Introducing perception and modelling of spatial randomness in classroom

José Renato De Nóbrega
Department of Biostatistics, School of Biology, Faculty of Science, Central University of Venezuela, Caracas, Venezuela
e-mail: renato.nobrega@ciens.ucv.ve

Summary A strategy to facilitate understanding of spatial randomness is described, using student activities developed in sequence: looking at spatial patterns, simulating approximate spatial randomness using a grid of equally-likely squares, using binomial probabilities for approximations and predictions and then comparing with given Poisson probabilities. Key questions are discussed with students on concepts needed to understand the approximate models and to generate predictions based on the approximations. The sequence is structured to be interactive to encourage student's interest and curiosity.

Keywords: Teaching statistics; spatial randomness; binomial; Poisson.

INTRODUCTION

I discuss an exercise designed to introduce students to the study of spatial randomness through activities developed in sequence. In the first, students examine three dispersion patterns and select the one they think is random and comment on features of randomness. These comments lead to discussion of characteristics of spatial randomness and the essential conditions of a model to simulate a random spatial pattern. Probability models are then used to simulate approximations to spatial randomness and to explore predictions and properties. This leads to the binomial for the approximate model. The activity is completed by comparing binomial with Poisson probabilities matched through means. The exercise requires student knowledge about the binomial and Poisson models. I have used the strategy with undergraduate biology students 18 years old.

PERCEPTION OF SPATIAL RANDOMNESS

Students are invited to inspect graphs representing dispersion patterns of 90 points in a closed area and to expose their perception about whether the points are distributed randomly or not. I just show three obvious and contrived patterns representing regular, clumped and random dispersion, with the intention of facilitating the choice for students so that they can focus on characteristics of randomness (Figure 1). In the realized experiences, the students choose correctly the random pattern, and the principal reason given to support a belief in the randomness is the irregularity of the pattern.

SIMULATING SPATIAL RANDOMNESS

In this activity, the minimum conditions required for a simple model representing random dispersion of 90 points in the area are discussed. The lack of structure in the arrangement of the points in the random pattern is conducive to a model with at least two conditions: each small equal-sized portion of the area should have an equal chance of receiving any point, and its position should not affect or be affected by the positions of the other points. Here, it is possible to reach an agreement with students in that an approximate model of random dispersion should represent the area as a grid of equally- sized and equally-likely squares, with each point allocated to a square unaffected by all other allocations.

For the simulation of random dispersion, each student is provided with a sheet in which the total area is subdivided into 100 squares of equal size (the grid), each one with coordinates indicating the row and column of its location (Figure 2), and 90 trials are conducted. In each trial, two

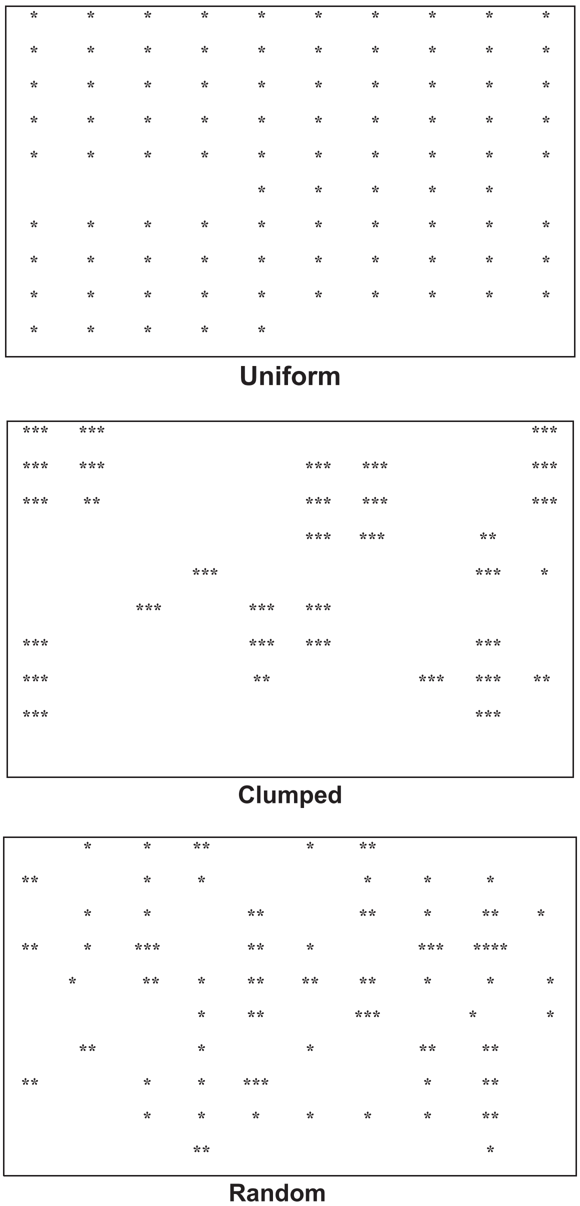

Fig. 1. The three patterns of dispersion

	0	1	2	3	4	5	6	7	8	9
0										
1										
2										
3										
4										
5										
6										
7										
8										
9										

Fig. 2. Area subdivided into 100 squares, each with coordinates indicating the row and column of its location

random digits from 0 to 9 are generated representing the coordinates of the selected square, and a point is marked in that square.

QUESTIONS, OBJECTIVES AND RESPONSES

The following questions help the students to understand the simulation and then to see the model being used for the approximation.

Question 1: For each of the 90 points, what is the probability that a particular square will be chosen?

The answer should not present any difficulty for students. In a population of 100 squares, with each having an equal probability of being selected in a trial, this probability is $1/100 = 0.01$. Students agree that this is a small probability. It is important that the teacher stresses this fact: the probability of 'the appearance of a point in any particular square' is low.

Question 2: How many points could a particular square contain once all 90 points have been randomly entered into the test area?

The aim of this question is to help the students recognize that we can consider a random variable defined as 'the number of points that will be placed in a square' and differentiate between the theoretically possible values for this variable and the likely values. In experiments I have conducted, students usually indicate low values (zero, one, two and up to three points) for the possible points in a particular square, but they hesitate to indicate higher numbers because they consider that these are 'not possible'. In these cases, I clarify that the question is requesting the theoretically possible values, highlighting the difference between possible values and the probabilities associated with these values.

Once this clarification is made, the following questions could be raised: is it possible that all 90 points will be positioned in a single square? and what is the chance of this happening? Given the previous explanation, students in general recognize that this is a theoretically possible event but with a negligible probability that is practically zero; in my personal experience, students demonstrate difficulty in determining the technical procedure for calculating this probability.

However, in finalizing the discussion, there is an agreement that the number of points that may be added to a single square is a random variable that has a value from zero (0) to ninety (90). This realization presents a good opportunity to introduce the definition of independent events and to calculate the required probability. Because the probability that a particular square is chosen in a selection trial is 0.01 and because this probability does not change for each of the 90 trials, the probability that the same square is chosen randomly in all 90 independent trials is obtained by raising 0.01 to the 90th power: $(0.01)^{90}$.

Question 3: How many squares will be empty once all 90 points have been entered into the area?

This question is posed at the beginning of the simulation exercise to provoke curiosity and interest. The students usually give a range of possible values. There is no problem with respect to the lower bound: given that there are 90 points to distribute among 100 squares, the number of empty squares cannot be less than 10. With respect to the upper bound, students usually show certain prudence, and the maximum value more frequently expressed in my experience has been 30 empty squares. At this point, I intervene with the prediction that for the simulation exercise, 'the average number of empty squares will be 40'. In general, this statement surprises students because of the discrepancy between this and the predictions made by them. I then state that the simulation will provide the opportunity to test the model prediction. The simulations are conducted, and each student must calculate the frequency in his/her simulation of squares with zero, one, two and three or more points; the students find that the average number of empty squares over all their simulations does not disagree significantly with the predicted value (40). We then proceed to discuss a probabilistic model as approximation to this process.

THE BINOMIAL MODEL

Given that the students know the binomial model, they can recognize this in the simulation. The probability that the variable X, 'number of points in a square', takes value x is given by a binomial probability b $(x;\ n,\ p)$ function, with $n=90$, $p=0.01$ and x: 0, 1, 2…90. At this moment, it is convenient to calculate, with the students, the binomial probability that a square is empty ($p(X=0)=0.4047$) and see that according to this model, the average number of empty squares in a sample of 100 squares is $p(X=0)*100=0.4047*100=40.47$. Students thus discover the rationale of the prediction made by the teacher in the simulation exercise.

INTRODUCING THE POISSON MODEL

The Poisson model is introduced by name as an approximation to the binomial model for a large number of trials ($n=90$) with a low probability of a successful event ($p=0.01$). The students are informed that the Poisson has just one parameter, namely, the average number of points per square, and that there are two ways to calculate this in the exercise: by dividing the total number of points among the total number of squares (90 points/100 squares = 0.9 points per square) or as the mean of a binomial variable (mean $=p*n=0.01*90=0.90$). The calculated binomial probability for $X=0$ is then compared with the corresponding Poisson probability, given via technology (Poisson function in Excel) or a formula according to the statistical level of the course.

Table 1. The binomial Poisson approximation

Variable X	Binomial probability			Poisson probability
	Number of squares: 100 $n=90$ $p=0.01$ Average: 0.90 points per square	Number of squares: 1000 $n=900$ $p=0.001$ Average: 0.90 points per square	Number of squares: 10000 $n=9000$ $p=0.0001$ Average: 0.90 points per square	Parameter: average number of points per square: 0.90
0	0.404732	0.406387	0.406551	0.406570
1	0.367938	0.366114	0.365933	0.365913
2	0.165386	0.164733	0.164668	0.164661
3	0.049003	0.049359	0.049394	0.049398
≥4	0.01294	0.013407	0.013454	0.013459

The overall exercise culminates by considering increasing the number of squares and the number of points by multiplying them by the same factor, thus maintaining the average number of points per square at a constant, and comparing binomial probabilities with the corresponding Poisson probabilities. This could be illustrated using Table 1, noting that the binomial probabilities became close to the Poisson. This final part of the exercise demonstrates the use of the Poisson probability function as an appropriate model for spatial randomness.

After the exercise, I usually recommend that biology students reinforce this knowledge by reading about spatial dispersion in plants and animals in several books such as Ricklefs and Miller (2000) and Starr et al. (2009). A very useful and recommended website is the STatistical Education through Problem Solving (STEPS) module Spatial Pattern of a Plant Population (http://www.stats.gla.ac.uk/steps/abstracts/biology.html#16). This module also illustrates the use of the chi-square goodness-of-fit test, which could be the basis for a further task (statistic inference about a random pattern) to include according the statistical level of the students.

Acknowledgement

I thank the reviewer and editor for their excellent comments.

References

Ricklefs, R.E. and Miller, G.L. (2000). Ecology, San Francisco: Freeman.

Starr, C., Taggart, R., Evers, C. and Starr, L. (2009). Biology. The Unity and Diversity, Belmont: Brooks/Cole, Cengage Learning.

Peter Holmes Prize Announcement

The article by Sarah Bebermeier and Katharina Reiss entitled 'Practicing Statistics by Creating Exercises for Fellow Students' has been awarded the Peter Holmes prize for 2016. The aim of this prize is to highlight excellence in motivating practical classroom activity. When Peter established Teaching Statistics, the core goal was to enable teachers to share best practice. Sarah and Katharina's article reports on a core active learning principle often espoused but less often authentically implemented in teaching statistics, namely, a developmental process of student group creation of assessment questions, incorporating peer feedback, full and constructive instructor feedback, student revision and presentation, further peer and instructor feedback and for the compilation into a class portfolio. The full process was developed for a first year psychology class, but the selected topic of descriptive statistics makes this article accessible for teaching at school level. Although developed for a psychology class, the authors' demonstration of thinking beyond standard approaches is an admirable exemplar for all disciplines to consider in teaching statistics. Moreover, the authors' candid reflections on the challenges, possible improvements in effectiveness and efficiency, and workable adaptations for other situations provide ideas for teachers to apply the principle in contexts beyond those described in the article.