数理解析研究所講究録
第1649巻 2009年 9-16

玉と壷のモデルにおける玉の衝突について

福岡教育大学・数学教育講座　中田 寿夫 (Toshio NAKATA)
Department of Mathematics,
Fukuoka University of Education

概 要

ランダムなアルゴリズムのモデルの基礎となっている「玉と壷の問題」において、Wendl (*Stat. Prob. Lett.*, (2003)) は2つの種類の玉の衝突のモデルを提案し、玉が衝突しない確率を計算した。ここでは、衝突の個数についての正確な分布と階乗モーメントを求め、ポアソン分布への収束を示した。さらには「Chen-Steinの方法」と呼ばれる方法を用いてポアソン近似の誤差を評価した。

キーワード：衝突確率, 玉と壷のモデル, ポアソン近似, Chen-Steinの方法

1 はじめに

いくつかの玉をいくつかの壷の中に無作為に入れる、いわゆる「玉と壷の問題」は理論計算機科学の分野ではランダムなアルゴリズムのモデルとして扱われている。例えば、ランダムなアルゴリズムの標準的教科書 ([10, 11]) には、バケットソートやハッシュなどの応用についての記述がある。

一方で、玉と壷の問題に関しての数学的なとり扱いについて、歴史的にはラプラスの時代まで遡り、多くの研究が知られている。1970年代までの内容はJohnson, Kotz [7] とKolchin et. al. [8] という標準的な本に纏められている。さらに、2000年以降でもアルゴリズムの解析のコミュニティや離散確率論のコミュニティにおいて、アルゴリズムへの応用や数学的な興味からの極限定理の研究など多くの研究がすすんでいる ([2, 6] とその文献参照)。このことは計算機科学やバイオサイエンスのモデルとなっていることが研究の動機付けとなっているが、その中の一つとして、Wendl [14] は玉と壷のモデルにおける衝突について次のモデルを提案した：

問題 $m, n, t \geq 1$ を整数とする。白 m 個のボールをグループ A、黒 n 個のボールをグループ B とし、それぞれ t 個の壷に

でたらめに入れるときに、t 個の壷のうちにグループ A, B の衝突が k 個起きる確率はいくらか？[1]

Wendl は飛行機が空中で衝突する問題、DNA クローンマッピング問題などのモデルとして上記を定式化し、衝突が起きない確率を正確に求めた。

この小文ではそれを発展させた内容 [12, 13] について紹介する。具体的には衝突の個数についての正確な分布と階乗モーメントを導き、ポアソン分布への収束を紹介する。さらに、必ずしも独立性をみたさない確率変数列に関する「Chen-Stein の方法」と呼ばれる手法を用いてポアソン近似の誤差を評価する。

1.1 この小文の内容

§2において衝突が起きない確率を求めた Wendl [14] による結果を紹介し、それ以降は [12, 13] による Wendl の結果の発展的な内容について紹介する。具体的には §3 において正確な確率分布ならびに階乗モーメントを記述した。§3.1においては [12, 13] に関しての補足的な証明を行った。§4においてはポアソン分布への収束に関しての結果を記述した。§4.1において強い条件の下でポアソン分布への収束の結果を記述し、§4.2において Chen-Stein の手法による誤差の評価について述べた。

2 玉の衝突の起きない確率 (Wendl の結果)

この報告で扱う重要な変数は以下のものである：

定義 2.1 $X = X(m, n, t)$ を 問題 における A と B の衝突の壷の数とする (確率変数である)。

Wendl は全く衝突が起きない確率 $\mathrm{P}(X = 0)$ を正確に求めた。証明にはグラフ理論の道具を用いている。

定理 2.1 ([14])

$$\mathrm{P}(X = 0) = \frac{1}{t^{m+n}} \sum_{i \geq 0} \sum_{j \geq 0} \left\{ {m \atop i} \right\} \left\{ {n \atop j} \right\} (t)_{i+j}. \tag{1}$$

[1] Wendl [14] では、直接的にこのような形で問題を提起しているわけではない。著者の興味を持つ「玉と壷のモデル」として説明するとこのようになる。

ただし、式(1)で用いられた記号は以下のとおりである：

- $\left\{ {n \atop k} \right\}$ はn個のものをk個のグループに分ける個数で、第2種スターリング数と呼ばれるもので、任意の整数n, kについて以下のように帰納的に定義される：$\left\{ {0 \atop 0} \right\} = 1$ かつ $n, k \neq 0$ について $\left\{ {n \atop 0} \right\} = \left\{ {0 \atop k} \right\} = 0$ であり、
$$\left\{ {n \atop k} \right\} = \left\{ {n-1 \atop k-1} \right\} + k \left\{ {n-1 \atop k} \right\}.$$
この記法はいくつかの表記方法があるが計算機科学の標準的教科書[9]に従った。

- $(t)_k$ は、順列の記号で整数$t, k \geq 0$について $(t)_k = \binom{t}{k} k! = t(t-1)\cdots(t-k+1)$ かつ $(t)_0 = 1$ for $t \geq 0$ と定義する。この記法は確率論の標準的教科書[3]に従った。

3 正確な確率分布についての結果

[12, 13]では、$\mathrm{P}(X=0)$だけではなく、一般の分布 $\mathrm{P}(X=k)$ を求めた。さらに平均を含む階乗モーメントについての厳密な形を導いた。

定理 3.1 m 個の玉と n 個の玉を t 個の壷に投げたとき、衝突する壷の個数が k となる確率 $\mathrm{P}(X=k)$ は $0 \leq k \leq \min\{m, n, t\}$ について

$$\mathrm{P}(X=k) = \frac{1}{t^{m+n}} \sum_{i \geq k} \sum_{j \geq k} \left\{ {m \atop i} \right\} \left\{ {n \atop j} \right\} \frac{(i)_k (j)_k (t)_{i+j-k}}{k!}. \tag{2}$$

階乗モーメントは

$$\mathrm{E}((X)_l) = (t)_l \left(\sum_{i=0}^{l} \binom{l}{i} (-1)^i \left(1 - \frac{i}{t}\right)^m \right) \left(\sum_{j=0}^{l} \binom{l}{j} (-1)^j \left(1 - \frac{j}{t}\right)^n \right). \tag{3}$$

特にその平均 $\mathrm{E}(X)$ は以下のとおり：

$$\mathrm{E}(X) = t\left(1 - \left(1 - \frac{1}{t}\right)^n\right)\left(1 - \left(1 - \frac{1}{t}\right)^m\right). \tag{4}$$

証明は[12, 13]に譲る。ただし、[12, 13]では直接的な手法で式(3)を示した。それにより、([5, 9]にも現れないような)面倒な二項係数の変形をいくつか用いた見通しの悪いものとなっている。この小文では、それを避

けるべく、§3.1において、指示関数を用いた手法で(比較的)簡易的な証明を与える。また、X のいくつかの母関数が明確に計算可能であり、母関数から式(3)を証明することもできるが、[2, 6, 8] にあるような複素解析の手法を用いるためここでは割愛する。

3.1 指示関数を用いた式(3)の簡易的な証明

$A(i)$ と $B(i)$ の集合をそれぞれ i 番目の壷において A と B からの玉が存在する(衝突が起きる)という事象とする。このとき、$A(i)$ と $B(i)$ は玉の投げ入れの独立性が遺伝して、独立な事象であることが分かる。結合事象の確率を計算すると

$$\begin{aligned} \mathrm{P}\left(\bigcap_{i=1}^{l} A(i)\right) &= 1-\mathrm{P}\left(\bigcup_{i=1}^{l} A(i)^{\mathrm{c}}\right) = 1-\sum_{i=1}^{l}\binom{l}{i}(-1)^{i+1}\mathrm{P}\left(\bigcap_{j=1}^{i} A(j)^{\mathrm{c}}\right) \\ &= 1-\sum_{i=1}^{l}\binom{l}{i}(-1)^{i+1}\left(1-\frac{i}{t}\right)^{m} = \sum_{i=0}^{l}\binom{l}{i}(-1)^{i}\left(1-\frac{i}{t}\right)^{m} \end{aligned} \tag{5}$$

である。同様に $\mathrm{P}\left(\bigcap_{i=1}^{l} B(i)\right)$ も計算できる。

さらに、$X=\sum_{i=1}^{t}\xi_i$ とする。ここで、

$$\xi_i = \begin{cases} 1, & i\text{番目の壷で衝突が起きる} \\ 0, & i\text{番目の壷で衝突が起きない} \end{cases} \tag{6}$$

である。よって、

$$\mathrm{E}((X)_l) = \mathrm{E}\left\{\prod_{j=0}^{l-1}\left(\sum_{i=1}^{t}\xi_i - j\right)\right\} = \sum_{\{i_1<\cdots<i_l\}\subset\{1,\cdots,t\}} \mathrm{E}(\xi_{i_1}\cdots\xi_{i_l}) = (t)_l\mathrm{E}(\xi_1\cdots\xi_l). \tag{7}$$

ここで、

$$\begin{aligned} \mathrm{E}(\xi_1\cdots\xi_l) &= \mathrm{P}\left(\xi_1=1,\cdots,\xi_l=1\right)\mathrm{P}\left(\bigcap_{i=1}^{l}(A(i)\cap B(i))\right) \\ &= \mathrm{P}\left(\left\{\bigcap_{i=1}^{l}A(i)\right\}\cap\left\{\bigcap_{i=1}^{l}B(i)\right\}\right) = \mathrm{P}\left(\bigcap_{i=1}^{l}A(i)\right)\mathrm{P}\left(\bigcap_{i=1}^{l}B(i)\right) \end{aligned}$$

である($A(i), B(i)$ の独立性を用いた)。最後に式(5)により証明が終わる。

注意 3.1 確率変数列 $\{\xi_i\}$ は独立でないが、可換(定義は [4, VII.4])であることにより式(7)の最後の等号が成立している。

4 ポアソン分布への収束に関する結果

まず、玉をたくさん投げ入れたときの $\mathrm{E}(X(m,n,t))$ について調べてみよう。ここでは次の記号を用いる。

- $a_n = \omega(b_n)$ は $\lim_{n\to\infty} \frac{a_n}{b_n} = \infty$ で定義される。
- $a_n \sim b_n$ は $\lim_{n\to\infty} \frac{a_n}{b_n} = 1$ で定義される。

O, o は通常の意味で用いられる。式 (4) により次がすぐにわかる。

(i) $\min\{m,n\} = \omega(t)$ ならば $\mathrm{E}(X(m,n,t))/t \sim 1$.

(ii) $\min\{m,n\} = o(t)$ ならば $\mathrm{E}(X(m,n,t))/t = o(1)$.

(iii) 特に $m \sim \alpha t$ かつ $m \sim \beta t$ $(\alpha, \beta > 0)$ ならば

$$\mathrm{E}(X(m,n,t))/t \sim (1-e^{-\alpha})(1-e^{-\beta}). \tag{8}$$

式 (8) をより詳しく言うと、X/t について下記のことが言える。

命題 4.1 (大数の弱法則) 玉の数が壷の数に比例するくらいのオーダであれば、詳しく言うと $m \sim \alpha t$ かつ $n \sim \beta t$ $(\alpha, \beta > 0)$ ならば、全体の壷の中での衝突の割合は以下のようになる：

$$\lim_{t\to\infty} \frac{X}{t} = (1-e^{-\alpha})(1-e^{-\beta}) \quad \text{(確率収束)}.$$

4.1 ポアソンの小数の法則

少し記号を準備する。

- $\mathcal{L}(X)$ を確率変数 X の分布と定義する。
- 2 つの分布の全変動距離を以下で定義する：

$$\mathrm{d_{TV}}(\mathcal{L}(X), \mathcal{L}(Y)) = \frac{1}{2}\sum_{k\geq 0} |\mathrm{P}(X=k) - \mathrm{P}(Y=k)|.$$

- 分布の収束について以下の記法を用いる：

$$\lim_{t\to\infty} \mathcal{L}(X_t) \stackrel{\mathrm{d}}{=} \mathcal{L}(Y).$$

これは $\lim_{t\to\infty} \mathrm{E}(f(X_t)) = \mathrm{E}(f(Y))$ が任意の有界連続関数 f でなりたつと定義される。

- $\mathrm{Poi}(\gamma)$ をパラメータ γ のポアソン分布とする：

$$\mathrm{P}(Y=k)=e^{-\gamma}\frac{\gamma^k}{k!} \quad \text{for } k=0,1,\cdots.$$

このとき次が言える。

定理 4.1 (ポアソンの小数の法則) 壷の数が玉の数と比べて多い場合、より詳しく言うと $\max\{m,n\}=o(t)$ かつ $mn\sim\gamma t$ ただし $\gamma>0$ という場合は次がなりたつ：

$$\lim_{t\to\infty}\mathrm{E}(X(m,n,t))=\gamma.$$
$$\lim_{t\to\infty}\mathcal{L}(X(m,n,t))\stackrel{\mathrm{d}}{=}\mathrm{Poi}(\gamma).$$

注意 4.1 独立で出現確率が小さいベルヌイ確率変数 (0,1 の値しかとらない確率変数) の和については、確率論の一般論を用いてポアソンの小数の法則が成立する。しかし、このケースは式 (6) で定義した $\{\xi_i\}$ の独立性がなりたたないのでそのような一般論は使えない。[13] では階乗モーメントからモーメントの収束を示して証明した。

4.2 収束の誤差の解析 —Chen-Stein の方法—

定理 4.1に関しては収束だけを示したわけであるが、係数を含めて収束のオーダが分かれば実用上ありがたいことが多い。このモデルでは、Chen-Stein の方法と呼ばれる、係数を含めたポアソン分布への収束のオーダを与える強力な手法が効果的に利用できるので紹介する。Chen-Stein の方法は 1970 年代に開発され、[1] は全章を通してそれについて書かれている (歴史的背景や発想のアイデアが第 1 章に詳しく記述されている)。

その前に記号 λ,Δ を準備する。

$$\lambda=\mathrm{E}(X)=t\left(1-\left(1-\frac{1}{t}\right)^m\right)\left(1-\left(1-\frac{1}{t}\right)^n\right),$$

$$\Delta = \frac{t(t-1)}{2}\left(1-2\left(1-\frac{1}{t}\right)^m+\left(1-\frac{2}{t}\right)^m\right)\left(1-2\left(1-\frac{1}{t}\right)^n+\left(1-\frac{2}{t}\right)^n\right)$$

t	λ	誤差の限界
10	4.24219	0.67558
50	1.67311	0.29416
100	0.91427	0.15247
200	0.47804	4.2523×10^{-2}
500	0.19643	7.2697×10^{-3}
1000	0.09910	1.8582×10^{-3}

表 1: 式 (9) について $m = n = 10$ としたときの t, λ と誤差の限界

定理 4.2 (Chen-Stein の評価式) λ, Δ の記号のもとで、下記の不等式がなりたつ：

$$\mathrm{d_{TV}}(\mathcal{L}(X), \mathrm{Poi}(\lambda)) \leq \min\left\{1, \frac{1}{\lambda}\right\}(\lambda^2 - 2\Delta). \tag{9}$$

注意 4.2 壷の数 t が十分大きければ、衝突が滅多に起こらずポアソン近似はうまくいくことがわかる。例えば、表 4.2 は 式 (9) について $m = n = 10$ としたときの t, λ と誤差の限界 (式 (9) の右辺) を表している。

5 おわりに

玉と壷のモデルにおいての衝突確率について議論し、衝突の個数についての正確な階乗モーメントを求め、ポアソン分布への収束を示した。さらには「Chen-Stein の方法」と呼ばれる方法を用いてポアソン近似の誤差を評価した。

ポアソンの小数の法則だけではなく、強いの条件の下では中心極限定理もなりたつことがわかっている。ただし、[6] の条件のような広い条件の下での中心極限定理を示すには、母関数に関して解析的に精緻な調査が必要となってきてそれほど容易なことではない。現在は中心極限定理についての一般論を構築中である。

参考文献

[1] Barbour, A. D., Holst, L. and Janson, S. (1992). *Poisson Approximation*, Oxford UP.

[2] FLAJOLET, P. AND SEDGEWICK, R. (2008). *Analytic Combinatorics*, Cambridge Univ. Press.

[3] FELLER, W. (1968). *An Introduction to Probability Theory and Its Applications*, Vol. 1, 3rd ed. Wiley.

[4] FELLER, W. (1971). *An Introduction to Probability Theory and Its Applications*, Vol. 2, 2rd ed. Wiley.

[5] グレアム, クヌース, パタシュニク著, 有澤他訳, (1993). コンピュータの数学, 共立出版.

[6] HWANG, H., JANSON, S.(2008). Local limit theorems for finite and infinite urn models, *Ann. Probab.* **36**, No. 3, 992-1022.

[7] JOHNSON, N. L., AND KOTZ, S. (1977). *Urn Models and Their Application*, John Wiley, New York.

[8] KOLCHIN, F., SEVASTYANOV, A. AND CHISTYAKOV, P. (1978). *Random allocations*, John Wiley, New York.

[9] KNUTH, D. (1973). *The Art of Computer Programming, Fundamental Algorithms*, Volume 1. 3rd edn. Addison-Wesley.

[10] MITZENMACHER, M., UPFAL, E. (2005). *Probability and Computing: Randomized Algorithms and Probabilistic Analysis*, Cambridge Univ. Press.

[11] MOTWANI, M., RAGHAVAN, P. (1995). *Randomized Algorithms*, Cambridge Univ. Press.

[12] NAKATA, T. (2008). Collision probability for an occupancy problem, *Stat. Prob. Lett.*, **78**, 1929–1932.

[13] NAKATA, T. (2008). A Poisson approximation for an occupancy problem with collisions, *J. Appl. Prob.* **45**, No. 2, 430–439.

[14] WENDL, M. (2003). Collision probability between sets of random variables, *Stat. Prob. Lett.* **64**, 249–254.