対して安全な疑似乱数生成器である．

定理 1 は次の定理 2 より従う．

定理 2

$\{0,1\}^{292}$ 上の一様分布 P' の下で，$\{g_k(\omega')\}_{k=1}^{524288}$ について

(i) 各 $g_k(\omega')$ は $\{0,1\}^{128}$ 上一様分布．

(ii) $k \neq l$ ならば $g_k(\omega')$ と $g_l(\omega')$ は独立(ペアごとに独立)．

(注意)

- (i)+(ii) は次の (iii) と同値．
  (iii) $k \neq l$ ならば，各 $a, b \in \{i/2^m; i = 0, 1, \ldots, 2^m - 1\}$ に対して

$$P'\left(g_k(\omega') = a,\ g_l(\omega') = b\right) = 2^{-m} \times 2^{-m}.$$

- 3 項 $g_k(\omega'), g_l(\omega'), g_m(\omega')$ は独立でない．

定理 2の証明

- 一般に周期1の有界関数 F, G に対して

$$\int_0^1 \int_0^1 F(x + k\alpha)G(x + l\alpha)dxd\alpha = \int_0^1 F(x)dx \int_0^1 G(\alpha)d\alpha.$$

- $a, b \in \{i/2^m; i = 0, 1, \ldots, 2^m - 1\}$, $C = [a, a + 2^{-m})$, $D = [b, b + 2^{-m})$ に対して

$$\begin{aligned} P'\left(g_k(\omega') = a,\ g_l(\omega') = b\right) &= E'\left[1_C(\tilde{z}_k)1_D(\tilde{z}_l)\right] \\ &= \int_0^1 \int_0^1 1_C(x + k\alpha)1_D(x + l\alpha)dxd\alpha \\ &= \int_0^1 1_C(x)dx \int_0^1 1_D(\alpha)d\alpha \\ &= 2^{-m} \times 2^{-m}. \end{aligned}$$

(q.e.d.)

“定理 2 $\Longrightarrow$ 定理 1” の証明

$X'_k := X(g_k(\omega'))$ とおく.

(平均の計算) 各 $g_k(\omega')$ は $\{0,1\}^m$ 上一様分布するから

$$E'\left[X'_k\right] = E\left[X_k\right] = p$$

従って

$$E'[S(g(\omega'))] = E'\left[\sum_{k=1}^{524288} X'_k\right] = \sum_{k=1}^{524288} E'\left[X'_k\right] = 524288p$$

よって

$$E'[S(g(\omega'))] = E[S(\omega)].$$

（分散の計算）

$$
\begin{aligned}
V'[S(g(\omega'))] &= E'\left[(S(g(\omega')) - 524288p)^2\right] \\
&= E'\left[\left(\sum_{k=1}^{524288}(X'_k - p)\right)\left(\sum_{l=1}^{524288}(X'_l - p)\right)\right] \\
&= \sum_{k=1}^{524288}\sum_{l=1}^{524288} E'\left[(X'_k - p)(X'_l - p)\right] \\
&= \sum_{k=1}^{524288} E'\left[(X'_k - p)^2\right] + \sum_{k\neq l}^{524288} \boxed{E'\left[X'_k - p\right]E'\left[X'_l - p\right] = 0} \\
&= \sum_{k=1}^{524288} E'\left[(X'_k - p)^2\right] \\
&= 524288\,\mathrm{E}'\left[(X'_1 - p)^2\right] = 524288\,p(1-p)
\end{aligned}
$$

よって，$E'[S(g(\omega'))] = E[S(\omega)]$.　　(q.e.d.)

数理統計学の視点から

- アリスが**自分の意思**で種$\omega' \in \{0,1\}^{292}$を選ぶという設定では，結果の**客観性**が脅かされる．
  - 実際，「ひどい種」によって故意にひどい結果を出せる．
  - 無作為抽出が必要．
- 解決策
  - 無作為抽出：**本物の硬貨**を292回投げて$\omega' \in \{0,1\}^{292}$を得る．
  - そのω'を種として$S(g(\omega'))$を計算する．
- 疑似乱数生成器gの役割
  - 本物の硬貨を67,108,864回も投げることはできない．しかし，gを用いれば292回で済み，実行可能である．

例題の解法まとめ(1)

- モンテカルロ法：「賭け」
  - アリスは $\omega \in \{0,1\}^{67108864}$ を自分の意思で選ぶ．
  - $S(\omega)$ の値を算出する．
  - $\omega \notin A$ ならアリスの勝ち，そうでなければアリスの負け．
  - 一様分布 P の下で $P(\omega \in A) < 1/32$．
- 乱数の問題
  - 乱数とは自身とほぼ同じ長さの入力を必要とする $\{0,1\}$-列．
  - $\{0,1\}^{67108864}$ の元の圧倒的多数は乱数である．
  - $P(\omega \in A) < 1/32$ に意味を持たせるには乱数から ω を選ぶべき．
  - アリスは自分の意思で乱数を選ぶことはできない．

例題の解法まとめ(2)

- 疑似乱数生成器 $g : \{0,1\}^{292} \to \{0,1\}^{67108864}$
  - アリスは自分の意思で種 $\omega' \in \{0,1\}^{292}$ を選ぶ．
  - 新しい「賭け」：$g(\omega') \notin A$ なら勝ち，$g(\omega') \in A$ なら負け．
  - $P'(g(\omega') \in A) < 1/32$ となる g(A に対して安全)が存在する．
- 数理統計学の視点
  - 本物の硬貨を投げて ω' を選べば無作為抽出が可能．

モンテカルロ積分

- 例題のように，確率変数 X の独立なコピー列 $X_1, X_2, \ldots, X_N$ の和 $S = X_1 + X_2 + \cdots + X_N$ をサンプリングし，大数の法則を根拠に $S(\omega)/N$ でもって X の平均 $E[X]$ を推定する方法を「モンテカルロ積分」という．
- モンテカルロ法のうち，およそ科学的目的を持つものはそのほとんどがモンテカルロ積分である．
- モンテカルロ積分の場合は,「安全な疑似乱数生成器 g」が存在する．
- 非常に大規模なモンテカルロ積分では，g の種 ω' の長さが巨大になり，これをプレーヤーの意思で自由に選ぶことが再び困難になる．そのような場合のモンテカルロ積分は現時点では完全には解決されていない．

一般のモンテカルロ積分

モンテカルロ積分以外のモンテカルロ法においても，その定式化はほぼ同様である．

- モンテカルロ法の目的は，ある確率変数Sの「一般的な値」をサンプリングすること．
- 乱数の問題を解決するために疑似乱数生成器を用いる．
- 一般には「計算量的に安全な疑似乱数生成器」が有効と思われるが，それが存在するかどうか不明．(とても難しい問題)

Thank you very much.

Please visit !

http://homepage.mac.com/hiroshi_sugita/mcm.html