ANZIAM J. 53(2011), 122–133
doi:10.1017/S1446181112000089

A MODIFIED PÓLYA URN PROCESS AND AN INDEX FOR SPATIAL DISTRIBUTIONS WITH VOLUME EXCLUSION

BENJAMIN J. BINDER ✉ [1], EMILY J. HACKETT-JONES[2], JONATHAN TUKE[1] and KERRY A. LANDMAN[2]

(Received 9 February, 2012; revised 4 April, 2012)

Abstract

Spatial data sets can be analysed by counting the number of objects in equally sized bins. The bin counts are related to the Pólya urn process, where coloured balls (for example, white or black) are removed from the urn at random. If there are insufficient white or black balls for the prescribed number of trials, the Pólya urn process becomes untenable. In this case, we modify the Pólya urn process so that it continues to describe the removal of volume within a spatial distribution of objects. We determine when the standard formula for the variance of the standard Pólya distribution gives a good approximation to the true variance. The variance quantifies an index for assessing whether a spatial point data set is at its most randomly distributed state, called the complete spatial randomness (CSR) state. If the bin size is an order of magnitude larger than the size of the objects, then the standard formula for the CSR limit is indicative of when the CSR state has been attained. For the special case when the object size divides the bin size, the standard formula is in fact exact.

2010 *Mathematics subject classification*: primary 62H11; secondary 47N30.
Keywords and phrases: pólya urn, spatial statistics, volume exclusion.

1. Introduction

Analysing the spatial distribution of a set of objects, for example, Lagrangian fluid particles [1, 12, 16, 19] or cellular automata agents [2, 4, 9, 14, 21], has received much attention in the past [7, 12, 16, 19, 20]. In particular, Phelps and Tucker [19] derived an index which quantified the deviation of a spatial point data set from uniformity, based on a scaled variance between bin counts. In their formulation of the index, a domain was partitioned into equally sized bins, while objects were idealized as points with no volume (for example, Lagrangian fluid particles). Limiting values for the index were

[1]School of Mathematical Sciences, The University of Adelaide, South Australia 5005, Australia;
e-mail: benjamin.binder@adelaide.edu.au, simon.tuke@adelaide.edu.au.
[2]Department of Mathematics and Statistics, The University of Melbourne, Victoria 3010, Australia;
e-mail: kerryl@unimelb.edu.au, emilyhackettjones@gmail.com.

obtained when the objects (or points) were distributed uniformly at random throughout a domain, termed complete spatial randomness (CSR) [7, 20].

More recently, Binder and Landman [3] derived CSR limits for the index [19] when the objects exclude volume within a domain (for example, cellular automata agents). In the derivation of the CSR limiting values, the bin counts followed a Pólya–Eggenberger (Pólya) distribution [8]. The CSR limiting values are based on the formulae for the moments of the standard Pólya distribution. This work relied on an analogy between removal of volume by spatial objects and the Pólya urn process. Here we are interested in cases when the Pólya urn process breaks down.

When formulated in 1923, the Pólya distribution was expressed in terms of random drawings of coloured balls from an urn [8]. Initially there are S white balls and $A - S$ black balls. At each trial, one ball is drawn at random and then replaced along with s balls of identical colour. This is repeated n times. If the rules of the drawings can be followed at every trial for all possible random combinations, then the Pólya urn process is called tenable [15, 17]. The process is tenable for all cases where $s \geq -1$.

We are interested here in the cases with $s < -1$, that is, when two or more balls are removed at each trial. If the number of white balls S and the number of black balls $A - S$ are divisible by $|s|$, then the Pólya urn process is actually tenable, although after a certain number of trials the distribution becomes degenerate. For example, if there are initially not enough white balls in the urn for a given number of trials, the probability of drawing a white ball is zero, and a black ball unity, or vice versa. If the number of white or black balls is not divisible by $|s|$, then the Pólya urn process is untenable, because it is possible to have a situation where there are white or black balls in the urn, but there are less than $|s|$ of them. Furthermore, the corresponding standard Pólya distribution [8] does not give a valid probability distribution. This situation occurs when we apply the analogy with removal of volume within a spatial distribution of objects with volume $|s|$. Therefore, we wish to extend the rules of the Pólya urn process after it becomes untenable.

There are a number of ways to do this. Assuming that a white ball is drawn when there are less than $|s|$ of them, either: (i) the remaining number are removed, leaving only black balls; or (ii) the white ball is replaced and this drawing is not included as a trial; in this case, only trials where black balls are drawn are counted as valid trials. Of course there are other possible rules. Here, the second case is the appropriate one in order to continue the analogy with spatial placement of objects.

We determine a probability distribution for this modified Pólya urn process. The variance of this distribution can be calculated numerically. In principle, the CSR limiting values for the index can also be calculated numerically. However, this significantly reduces the usability of the index, and its corresponding standard CSR limiting values [3], as a simple tool for spatial point data analysis. Fortunately, for typical parameter values of interest, we find that the variance given by the formulae for the standard Pólya distribution is a good approximation to the variance of the modified Pólya urn process. This then permits the use of the standard formulae for the CSR limiting values [3], which are indicative of when the CSR state has been attained.

2. The untenable Pólya urn model

We now consider the standard and modified distributions that describe the Pólya urn process.

2.1. The standard Pólya distribution If there are n trials of the Pólya urn process, and K represents the total number of times a white ball is drawn, then the distribution of K is the standard Pólya distribution [8, 10, 11, 15, 17] with parameters n, S, $A-S$ and s:

$$\begin{aligned}\mathrm{P_s}(K=k,n) &= \binom{n}{k}\frac{S(S+s)\cdots(S+(k-1)s)}{A(A+s)\cdots(A+(k-1)s)}\\ &\quad\times\frac{(A-S)(A-S+s)\cdots(A-S+(n-k-1)s)}{(A+ks)(A+(k+1)s)\cdots(A+(n-1)s)}.\end{aligned} \tag{2.1}$$

Equation (2.1) can be written more concisely as

$$\mathrm{P_s}(K=k,n) = \binom{n}{k} f(k,n-k),$$

where the bivariate function f is defined as

$$f(r,t) = \frac{\prod_{i=0}^{r-1}(P+i\alpha)\prod_{i=0}^{t-1}(Q+i\alpha)}{\prod_{i=0}^{r+t-1}(1+i\alpha)}, \tag{2.2}$$

with

$$P=\frac{S}{A},\quad Q=1-P,\quad \alpha=\frac{s}{A},$$

and $\prod_{i=0}^{-1}=1$. The mean and variance of the standard Pólya distribution are

$$\mu_{\mathrm{s}} = nP,\quad \sigma_{\mathrm{s}}^2 = nPQ\frac{1+n\alpha}{1+\alpha}. \tag{2.3}$$

It is instructive to illustrate the probability tree for $n=2$ trials (Figure 1). Let the probability of success be p at each trial, with the corresponding probability of failure being $q=1-p$. At the first trial $p=S/A$, but subsequent p (and q) values depend on the outcomes of previous trials. Therefore, the p and q values are contagious. Crucially, we observe that the probability of a success followed by a failure is the same as the probability of a failure followed by a success, even though the individual p and q values are not the same in each sequence. This accounts for the form of the probability mass function (PMF) in (2.1): first consider all k successes followed by $n-k$ failures and then multiply by the binomial coefficient to account for all possible combinations of sequences of successes and failures.

When $s\ge 0$, the total number of balls in the urn either increases ($s>0$) or remains a constant value A ($s=0$) with each ball drawn. In this case, (2.1) always gives a valid distribution, where all probabilities are positive and sum to unity. For the case $s=-1$,

https://doi.org/10.1017/S1446181112000089 Published online by Cambridge University Press

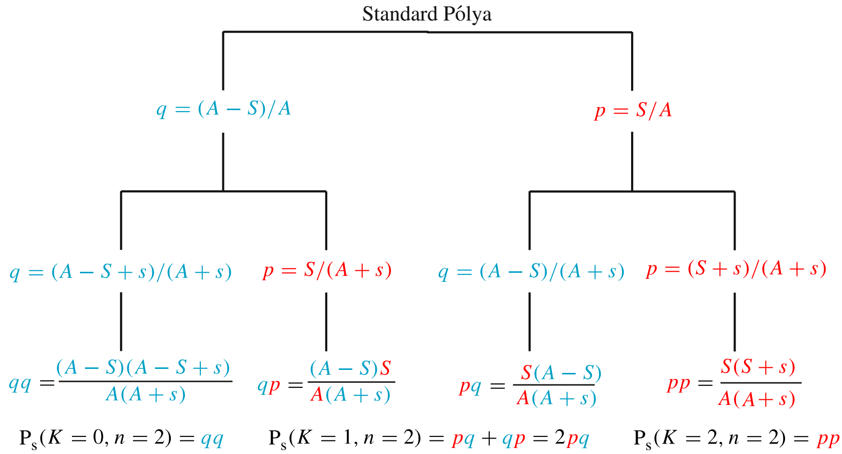

FIGURE 1. Probability tree for the standard Pólya distribution, $P_s(K = k, 2)$.

the urn is sampled without replacement, so the number of balls in the urn decreases by one at each trial. This is the standard example of the classical hypergeometric distribution [10, 11].

Here we are interested not only in the case when $s = -1$, but also in the more general case of $s < 0$, where the total number of balls in the urn decreases by $|s|$ with each trial. For the urn problem to remain tenable when $s < 0$, there are restrictions on the number of trials n, for given values of the total number of balls A, number of white balls S and number of black balls $A - S$ that are initially in the urn. In the urn problem, as posed by Pólya, there is a nonzero probability that only balls of one colour (either white or black) can be drawn. These restrictions impose the following inequalities:

$$A + s(n-1) > 0, \tag{2.4}$$

$$S + s(n-1) \geq 0, \tag{2.5}$$

$$A - S + s(n-1) \geq 0. \tag{2.6}$$

These inequalities ensure that each probability in every trial is nonzero and (2.1) is a valid distribution. Various authors have investigated the conditions when (2.1) remains a valid distribution [5, 6, 10, 11, 13, 18]. However, we are interested in cases when (2.1) defines probabilities that are no longer positive. This occurs when the inequality (2.4) holds, but (2.5) or (2.6) (or both) do not hold.

It is useful to first consider an example ($S = 3$, $A = 9$, $s = -2$ and $n = 3$) and its corresponding probability tree (Figure 2(a)). Since $S + s(n-1) < 0$, the probabilities at each trial become invalid when the (nonzero) numerator in any p value is less than $|s|$ but greater than zero. After that stage, there would be a negative number of white balls in the urn. Interestingly, even when the PMF (2.1) is an invalid distribution and should not be treated as a distribution, the sum of all probabilities equals one and the variance formula (2.3) holds (provided $A + sn > 0$ in the expression for the variance).

https://doi.org/10.1017/S1446181112000089 Published online by Cambridge University Press

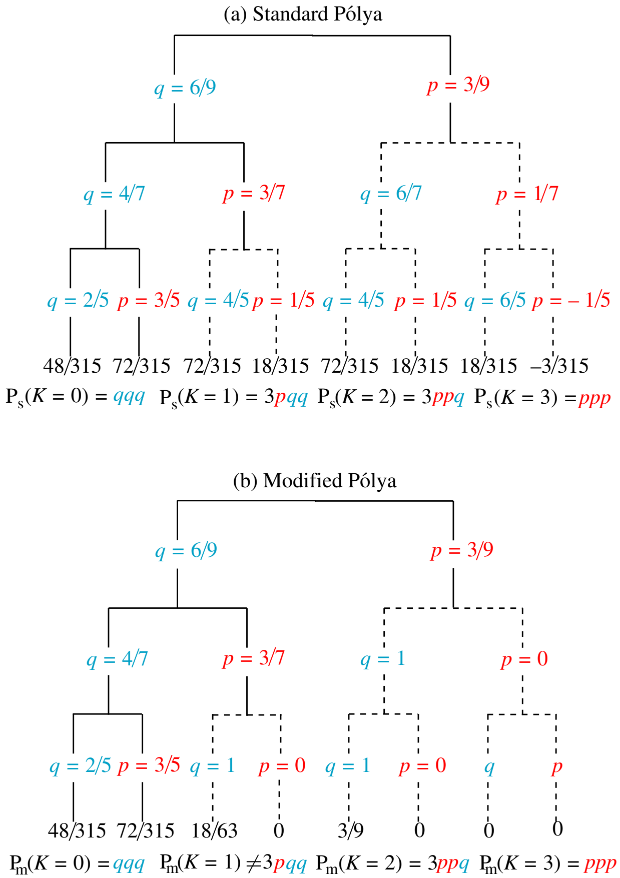

Figure 2. Probability trees with $P = 3/9$, $Q = 6/9$, $s = -2$ and $n = 3$. (a) Standard Pólya distribution, $\mathrm{P_s}(K = k, 3)$. The distribution is invalid when the numerator of any p or q value is less than $|s|$, indicated by the dashed lines. (b) Modified Pólya distribution, $\mathrm{P_m}(K = k, 3)$. The combinatorial advantage in this distribution is lost: for example, $pqq \neq qpq \neq qqp$. The dashed lines indicate when the distribution becomes degenerate.

How can the PMF (2.1) be corrected or modified to give a valid distribution when there are insufficient white or black balls? Moreover, how do the moments of such a modified PMF compare to the moments of the Pólya distribution (2.3)? We tackle this by defining a modified distribution which reflects the physical problem of consecutive draws of balls of the same colour.

2.2. The modified Pólya distribution The Pólya urn process can be extended by adding a new rule when the drawing becomes untenable. By means of example, we consider the situation when there are not enough white balls in the urn. A similar description holds when there are insufficient black balls in the urn.

https://doi.org/10.1017/S1446181112000089 Published online by Cambridge University Press

Assuming a white ball is drawn when there are less than $|s|$ of them, we replace the white ball and do not include this drawing as a trial. Only trials where black balls are drawn are counted as valid trials. A probability tree given by this extended process for our example is illustrated in Figure 2(b). We observe that the ordering of the p and q values in a given sequence is now important, so that the combinatorial advantage is lost, and therefore all possible permutations have to be considered. The formulation of this PMF gives rise to a modified Pólya distribution:

$$\mathrm{P_m}(K=k,n) \\ = \begin{cases} 0 & \text{if } k > \lfloor S/|s| \rfloor \text{ or } n-k > \lfloor (A-S)/|s| \rfloor \\ \binom{n}{k} f(k, n-k) & \text{if } k < \lfloor S/|s| \rfloor \text{ and } n-k < \lfloor (A-S)/|s| \rfloor \\ \sum_{i=0}^{n-k} \binom{i+k-1}{k-1} f(k,i) & \text{if } k = \lfloor S/|s| \rfloor \text{ and } n-k < \lfloor (A-S)/|s| \rfloor \\ \sum_{i=0}^{k} \binom{i+n-k-1}{n-k-1} f(i,n-k) & \text{if } k < \lfloor S/|s| \rfloor \text{ and } n-k = \lfloor (A-S)/|s| \rfloor \\ 1 & \text{if } k = \lfloor S/|s| \rfloor \text{ and } n-k = \lfloor (A-S)/|s| \rfloor. \end{cases} \tag{2.7}$$

The mean and variance are

$$\mu_{\mathrm{m}} = \sum_{k=0}^{n} k\mathrm{P_m}(K=k,n), \qquad \sigma_{\mathrm{m}}^2 = \sum_{k=0}^{n} (k-\mu_{\mathrm{m}})^2 \mathrm{P_m}(K=k,n). \tag{2.8}$$

As expected, the modified Pólya PMF (2.7) is significantly more complicated than the previous PMF (2.1). We describe each of the five cases (from top to bottom) in (2.7) as follows. For a given number of trials n there are: (i) an insufficient number of either white or black balls in the urn; (ii) a sufficient number of white and black balls in the urn (the standard Pólya distribution); (iii) a sufficient number of black balls in the urn, with the distribution becoming degenerate at the kth success due to an insufficient number of white balls; (iv) a sufficient number of white balls in the urn, with the distribution becoming degenerate at the kth success due to an insufficient number of black balls; (v) precisely enough white and black balls in the urn, with only one possible nonzero outcome.

The PMF (2.7) is a valid distribution if each individual outcome is nonnegative (namely $\mathrm{P_m}(K=k,n) \geq 0$) and the sum of all possible outcomes is unity (namely $\sum_{k=1}^{n} \mathrm{P_m}(K=k,n) = 1$). Both of these conditions are satisfied by the way in which we constructed the distribution. The first condition is obviously true as the bivariate function (2.2) is now not permitted to take negative values. It is less obvious that the sum of all possible outcomes is unity, but we did verify this using the statistical software program R. A formal proof of this second condition relies on proving (by induction) a more general result associated with the tree structure of the distribution; namely, the sum of the product of the probabilities along all paths in a (binary) tree is unity, provided the probabilities at each (nondegenerate) node sum to unity.

So far we have established that the PMF (2.1) gives an invalid distribution when the (nonzero) numerator in any p or q value is less than $|s|$. However, there is a special case when the number of white and the number of black balls initially in the urn are both divisible by s (namely $S \equiv 0 \bmod |s|$ and $A - S \equiv 0 \bmod |s|$). Given sufficient trials n, the distribution will eventually become degenerate. However, in any sequence of p or q values, the numerators are positive multiples of $|s|$ or zero, as seen in (2.1). Hence, in this case the PMF (2.1) will give probabilities that are either zero or positive and that sum to unity. In this special case, the two distributions (2.1) and (2.7) are in fact identical, and their moments must, of course, be given by the standard formula (2.3).

2.3. Comparisons We wish to compare the variances of the two distributions (2.1) and (2.7). To check our results we also simulate the Pólya urn. In Figures 3 and 4 we present results where the total number of balls initially in the urn is an integer multiple M of the number of white balls S initially in the urn. The corresponding number of black balls is then $(M-1)S$. The findings are characterized using arithmetic modulo s.

As a way of checking the formulation (2.7), we first considered the special case when the initial numbers of white and black balls are both divisible by s (namely $S \equiv 0 \bmod |s|$ and $A - S \equiv 0 \bmod |s|$). As expected, the standard formula (2.3) accurately predicts the simulation results (not shown here). Furthermore, the corresponding variance (2.8), which is obtained numerically, is identical to the standard variance (2.3).

We now consider a more interesting situation, when the initial numbers of white and black balls are no longer congruent to 0 mod $|s|$ (Figure 3(a)). The results confirm that (2.8) accurately predicts the results from the Pólya urn simulations. However, the standard formula for the variance (2.3) is not a good approximation to the modified Pólya variance (2.8). The noticeable differences between the two formulae for the variance occur because the remainder ($S \equiv 24 \bmod 26$) is comparable in size to the number of white balls initially in the urn ($S = 50$). These differences are almost nonexistent when we examine the results (Figure 3(c)) with a much reduced remainder ($S \equiv 2 \bmod 26$), an order of magnitude smaller than the number of white balls initially in the urn ($S = 54$). These results show that (for a given value of s) the error in the approximations of the modified Pólya variance tends to zero (when $S = 52$ with $S \equiv 0 \bmod 26$) in a nonuniform way. Note that even though the number of black balls initially in the urn ($A - S = (M-1)S$) is not congruent to 0 mod 26 (in either of Figures 3(a) or (c)), the remainders are at least an order of magnitude smaller than $(M-1)S$. Hence, it is the congruency of S modulo s that dominates the error or differences observed in Figure 3.

Next we examine the case when the initial number of white balls in the urn increases, for a fixed nonzero value of the remainder. Figure 4(a) shows that the

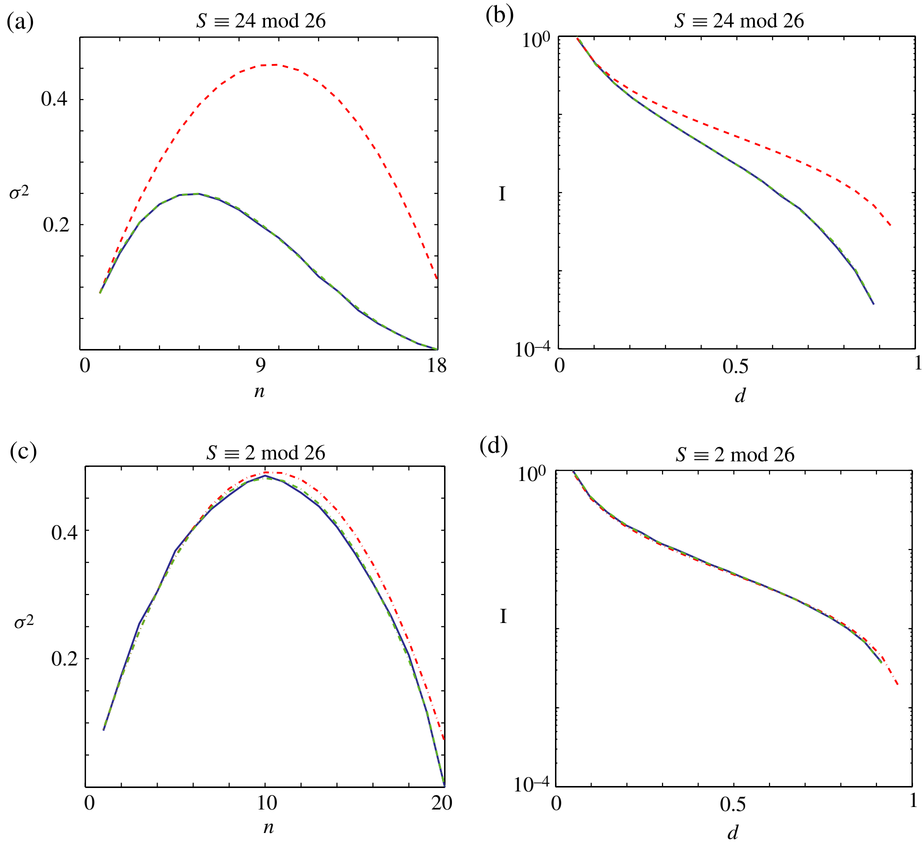

FIGURE 3. Variance and index for two distributions, with $s = -26$, $M = 10$, $A = MS$, $P = 1/10$ and $Q = 9/10$. (a) and (c) Variance versus number of trials: variance from 10 000 simulations of the Pólya urn (blue curves) and from (2.3) (red broken curve) and (2.8) (green broken curve) (colour available online), for (a) $S = 50$ with $S \equiv 24 \bmod 26$, (c) $S = 54$ with $S \equiv 2 \bmod 26$. (b) and (d) Index and CSR limits versus density d, for the same parameter values as in (a) and (c), respectively: index from 10 000 simulations of the Pólya urn (blue curves) and CSR limiting values based on the variance formulae (2.3) (red broken curve) and (2.8) (green broken curve).

approximation to the modified Pólya variance improves as S increases ($S \equiv 4 \bmod 6$). Note that we were careful to choose parameter values so that $A - S = (M - 1)S \equiv 0 \bmod 6$, eliminating any (small) errors in the approximations due to the initial number of black balls.

Let us summarize our findings. (i) Both PMFs related to the Pólya urn process are identical when $s < 0$ and $S \equiv A - S \equiv 0 \bmod |s|$, even though the distributions may become degenerate. In this case, the Pólya urn process is tenable. (ii) When either $S \not\equiv 0 \bmod |s|$ or $A - S \not\equiv 0 \bmod |s|$, and when there are initially insufficient white or black balls for a given number of trials, the Pólya urn process is untenable and the standard PMF (2.1) is an invalid distribution. (iii) We have determined a valid

https://doi.org/10.1017/S1446181112000089 Published online by Cambridge University Press

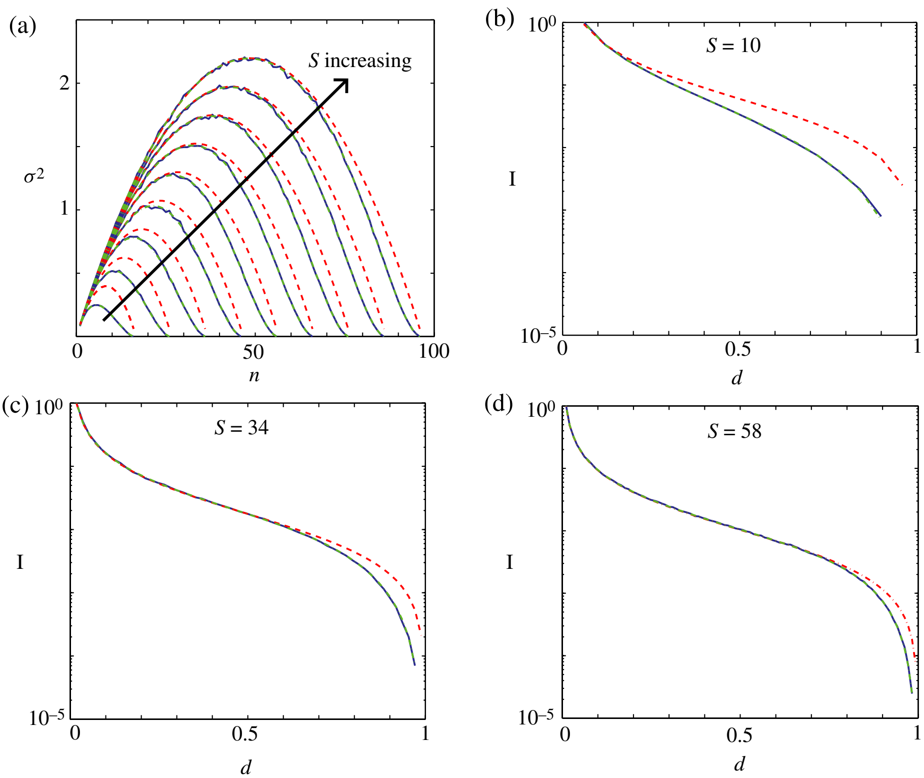

FIGURE 4. Variance and index for three distributions, with $s = -6$, $M = 10$, $A = MS$, $P = 1/10$ and $Q = 9/10$, with varying values of S with $S \equiv 4 \bmod 6$. (a) Variance versus number of trials: variance from 10 000 simulations of the Pólya urn (blue curves) and from (2.3) (red broken curve) and (2.8) (green broken curve) (colour available online). The groupings of the three curves (blue, red and green) are for increasing values of $S = 10, 16, 22, 28, 34, 40, 46, 52, 58$. (b)–(d) Index and CSR limits versus density d: index from 10 000 simulations of the Pólya urn (blue curves) and CSR limiting values based on the variance formulae (2.3) (red broken curve) and (2.8) (green broken curve), for (b) $S = 10$, (c) $S = 34$, (d) $S = 58$.

modified Pólya distribution (2.7). (iv) In situations when the PMF (2.1) gives an invalid distribution, the standard formula for the variance (2.3) is a good approximation to the modified Pólya variance (2.8), provided $|s| \ll S$ and $|s| \ll A - S$. These results are pertinent for determining the CSR limits for the index (based on the variance) used to characterize spatial distributions of objects which exclude volume.

3. Index and CSR limit

Consider a domain of volume A that is populated with a total of n objects each of volume $\hat{s}$. The domain is divided into M equally sized bins each with volume S. Following Phelps and Tucker [19], we quantify the deviation between each bin

https://doi.org/10.1017/S1446181112000089 Published online by Cambridge University Press

count b_j, for $j = 1, \ldots, M$, and the uniformly distributed state by

$$\sigma^2 = \frac{1}{M} \sum_{j=1}^{M} \left(b_j - \frac{n}{M} \right)^2, \tag{3.1}$$

where n/M is the average bin count. The variance is scaled by

$$\sigma_0^2 = n^2 \left(\frac{M-1}{M^2} \right),$$

and defines an index

$$I = \frac{\sigma^2}{\sigma_0^2}. \tag{3.2}$$

If $\sigma^2 = 0$, the objects are said to be evenly distributed throughout the domain, referred to as an evenly distributed state, giving $I = 0$. At the other extreme, if all the objects are placed in a single bin, then $I = 1$. The index (3.2) is therefore a quantitative measure of the spatial distribution of a set of objects from the evenly distributed state. A more interesting and likely scenario to consider is when the objects are distributed uniformly at random throughout the domain. This is known as the CSR state [3, 7, 12, 19, 20].

The CSR limit for the index is obtained by considering the random placement of n objects each of size $\hat{s} > 0$ throughout a domain. For each trial or object placement, the probability of the object being placed in a bin is then equal to the unoccupied space within the bin divided by the total amount of unoccupied space in the domain. The probability of success for a given trial depends on the outcomes of the previous object placements, in a similar way to that of the Pólya distribution. Following Binder and Landman [3], there is a direct analogy between removing balls from a Pólya urn and placing finite-sized objects within a bin. At each trial, instead of removing balls with $s < 0$ from an urn, we are removing volume $s = -\hat{s}$ from the domain. The number of white balls S initially in the urn corresponds to the volume of just one bin. The number of black balls initially in the urn, $A - S = S(M-1)$, then represents the remaining volume in the domain. We assume that each bin count b_j can be represented by a random variable that follows the modified Pólya distribution (2.7) with parameters n, S, $A - S = S(M-1)$ and $s = -\hat{s}$. Taking the expectation of (3.1), we find that the expected value for the variance of the bin counts is given by (2.8).

As discussed, the Pólya distribution may become untenable, but when $\hat{s} \ll S < A - S$, the variance is well approximated by (2.3). Therefore, we approximate the expected value for the variance of the bin counts by

$$\sigma_{\text{csr}}^2 = nPQ \frac{1 + n\alpha}{1 + \alpha} \approx \frac{n}{M} \left(1 - \frac{1}{M} \right) \left(1 - \frac{n\hat{s}}{A} \right). \tag{3.3}$$

Scaling (3.3) by σ_0^2 ,we obtain the CSR limit

$$I_{\text{csr}} = \frac{\sigma_{\text{csr}}^2}{\sigma_0^2} = \frac{1-d}{n} \quad \text{where } d = \frac{n\hat{s}}{A}. \tag{3.4}$$

https://doi.org/10.1017/S1446181112000089 Published online by Cambridge University Press

Here d is the density or proportion of volume in the domain that is occupied by objects.

4. Discussion

To validate the CSR limit for the index, we simulate the Pólya urn and calculate the average index $\langle I \rangle$ using (3.2) when $S \not\equiv 0 \bmod \hat{s}$ and $A - S \not\equiv 0 \bmod \hat{s}$ (Figures 3(b) and (d); 4(b)–(d)). We compare this average index to the two CSR limiting values based on the two variances (2.3) and (2.8). As we already know, the formula (3.4) is based on the variance (2.3). The other CSR limiting value is obtained numerically, by substituting the calculated values of the variance $\sigma^2_{\mathrm{csr}} = \sigma^2_{\mathrm{m}}$ into (3.4). The results demonstrate that the explicit formula for the CSR limit (red) is a good approximation to the numerically obtained CSR limit, when $\hat{s} \ll S$ and $\hat{s} \ll A - S$.

In conclusion, if the bin size is an order of magnitude larger than the size of the objects, the explicit formula for the CSR limit should be indicative of when the CSR state has been attained. This easy-to-implement measure should prove useful in assessing the spatial distribution of objects that arise in many physical applications (examples are given by Binder and Landman [3]).

Acknowledgements

This work was supported by an Australian Research Council Discovery Grant (Kerry Landman). Kerry Landman is an ARC Professorial Fellow.

References

[1] B. J. Binder, "Ghost rods adopting the role of withdrawn baffles in batch mixer designs", *Phys. Lett. A* **374** (2010) 3483–3486; doi:10.1016/j.physleta.2010.06.039.

[2] B. J. Binder and K. A. Landman, "Exclusion processes on a growing domain", *J. Theor. Biol.* **259** (2009) 541–551; doi:10.1016/j.jtbi.2009.04.025.

[3] B. J. Binder and K. A. Landman, "Quantifying evenly distributed states in exclusion and non-exclusion processes", *Phys. Rev. E.* **83** (2011) 041914; doi:10.1103/PhysRevE.83.041914.

[4] C. A. Chung, T.-H. Lin, S.-D. Chen and H.-I. Huang, "Hybrid cellular automaton modeling of nutrient modulated cell growth in tissue engineering constructs", *J. Theoret. Biol.* **262** (2010) 267–278; doi:10.1016/j.jtbi.2009.09.031.

[5] O. L. Davies, "On asymptotic formulae for the hypergeometric series: I. Hypergeometric series in which the fourth element, *x*, is unity", *Biometrika* **25** (1933) 295–322; doi:10.1093/biomet/25.3-4.295.

[6] O. L. Davies, "On asympotic formulae for the hypergeometric series: II. Hypergeometric series in which the fourth element, *x*, is not necessarily unity", *Biometrika* **26** (1934) 59–107; doi:10.1093/biomet/26.1-2.59.

[7] P. J. Diggle, *Statistical analysis of spatial point patterns* (Academic Press, London, 1983).

[8] F. Eggenberger and G. Pólya, "Über die Statistik verketteter Vorgänge", *Z. angew. Math. Mech.* **3** (1923) 279–289; doi:10.1002/zamm.19230030407.

[9] H. Enderling, L. Hlatky and P. Hahnfeldt, "Migration rules: Tumours are conglomerates of self-metastases", *Brit. J. Cancer* **100** (2009) 1917–1925; doi:10.1038/sj.bjc.6605071.

[10] N. L. Johnson, A. W. Kemp and S. Kotz, *Univariate discrete distributions*, 3rd edn (John Wiley & Sons, Hoboken, NJ, 2005).

[11] N. L. Johnson and S. Kotz, *Urn models and their applications* (John Wiley & Sons, New York, 1977).

[12] S. W. Jones, "The enhancement of mixing by chaotic advection", *Phys. Fluids A.* **3** (1991) 1081–1086; doi:10.1063/1.858089.

[13] C. D. Kemp and A. W. Kemp, "Generalized hypergeometric distributions", *J. R. Stat. Soc. B* **18** (1956) 202–211; http://www.jstor.org/stable/2983704.

[14] E. Khain, L. M. Sander and C. M. Schneider-Mizell, "The role of cell–cell adhesion in wound healing", *J. Stat. Phys.* **128** (2007) 209–218; doi:10.1007/s10955-006-9194-8.

[15] S. Kholfi and H. M. Mahmoud, "The class of tenable zero-balanced Pólya urns with an initially dominant subset of colors", *Stat. Prob. Lett.* **82** (2012) 49–57; doi:10.1016/j.spl.2011.08.006.

[16] T. Li and I. Manas-Zloczower, "A study of distributive mixing in counterrotating twin screw extruders", *Int. Polym. Process.* **10** (1995) 314–320; http://www.polymer-process.com/IPP950314.

[17] H. M. Mahmoud, *Pólya urn models* (Chapman & Hall/CRC, Boca Raton, FL, 2008).

[18] A. Noack, "A class of random variables with discrete distributions", *Ann. Math. Stat.* **21** (1950) 127–132; doi:10.1214/aoms/1177729894.

[19] J. H. Phelps and C. L. Tucker, "Lagrangian particle calculations of distributive mixing: Limitations and applications", *Chem. Eng. Sci.* **61** (2006) 6826–6836; doi:10.1016/j.ces.2006.07.008.

[20] B. D. Ripley, *Spatial statistics* (John Wiley & Sons, New York, 1981).

[21] D. Zhang, I. M. Brinas, B. J. Binder, K. A. Landman and D. F. Newgreen, "Neural crest regionalisation for enteric nervous system formation: implications for Hirschsprung's disease and stem cell therapy", *Dev. Biol.* **339** (2010) 280–294; doi:10.1016/j.ydbio.2009.12.014.