Characterisation of equalisation problems via random walks

TOSHIO NAKATA

1. *Introduction*

This is a follow-up article to a recent Gazette article by Abel [1]. We investigate probabilistic equalisation problems proposed by [1] using the *Pólya urn*, and characterise them through simple random walks.

Let us consider that Alice has a box with 50 blue pens and 50 red pens. She writes a one-page daily diary for 40 days using either colour pen chosen at random. For each setting (I)–(III) below, calculate the probability that the number of pages written with the red pen is 20.

(I) She does not put the pen back in the box but throws it away every day.

(II) She puts the pen back in the box every day.

(III) She puts the pen back in the box every day. In addition, her father buys another pen of the same colour as a reward for her efforts, and she also puts it in the box.

The probabilities are respectively

$$\begin{array}{rll} \text{(I)} & \binom{50}{20}^2\binom{2\cdot 50}{40}^{-1} & \approx 0.16158, \\ \text{(II)} & \binom{40}{20}2^{-40} & \approx 0.12537, \\ \text{(III)} & \binom{50+20-1}{20}^2\binom{2\cdot 50+40-1}{40}^{-1} & \approx 0.10588. \end{array}$$

Problems (I) and (II), from [1], are straightforward, but Problem (III), added here, is not so obvious. When dealing with such problems, it is better to formulate them generally. We suppose that there are n blue pens and n red pens in the box, and she writes her diary for $2k$ days. Let $X^-_{2k,2n}, X_{2k}$ and $X^+_{2k,2n}$ be the number of pages written in red for cases (I), (II) and (III), respectively. Then we have

$$\begin{cases} u^-_{2k,2n} & = \binom{n}{k}^2\binom{2n}{2k}^{-1}, \\ u_{2k} & = \binom{2k}{k}2^{-2k}, \\ u^+_{2k,2n} & = \binom{n+k-1}{k}^2\binom{2n+2k-1}{2k}^{-1}, \end{cases} \tag{1}$$

where

$$u^-_{2k,2n} = \mathrm{P}(X^-_{2k,2n} = k), \quad u_{2k} = \mathrm{P}(X_{2k} = k), \quad u^+_{2k,2n} = \mathrm{P}(X^+_{2k,2n} = k).$$

In Section 2, we state that all equations in (1) will be calculated by the Pólya urn. In Section 3, we characterise $u^-_{2k,2n}, u_{2k}$ and $u^+_{2k,2n}$ via random walks. In fact, the probabilities for Problems (I)–(III) are explained through events that one dimensional simple random walks return to the origin. As a related study, there is *the Feller game* that considers the recursion time of the random walk as a game. (See [2, p.246, Sec X.1], [3, 4]). While Abel [1] investigated the relative error of $u^-_{2k,2n}$ with

$$r_-(2k, 2n) = \frac{u^-_{2k,2n} - u_{2k}}{u_{2k}},$$

in Section 4, we study the relative error of $u^+_{2k,2n}$ with

$$r_+(2k, 2n) = \frac{u^+_{2k,2n} - u_{2k}}{u_{2k}}.$$

Finally, in Section 5, we give some concluding remarks.

2. *The Pólya urns*

The key point of Problems (I)–(III) is that the number of pens will increase or decrease randomly according to each setting, but there is a general model called the *Pólya urn* (See [2, 5, 6]).

For an arbitrary integer s, let $X^{(s)}_{2k,2n}$ be the number of pages written in red when s additional pens are added after each draw for our n and k settings. It follows that

$$u^{(s)}_{2k,2n} = \binom{2k}{k} \frac{\{(n)_{s;k}\}^2}{(2n)_{s;2k}}, \tag{2}$$

where $u^{(s)}_{2k,2n} = \mathrm{P}(X^{(s)}_{2k,2n} = k)$ and $(z)_{s;k}$ denotes *s-shifted factorial*

$$(z)_{s;k} = \begin{cases} 1, & \text{if } k = 0, \\ z(z+s)\cdots(z+(k-1)s), & \text{if } k = 1, 2, \ldots. \end{cases}$$

Note that (2) is a special case of the distribution of the Pólya urn (See [2, p.120, Equation (2.3)], [5, p.177, Equation (4.1)] and [6, p.51, Theorem 3.1]. Hence we have

$$u^-_{2k,2n} = u^{(-1)}_{2k,2n}, \quad u_{2k} = u^{(0)}_{2k,2n}, \quad u^+_{2k,2n} = u^{(1)}_{2k,2n}.$$

Equations

$$\begin{cases} (z)_{-1;k} &= z(z-1)\cdots(z-k+1) = k!\binom{z}{k}, \\ (z)_{0;k} &= z^k, \\ (z)_{1;k} &= z(z+1)\cdots(z+k-1) = k!\binom{z+k-1}{k} \end{cases}$$

yield

$$\begin{cases} u^{(-1)}_{2k,2n} = \binom{2k}{k}\frac{\left\{(n)_{-1;k}\right\}^2}{(2n)_{-1;2k}} &= \binom{2k}{k}\frac{\left\{k!\binom{n}{k}\right\}^2}{(2k)!\binom{2n}{2k}}, \\ u^{(0)}_{2k,2n} = \binom{2k}{k}\frac{\{(n)_{0;k}\}^2}{(2n)_{0;2k}} &= \binom{2k}{k}2^{-2k}, \\ u^{(1)}_{2k,2n} = \binom{2k}{k}\frac{\left\{(n)_{1;k}\right\}^2}{(2n)_{1;2k}} &= \binom{2k}{k}\frac{\left\{k!\binom{n+k-1}{k}\right\}^2}{(2k)!\binom{2n+2k-1}{2k}}, \end{cases}$$

which establishes (1). It seems to be worthwhile that the Pólya urn can handle these problems in a unified way.

3. *Characterisation via random walks*

We define a random walk $\{S_n\}$ as $S_0 = 0$ and $S_n = \sum_{i=1}^n \xi_i$ for $n = 1, 2, \ldots$, where $\xi_1, \ldots, \xi_n$ are independent and identically distributed random variables with $\mathrm{P}(\xi_i = 1) = \mathrm{P}(\xi_i = -1) = 1/2$. Then $u^-_{2k,2n}, u_{2k}$ and $u^+_{2k,2n}$ are characterised by the following.

Theorem: Letting n be a positive integer, we have

$$u^-_{2k,2n} = \mathrm{P}(S_{2k} = 0 \mid S_{2n} = 0) \quad \text{for } k = 0, 1, \ldots, n, \tag{3}$$

$$u_{2k} = \mathrm{P}(S_{2k} = 0) \quad \text{for } k = 0, 1, \ldots, \tag{4}$$

$$u^+_{2k,2n} = \frac{n}{n+k}\mathrm{P}(S_{2k} = 0 \mid S_{2n+2k} = 0) \quad \text{for } k = 0, 1, \ldots. \tag{5}$$

Proof: First of all, we check (4). The event $\{S_{2k} = 0\}$ means that the random walk returns to the origin with $2k$ steps, and is presented by

$$|\{1 \le i \le 2k : \xi_i = 1\}| = |\{1 \le i \le 2k : \xi_i = -1\}| = k,$$

where $|A|$ denotes the number of elements of the set A. Counting paths of the random walks implies (4). It is also described in [2, p. 75, III.2, Equation (2.3)] including the symbol u_{2k}.

Next we show (3) and (5) using u_{2k}. The right-hand side of (3) is

$$\mathrm{P}(S_{2k} = 0 \mid S_{2n} = 0) = \frac{\mathrm{P}\left(S_{2k} = 0\right)\mathrm{P}\left(S_{2n-2k} = 0\right)}{\mathrm{P}\left(S_{2n} = 0\right)},$$

because

$$\mathrm{P}\left(S_{2k}=0, S_{2n}=0\right)=\mathrm{P}\left(\sum_{i=1}^{2k}\xi_i=0, \sum_{j=2k+1}^{2n}\xi_j=0\right)$$
$$= \mathrm{P}\left(\sum_{i=1}^{2k}\xi_i=0\right)\mathrm{P}\left(\sum_{j=2k+1}^{2n}\xi_j=0\right)=\mathrm{P}\left(S_{2k}=0\right)\mathrm{P}\left(S_{2n-2k}=0\right),$$

which follows from the independent and stationary increments property of $\{S_n\}$. Similarly, the second factor of the right hand side of (5) is also

$$\mathrm{P}(S_{2k}=0 \mid S_{2n+2k}=0)=\frac{\mathrm{P}\left(S_{2k}=0\right)\mathrm{P}\left(S_{2n}=0\right)}{\mathrm{P}\left(S_{2n+2k}=0\right)}.$$

Consequently, if

$$u^-_{2k,2n} = \frac{u_{2k}u_{2n-2k}}{u_{2n}}, \tag{6}$$
$$u^+_{2k,2n} = \frac{n}{n+k}\frac{u_{2k}u_{2n}}{u_{2n+2k}} \tag{7}$$

hold, then (3) and (5) follow because of (4). Therefore, we check (6) and (7). Noting

$$u_{2n}=\binom{n-\frac{1}{2}}{n} \tag{8}$$

(see [7, p.186, Equation (5.36)]), we transform $u^-_{2k,2n}$ and $u^+_{2k,2n}$. Since

$$\frac{u^-_{2k,2n}}{u_{2k}}=\frac{\binom{n}{k}}{\binom{n-\frac{1}{2}}{k}} \tag{9}$$

(see [1, Equation (3)]), and

$$\frac{\binom{n}{k}}{\binom{r}{k}}=\frac{\binom{r-k}{n-k}}{\binom{r}{n}} \quad \text{for } r\in\mathbb{R} \text{ and integers } n\geqq k\geqq 0 \tag{10}$$

(see [7, p.167, Equation (5.21)]), we have

$$\frac{u^-_{2k,2n}}{u_{2k}}\overset{(9)}{=}\frac{\binom{n}{k}}{\binom{n-\frac{1}{2}}{k}}\overset{(10)}{=}\frac{\binom{n-k-\frac{1}{2}}{n-k}}{\binom{n-\frac{1}{2}}{n}}\overset{(8)}{=}\frac{u_{2n-2k}}{u_{2n}},$$

which implies (6). It is clear that

$$\binom{n-\frac{1}{2}}{n-1} = 2nu_{2n} \tag{11}$$

from (8), and

$$\binom{2n+2k-1}{2k}\binom{2k}{k}2^{-2k} = \binom{n+k-\frac{1}{2}}{k}\binom{n+k-1}{k} \tag{12}$$

from [7, p.186, Equation (5.35)] with $r = n+k-1/2$. Hence it follows that

$$\begin{aligned} \frac{u^{+}_{2k,2n}}{u_{2k}} &\overset{(1)}{=} \frac{\binom{n+k-1}{k}^2}{\binom{2k}{k}} \frac{2^{2k}}{\binom{2n+2k-1}{2k}} \overset{(12)}{=} \frac{\binom{n+k-1}{k}}{\binom{n+k-\frac{1}{2}}{k}} \\ &\overset{(10)}{=} \frac{\binom{n-\frac{1}{2}}{n-1}}{\binom{n+k-\frac{1}{2}}{n+k-1}} \overset{(11)}{=} \frac{nu_{2n}}{(n+k)u_{2n+2k}}, \end{aligned} \tag{13}$$

which implies (7).

4. Relative errors

Abel [1] showed $r_-(2k, 2n) \geqq 0$ for $k = 0, 1, \ldots, n$, and the equality occurs only in the case of $k = 0$. Besides this, we also see $r_+(2k, 2n) \leqq 0$ for $k = 0, 1, \ldots$, and the equality also occurs only in the case of $k = 0$ because $\binom{n+k-1}{k} \leqq \binom{n+k-\frac{1}{2}}{k}$ in (13).

Although Abel [1] studied the asymptotic behaviour of $r_-(2k, 2k) = \frac{u^{-}_{2k,2k}}{u_{2k}} - 1$ for $k \to \infty$ using the Catalan numbers, we use

$$u_{2k} \sim \frac{1}{\sqrt{\pi k}}, \tag{14}$$

which is well-known when showing the recurrence of random walks (see [2, p.75, Equation (2.4)]), where $a_k \sim b_k$ denotes $\lim_{k\to\infty} a_k/b_k = 1$. Then (14) directly implies

$$r_-(2k, 2k) + 1 = \frac{u^{-}_{2k,2k}}{u_{2k}} \overset{(9)}{=} \frac{1}{\binom{k-\frac{1}{2}}{k}} \overset{(8)}{=} \frac{1}{u_{2k}} \sim \sqrt{\pi k}.$$

Since $u^{-}_{2k,2k} = 1$ by (1), the asymptotic behaviour of $r_-(2k, 2k) + 1$ is the same as the reciprocal of (14). In contrast, when the process with (III) is started from one blue pen and one red pen, it turns out that

$$r_+(2k, 2) + 1 = \frac{u^+_{2k,2}}{u_{2k}} \overset{(13)}{=} \frac{1}{\binom{k+\frac{1}{2}}{k}} \overset{(8),(11)}{=} \frac{1}{(2k+1)u_{2k}} \sim \frac{1}{2}\sqrt{\frac{\pi}{k}}.$$

Since $u^+_{2k,2} = (2k + 1)^{-1}$ by (1), the asymptotic behaviour of $r_+(2k, 2) + 1$ is the same as the reciprocal of (14) times $(2k + 1)^{-1}$.

Moreover, [1, Equation (7)] tells us the asymptotic formula of $r_-(2nx, 2n)$ for $x \in (0, 1)$ as follows.

$$r_-(2nx, 2n) = \frac{1 - \sqrt{1 - x}}{\sqrt{1 - x}} - \frac{x}{8n(1 - x)^{3/2}} + O(n^{-2}).$$

Besides this, we remark for $x > 0$

$$r_+(2nx, 2n) = -\frac{\sqrt{1 + x} - 1}{\sqrt{1 + x}} - \frac{x}{8n(1 + x)^{3/2}} + O(n^{-2}), \tag{15}$$

the proof of which is the same as [1], but we briefly confirm (15). For the equation

$$r_+(2k, 2n) - 1 = \frac{u^+_{2k,2n}}{u_{2k}} \overset{(13)}{=} \frac{\Gamma\left(n + \frac{1}{2}\right)}{\Gamma(n)} \frac{\Gamma(n + k)}{\Gamma\left(n + k + \frac{1}{2}\right)},$$

it follows that for $k = nx$

$$\begin{cases} \frac{\Gamma\left(n+\frac{1}{2}\right)}{\Gamma(n)} & \sim n^{\frac{1}{2}} \left(1 - \frac{1}{8n} + \frac{1}{128n^2} + O(n^{-3})\right), \\ \frac{\Gamma(n(1+x))}{\Gamma\left(n(1+x)+\frac{1}{2}\right)} & \sim \{n(1 + x)\}^{-\frac{1}{2}} \left\{1 + \frac{1}{8n(1+x)} + \frac{1}{128n^2(1+x)^2} + O(n^{-3})\right\}, \end{cases} \tag{16}$$

because

$$z^{b-a}\frac{\Gamma(z + a)}{\Gamma(z + b)} \sim 1+\frac{(a - b)(a + b - 1)}{2z}+\binom{a - b}{2} \left\{3(a + b - 1)^2 - a + b - 1\right\} \frac{1}{12z^2}+\cdots ,$$

where $\Gamma(\cdot)$ is the standard gamma function. Multiplying the two equations in (16) yields (15).

5. Concluding remarks

This Article examined simple probabilistic problems by using the Pólya urns, and interrupted them through random walks. Giving further interpretations, we conclude this article.

The value $u_{2k}u_{2n-2k}$ of the numerator of (6) is considered as the probability distribution that in the time interval from 0 to $2n$ a random walk spends $2k$ time units on the positive side and $2n-2k$ time units on the negative side (see Feller [2, p. 82, Theorem 2]), which is called *the discrete arcsine distribution of order* n (See [2, p. 79]). Hence $u^-_{2k,2n}$ is regarded as the discrete arcsine distribution conditioned by the random walk returning to 0 with $2n$ steps. A similar interpretation is also possible for $u^+_{2k,2n}$.

Acknowledgements

The author would like to thank the referee and the editor for helpful suggestions that have improved this article. This research was supported by KAKENHI 19K03622 of Japan Society for the Promotion of Science.

References

1. U. Abel, On the relative error between the binomial and the hypergeometric distribution, *Math. Gaz.*, **104** Issue 559, (2020) pp. 136-142.

2. W. Feller, *An introduction to probability theory and its applications*, Vol. I, (3rd edn.), Wiley (1968).

3. K. Matsumoto, T. Nakata, Limit theorems for a generalized Feller game, *J. Appl. Prob.*, **50** (2013) pp. 54-63.

4. T. Nakata, Approximations for the Feller games, *Math. Gaz.*, to appear.

5. N. Johnson, S. Kotz, *Urn Models and Their Application: An Approach to Modern Discrete Probability Theory*, Wiley (1977).

6. H. Mahmoud, *Pólya Urn Models*, Chapman and Hall/CRC, (2008).

7. R. Graham, D. Knuth, O. Patashnik, *Concrete Mathematics: A Foundation for Computer Science*, (2nd edn.) Addison-Wesley (1994).

TOSHIO NAKATA
Department of Mathematics,
University of Teacher Education Fukuoka,
Munakata, Fukuoka, 811-4192, Japan
e-mail: *nakata@fukuoka-edu.ac.jp*