酔歩の最長到達距離更新回数の統計的性質

奥田 一成 [1], 日永田 泰啓 [2]

[1] 佐賀大学 大学院 工学系研究科 知能情報システム学専攻
[2] 佐賀大学 総合情報基盤センター

概要

1 次元空間上を移動する、離散時間の酔歩を考える。$n(\geq 1)$ ステップ後の酔歩の位置 x_n は ($x_0 = 0$ を始点として) $x_n = x_{n-1} + \xi_n$ で決める。ただし、ξ_n は連続で対称な確率密度関数 $\phi(\xi_n)$ によって決める。位置 x_i が過去の全ての位置 $\{x_k\}_{0\leq k<i}$ よりも大きくなることを、i ステップで「一方向最長到達距離が更新された」という事にする。このような更新が起こる回数の厳密な期待値は ($\phi(\cdot)$ に依らず) n のみに依存する事が知られている。では、(i ステップで) 更新したという判断を、「〈始点からの距離〉$|x_i|$ が過去の〈始点からの距離〉$\{|x_k|\}_{0\leq k<i}$ よりも大きくなる事」で行ったなら、その更新回数の期待値の依存性はどうなるのか？　更に、酔歩する空間を 2 次元へと変更し、〈始点からのユークリッド距離〉で更新するか否かを判断したなら、その更新回数の期待値の依存性はどうなるのか？　以上の 2 つの依存性を、シミュレーションによって調べる。

Statistics of the number of records in a sequence of positions of a random walker

Kazunari Okuda [1], Yasuhiro Hieida [2]

[1] Department of Information Science, Saga University
[2] Computer and Network Center, Saga University

Abstract

We consider a discrete-time random walker on a continuous line. The position x_n of the walker at $n(\geq 1)$ steps evolves according to $x_n = x_{n-1} + \xi_n$ and $x_0 = 0$, where ξ_n is drawn from a probability density function $\phi(\xi)$ which is symmetric and continuous. We call an event that $|x_i| > |x_k|$ for $0 \leq \forall k < i$ a "record" which happens at the i-th step. By performing simulations, we study what the number of records depends upon for a one-dimensional random walker mentioned above. We study such dependence also for a two-dimensional random walker.

1　はじめに

時間毎の値がでたらめに決まるモデルを酔歩という。例えば、拡散粒子の時間毎の位置や企業の秒毎の株価の値などの動きが酔歩に該当する。酔歩は物理学、化学、数学と幅広い範囲で応用されており、多くの未解決問題が存在している。本研究では、酔歩の「最長到達距離更新回数」に注目し、その統計的性質を調べる。

1 次元空間上を移動する離散時間の酔歩を考える。$n(\geq 1)$ ステップ後の酔歩の位置 x_n を

$$x_n = x_{n-1} + \xi_n \text{ (始点は } x_0 = 0 \text{ とする)} \tag{1}$$

で決める。(1) 式の ξ_n の値は連続で対称的な確率密度関数 $\phi(\xi)$ によって決める。$\phi(\xi)$ の具体例は次の

4 式である。

$$\phi(\xi)=\begin{cases}\dfrac{1}{2a} & \left(|\xi|\le a,\ a=\dfrac{1}{2}\right)\\ 0 & (\text{それ以外})\end{cases} \tag{2}$$

$$\phi(\xi)=\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left[-\frac{\xi^2}{2\sigma^2}\right]\quad(\sigma=1) \tag{3}$$

$$\phi(\xi)=\frac{1}{2}e^{-|\xi|} \tag{4}$$

$$\phi(\xi)=\frac{\alpha}{\pi(\xi^2+\alpha^2)}\quad(\alpha=1) \tag{5}$$

以下では、(2) 式 (一様分布)、(3) 式 (正規分布)、(4) 式 (「対称化した」指数分布)、(5) 式 (コーシー分布) の各 $\phi(\xi)$ に従う酔歩を、それぞれ **U**、**N**、**E**、**C** と略記する。(1) 式に従うサンプルを生成する時に用いる擬似乱数生成法としては、線形合同法[1]もしくは Mersenne Twister 法 (以下、**MT** 法と略す) あるいはその両方を用いた。以下のすべてのシミュレーションのサンプル数は 10^6 である。すべての対数の底は 10 である。

2 最長到達距離の更新回数

2.1 一方向最長到達距離の更新回数

位置 x_i が過去の全ての位置 $\{x_k\}_{0\le k<i}$ よりも大きくなる事を i ステップで「一方向最長到達距離が更新された」という事にする。n ステップ目までの更新回数を $R(n)$ と書く事にすると、その厳密な期待値 $\langle R(n)\rangle$ は (全ての自然数 n に対して)

$$\langle R(n)\rangle=(2n+1)\binom{2n}{n}2^{-2n} \tag{6}$$

となる [1]。ただし、$R(0)=1$ とする。「一方向最長到達距離更新回数」は気象学（地球の平均気温の最高、最低気温の更新）、経済学（株価の安値または高値更新）といった幅広い分野で見ることができる。$\langle R(n)\rangle$ は、酔歩の移動時間 n にのみ依存し、($\phi(\xi)$ が連続で対称的な確率密度関数である限り) $\phi(\xi)$ によらない (従って、(2) 式、(3) 式、(5) 式 のパラメタ a、σ、α にも依らない)[1]。(6) 式から十分大きな n に対して

$$\langle R(n)\rangle\simeq\frac{2}{\sqrt{\pi}}\sqrt{n} \tag{7}$$

となる事が分かる。

2.2 両方向最長到達距離の更新回数

この節でも 2.1 節同様、酔歩の位置は (1) 式で決める。「〈始点からの距離〉$|x_i|$ が過去の〈始点からの距離〉$\{|x_k|\}_{0\le k<i}$ よりも大きくなる事」を、i ステップで「両方向最長到達距離が更新された」と言うことにする。n ステップ目までの更新回数を $r(n)$ と書く事にする (ただし $r(0)=1$ とする) と、その期待値 $\langle r(n)\rangle$ はどのような依存性を持つのか (たとえば、(6) 式のように $\phi(\xi)$ によらないのか) をシミュレーションによって調べる。

まず、各 $\phi(\xi)$ に対して $\langle r(n)\rangle$ の区間推定を行ったのが、図 1 〜図 3 である。これらの図は、シミュレーションから求めた $\langle r(n)\rangle$ の one σ の信頼区間をプロットしたものである。図 1 によれば、各分布の $\langle r(10)\rangle$ は $\phi(\xi)$ に依存している。しかし、n を大きくしていくと $n=10^6$ では図 2、図 3 のように (C 以外は) 重なる。

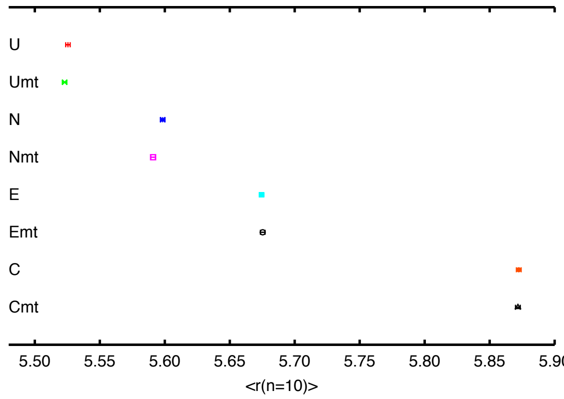

図 1: 各分布の $\langle r(n=10)\rangle$ の信頼区間の比較。縦軸の各記号は、(2) 式〜(5) 式のすぐ下で説明したもの。これらの記号に MT とついているものは MT 法、ついていないものが線形合同法を用いている。

次に、U、N、E、C のそれぞれに対する $r(n)$ の標本平均が (7) 式のように $\sqrt{n}$ に比例するかどうかを調べる。$n=10$、50、100、500、1000、10^4、10^5、10^6(ただし、U のみは $n=5000$ のデータも用いる) における標本平均 $\bar{r}(n)$ の対数をとった値を最小二乗法によって一次関数で fitting した結果は図 4 のよ

[1]Oracle の Java の `java.util.Random` クラス

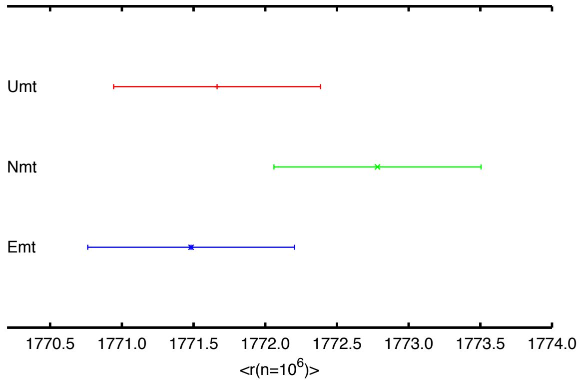

図 2: 一様分布、正規分布、指数分布の $\langle r(n = 10^6)\rangle$ の信頼区間の比較（コーシー分布を含めたものは図 3)。縦軸の記号については、図 1 を参照のこと。

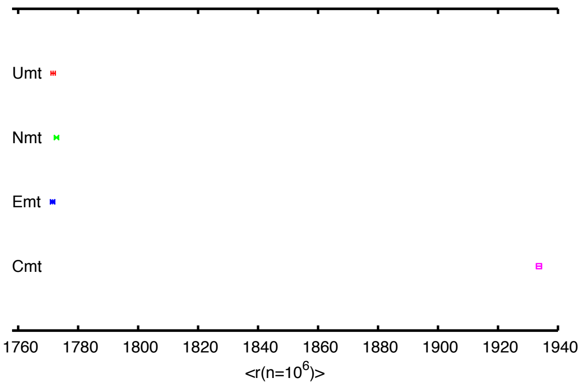

図 3: 各分布の $\langle r(n = 10^6)\rangle$ の信頼区間の比較。縦軸の記号については、図 1 を参照のこと。

うになった。以下は fitting 結果の式である。

$$\mathrm{U}: \log\langle r(n)\rangle = 0.501396 \log n + 0.241452 \tag{8}$$

$$\mathrm{N}: \log\langle r(n)\rangle = 0.500497 \log n + 0.245903 \tag{9}$$

$$\mathrm{E}: \log\langle r(n)\rangle = 0.499229 \log n + 0.251947 \tag{10}$$

$$\mathrm{C}: \log\langle r(n)\rangle = 0.503176 \log n + 0.271026 \tag{11}$$

2.3 2 次元最長到達距離の更新回数

前節では、更新するか否かを原点 (始点) からの距離で判断したのだった。では、その距離 (始点からのユークリッド距離) の測り方を用いて、かつ、酔歩する空間を 2 次元に変更したなら、n ステップまで移動した時の更新回数 $P(n)$ (ただし $P(0) = 1$ とする) の n 依存性等はどうなるのか？ このことを調べる。

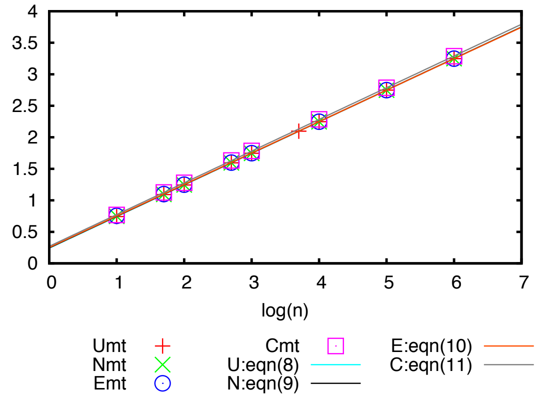

図 4: U、N、E、C それぞれの標本平均 $\bar{r}(n)$ の対数のデータ点 (乱数生成は MT 法による) と fit した直線 ((8) 式〜(11) 式) との比較。記号同士、直線同士はほとんど重なっている。縦軸は $\log(\bar{r}(n))$。すべての対数の底は 10。

$n(\geq 1)$ ステップ後の位置は ((1) 式の代わりに)

$$\vec{x}_n = \vec{x}_{n-1} + \vec{\xi}_n \tag{12}$$

で決める。ただし $\vec{x}_n = \vec{0}$ を始点とする。(12) 式の $\vec{\xi}_n$ は、単位円内に一様分布するもの (**UC** と略記) と、$\vec{\xi}_n$ の x 成分、y 成分それぞれが独立に標準正規分布に従うもの (**N2D** と略記) の 2 種類を考える。

UC の場合と N2D の場合とで、$\langle P(n)\rangle$ の値がどのように変化するかを見るために、(12) 式のシミュレーションを行い、$\langle P(n)\rangle$ の区間推定を行った。図 5 〜図 7 は、$n = 10, 10^2, 10^3, 10^4, 10^5$ と n を変えたとき、$\langle P(n)\rangle$ の one σ の信頼区間をプロットしたものである。$n = 10$ のとき UC と N2D の信頼区間は重ならない (図 5) が、$n = 10^4$ においては信頼区間は重なる (図 6)。図 7 は $n = 10, 10^2, 10^3, 10^4, 10^5$ と変えたときの各分布の $\langle P(n)\rangle$ の信頼区間の比較を行ったものである。以上の図から、n が大きい場合は、$P(n)$ の標本平均は、UC であるか N2D であるかに依存しないように見える。

次に、$P(n)$ の標本平均が (7) 式のような振舞いをするのかどうか調べる。そのために、$n = 10, 10^2, 10^3, 10^4, 10^5$ における $P(n)$ の標本平均の対数をとった値を最小二乗法によって一次関数で fitting した。その結果は図 8 のようになった。fit に用いた直線の式は

$$\mathrm{UC}: \log\langle P(n)\rangle = 0.51705 \log n + 0.294835 \tag{13}$$

$$\mathrm{N2D}: \log\langle P(n)\rangle = 0.516678 \log n + 0.296273 \tag{14}$$

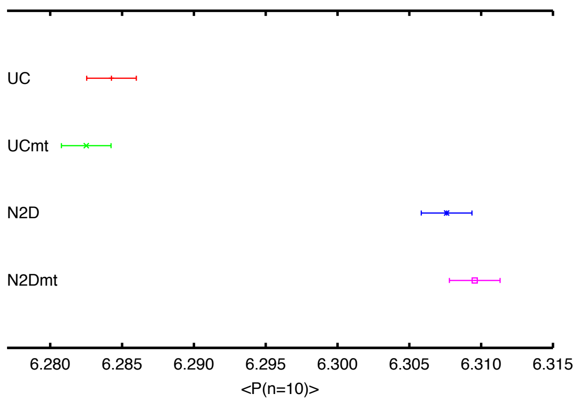

図 5: 各分布の $\langle P(n=10)\rangle$ の信頼区間の比較。縦軸の各記号は、本文の 2.3 節を参照の事。各記号に MT とついているものが MT 法、ついていないものが線形合同法を用いている。

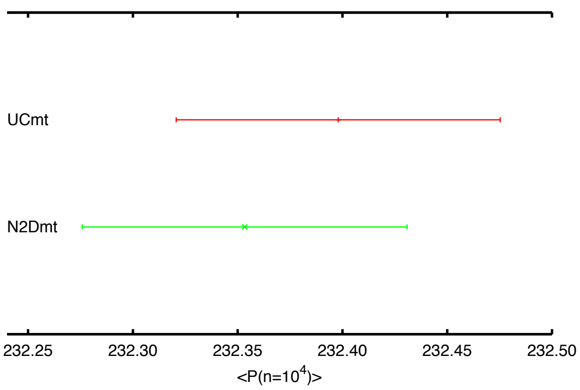

図 6: 各分布の $\langle P(n=10^4)\rangle$ の信頼区間の比較。縦軸の各記号は、図 5 と同じもの。

であり、傾きは 1/2 に近い値となっている。

3　まとめ

本研究では、酔歩の (始点からの距離に関する) 最長到達距離の更新回数をシミュレーションによって調べた。その結果、(1 次元空間での) 両方向最長到達距離の更新回数 (2.2 節) の標本平均は、10^5 ステップまでは (1) 式の ξ_n を決める個々の確率密度関数 ((2) 式〜(5) 式) に依存することがわかった。しかし、ステップ数 n を大きくすると、たとえば、$n=10^6$ では、(コーシー分布 (5) 式以外すなわち) 一様分布 (2) 式、正規分布 (3) 式、指数分布 (4) 式のそれぞれの信頼区間が重なるような振舞いを示した (図 2、図 3)。標本平均 $\bar{r}(n)$ は、一方向最長到達距離の更新回数の場合 (2.1 節) の (7) 式のように、$\sqrt{n}$ に比例するように見える (図 4)。

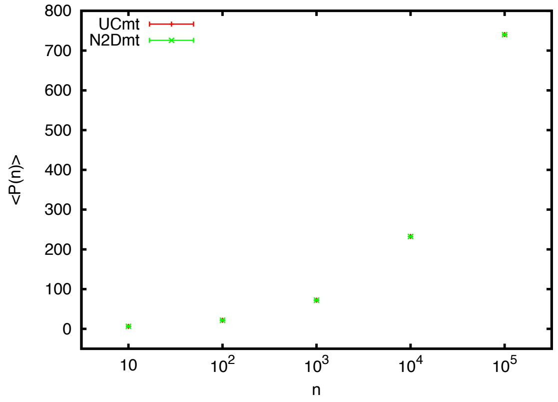

図 7: $n=10, 10^2, 10^3, 10^4, 10^5$ と変えたときの各分布の $\langle P(n)\rangle$ の信頼区間の比較 (2 つの信頼区間はほとんど重なっている)。`UCmt`、`N2Dmt` という記号については図 5 を参照の事。

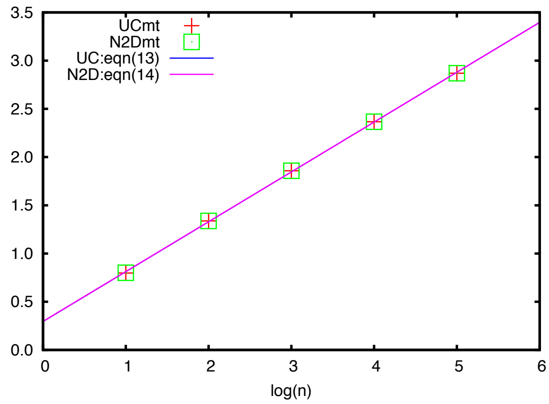

図 8: $n=10, 10^2, 10^3, 10^4, 10^5$ における各分布の標本平均 $\bar{P}(n)$ の対数のデータ点 (乱数生成は MT 法による) と fit した直線である (13) 式と (14) 式 (これら 2 直線はほとんど重なっている) の比較。縦軸は $\log\left(\bar{P}(n)\right)$。すべての対数の底は 10。

2 次元空間での (n ステップまでの) 最長到達距離の更新回数 $P(n)$ (2.3 節) も、$n=10$ では $\vec{\xi}_n$ の分布に依存するが (図 5)、n が大きくなると UC であるか N2D であるかに依存しないことがわかった (図 6、図 7)。$P(n)$ の標本平均は ((7) 式のように)n^b に比例する事が分かった (図 8)。ここで、b は 1/2 に近い数 ((13) 式、(14) 式それぞれの右辺第 1 項の log の係数) である。

参考文献

[1] Satya N. Majumdar, Physica A **389** (2010) 4299-4316.