PRACTICAL ACTIVITIES

A 'Striking' Demonstration of the Poisson Distribution

KEYWORDS:
Teaching;
Intellectual excitement;
Long-term learning;
Poisson demonstration.

Vincent Buonaccorsi and Amy Skibiel
Juniata College, Huntingdon, Pennsylvania, USA.
e-mail: buonaccorsi@juniata.edu

Summary
This article describes a simple classroom activity that helps students immediately visualize and understand the meaning and mathematical properties of the Poisson distribution.

◆ INTRODUCTION ◆

Striking interactive demonstrations add excitement to teaching and learning statistics (Sowey 2001), improve statistical intuition (Shaughnessy 1992) and may enhance long-term learning (Sowey 2001). Without illustration or first-hand experience, probability distributions may remain abstract theoretical constructs in an introductory statistics class. To give students experience in using the Poisson distribution, plot-sampling methods for plants or animals may be used to determine whether a distribution is random, clumped or uniform (e.g. Brower et al. 1997). However, such methods are typically time intensive, and may require a laboratory section or specialized equipment. Here we describe a simple interactive demonstration that was successfully used for two years in an undergraduate biostatistics lecture setting. The demonstration immediately and clearly illustrated the Poisson distribution, and a student survey ($n = 7$, self selection) indicated strong agreement that the activity had educational value and should be kept in future years.

A Poisson distribution of split peas per sampling quadrat was generated as a student dropped a handful of split peas onto a grid that was projected on an overhead projector. Data analysis consisted of (1) tallying the observed frequencies of peas per quadrat, (2) calculating the mean number of peas per quadrat, (3) calculating the Poisson estimated probabilities for each outcome, (4) calculating the Poisson expected frequencies, (5) calculating and comparing the mean and variance of the distribution, (6) calculating the coefficient of dispersion (CD) and (7) evaluating the fit of observed and expected frequencies using a chi-square goodness-of-fit test.

◆ CONSTRUCTION OF GRID ◆

A 20 cm × 20 cm grid comprised of one hundred 2 cm^2 quadrats was drawn with a permanent marker on a transparency and taped to a 36 cm by 36 cm cardboard box lid (figure 1). The grid was placed directly over a 20 cm × 20 cm square cutout in the centre of the box lid to allow the grid to be visualized

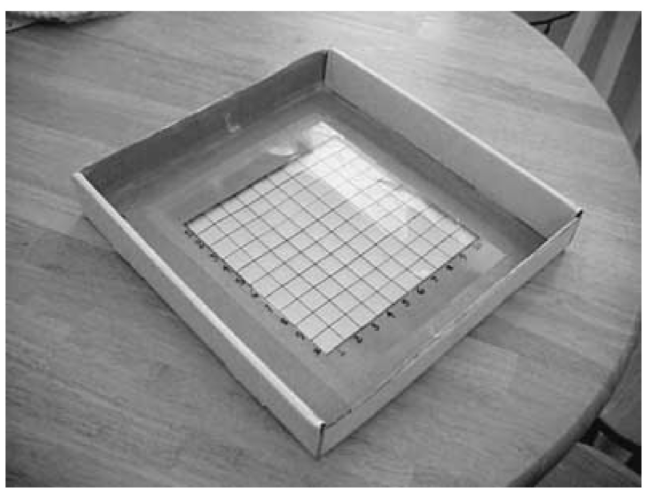

Fig 1.

on a screen using an overhead projector. The grid covered approximately one half of the lid area, allowing enough space between the transparency and lid edges to prevent split peas, dropped from a height of one foot, from hitting lid edges and rebounding onto the grid. The lid edges also prevented peas from falling onto the floor, facilitating clean-up.

◆ PROCEDURE ◆

At the beginning of the semester, each student in the class received a numbered syllabus. In class, a number was drawn using a random numbers table and the student with that number on his/her syllabus was selected. This student dropped 150 pre-counted split peas from a height of one foot above the centre of the grid that had been placed on an overhead projector. Split peas were used in the experiment because they are small relative to quadrat size (maximum possible number of peas/quadrat is about 20) and they do not roll. Peas that landed on the border between quadrats were pushed into the quadrat containing the largest fraction of the pea area. If it was too difficult to determine which quadrat contained the largest fraction of the pea area, the pea was removed from the grid and not counted in the experiment.

◆ RESULTS AND DISCUSSION ◆

Results from the Fall 2002 semester are presented (table 1). Approximately half of the peas landed on the grid. In contrast to what one might expect in a uniform or clumped distribution, most quadrats had no peas, some quadrats had one pea, and a few quadrats had two or three peas (table 1). Estimated Poisson probabilities were calculated for each observation category, using the equation

$$p(x_i) = \frac{\bar{x}^{x_i} e^{-\bar{x}}}{x_i!}$$

where the mean number of peas per quadrat, $\bar{x}$ (0.77), was calculated as the total number of peas observed (77) divided by the number of quadrats sampled (100), and x_i refers to the number of peas per quadrat. To avoid confusion in calculation of the mean, it was emphasized that sample size in this case refers to the total number of units of space or time (i.e. quadrats), rather than the number of peas that fell onto the grid. The probability for the tail of the distribution was calculated as 1 – the sum of the previous probabilities. Poisson frequencies were then obtained by multiplying the sample size (100) by the estimated probabilities.

The Poisson distribution is a probability distribution of the number of times a random event occurs. In this experiment the observed frequencies were close to the probabilities predicted by a Poisson distribution (table 1).

The coefficient of dispersion (variance/mean) can be used to determine whether occurrences of an event are approximately random (CD $\approx$ 1), clumped (CD > 1), or uniform (CD < 1) (Sokal and Rohlf 1995). From a biological viewpoint, schooling organisms (for example skipjack tuna) may exhibit clustered distributions whereas territorial competitive organisms may exhibit uniform or repulsed distributions (for example sea anemones). In this experiment, the sample variance (0.68) was similar to the mean (0.77), yielding a CD of 0.89, and a reasonable fit of observed frequencies of peas per quadrat to a Poisson (random) distribution. However, without a statistical test it is difficult to interpret whether deviations from Poisson expectations are due to chance alone.

To determine whether the deviation of observed frequencies from Poisson expected frequencies was unlikely to be due to chance alone, a chi-square

Peas/quadrat	*Frequency (number of quadrats)*	*Number of peas observed*	*Poisson probabilities*	*Poisson frequency (number of quadrats)*
0	45	0	0.463	46.3
1	36	36	0.357	35.7
2	16	32	0.137	13.7
3	3	9	0.035	3.5
≥4	0	0	0.008	0.8
	Σ = 100	Σ = 77		

Table 1. Observed frequencies of peas per quadrat, total number of peas observed per category, estimated Poisson probabilities and expected Poisson frequencies

Peas/quadrat	*Observed frequency*	*Poisson expected frequency*
0	45	46.3
1	36	35.7
≥2	19	18.0

Table 2. Observed and expected Poisson frequencies for each observation category

goodness-of-fit test was performed. In this experiment, the observation category of ≥4 peas per quadrat was lumped with the 2 and 3 peas per quadrat categories, following the suggestion for chi-square analysis that no expected frequency can be less than five (e.g. Sokal and Rohlf 1995). A chi-square goodness-of-fit test did not detect a significant difference between observed pea frequencies and expected Poisson frequencies, suggesting that the peas followed a Poisson distribution (table 2, X^2 test statistic = 0.095, $p = 0.76$). In other words, the difference between observed and Poisson expected frequencies was small enough to be attributable to chance alone.

The following assessment questions have been used for class discussion and examinations.

1. If a piece of double-sided tape was placed on the grid, how would you expect it to affect the CD of the distribution of peas per quadrat?
2. If the mean were 0.5 peas per quadrat, how many empty quadrats would you expect?
3. How would you expect the mean, variance and CD of the distribution of peas per quadrat to change if we increased the number of peas dropped from 150 to 175, assuming the distribution is still Poisson?
4. Suppose that every pea landed in a random fashion and stayed on the grid. How many peas would have to be dropped to ensure that not more than 1% of the quadrats remained empty (adapted from Hampton 1994)?

Acknowledgements

Special thanks to toddler Max Hosler for throwing a handful of split peas at VPB, Jeff Demarest for supporting this project, and Randy Bennett for useful discussion.

References

Brower, J.E., Zar, J.H. and von Ende, C.N. (1997). *Field and Laboratory Methods for General Ecology*. New York, NY: McGraw-Hill.

Hampton, R.E. (1994). *Introductory Biological Statistics*. New York: McGraw-Hill.

Shaughnessy, J.M. (1992). Research in probability and statistics: reflections and directions. In: D.A. Grouws (ed.), *Handbook of Research on Mathematics Teaching and Learning*, pp. 465–94. New York: Macmillan.

Sokal, R.R. and Rohlf, F.J. (1995). *Biometry*. New York, NY: Freeman.

Sowey, E.R. (2001). Striking demonstrations in teaching statistics. *Journal of Statistics Education*, **9**(1).