CONDITIONAL PROBABILITY: FRIEND OR FOE?

Marie Fitch and Stephanie Budgett
The University of Auckland
m.fitch@auckland.ac.nz

Many researchers in the probability education research field have noted that students frequently struggle to understand concepts related to conditioning. In order to support the development of these concepts in students enrolled in our first-year probability course we use interactive visual representations. We briefly discuss the theoretical ideas underpinning their use and then outline where and how their use has been incorporated into our teaching. Issues we have encountered, student reflections and ongoing areas for improvement and future research will also be discussed.

INTRODUCTION

Is this a joint probability or conditional probability? Am I supposed to be finding $\mathbb{P}(A|B)$ or $\mathbb{P}(B|A)$? These are such common questions amongst our students that when we read that students struggle to understand the concepts around conditioning, we know that researchers are expressing what we experience in our classrooms. Identifying the struggle many have with probability, and more specifically conditional probability, is not new (e.g. Chernoff & Sriramen, 2014). Visual representations have been proposed as a way to help students to better understand mathematical concepts in general (Arcavi, 2003) and more specifically statistical and probability concepts (Arnold, Pfannkuch, Wild, Regan & Budgett, 2011; Konold & Kazak, 2008; Pfannkuch, Budgett & Arnold, 2015).

In the context of conditional probability there are two very common misconceptions, the base-rate fallacy and confusion of the inverse (Villejoubert & Mandel, 2002). In both cases the result is all too often an inflated idea of the size of probability of interest. Gigerenzer (2014) found that using natural frequencies rather than probabilities led to a decrease in both misconceptions. Garcia-Retamero and Hoffrage (2013) and Binder, Krauss and Bruckmaier (2015) found that including visuals increased understanding and this difference was more pronounced when supporting information was given as natural frequencies.

As a means of developing ideas of probability based on random variables rather than events, Oldford and Cherry (2006) proposed the visual eikosogram representation. Pfannkuch and Budgett (2017) highlighted the potential benefits of using the eikosogram to promote probability reasoning. The pachinkogram (Budgett & Pfannkuch, 2016) has been specifically developed as a tool to tackle conditional probability misconceptions. They found the tool to be useful in increasing student understanding of the effects of changing the base rate and developing some probabilistic intuition. Student responses suggested that the visual aspects of the tool provide additional support for student understanding.

In this paper we focus on our experiences of including an activity based on the pachinkogram in our first-year probability course. Its use is preceded by a short activity using the eikosogram which we also briefly discuss.

OUR TEACHING CONTEXT

Our first-year probability course has a mathematics (algebra/calculus) co-requisite and about 120 students per semester. Students are a mix of very able mathematicians taking this as a mathematical statistics paper, and statistics majors taking the course because they have to. There are three lectures and one tutorial per week, which all last 50 minutes. The visual representations discussed in this paper are introduced and used during lectures. From the first lecture students are encouraged to make connections between words, mathematical symbols and visual representations and not to just stick with the representation that they are most comfortable with. Opportunities for active learning, which engage students through activities and/or discussions in class as opposed to passively listening to the lecturer (Freeman et al., 2014), are not restricted to tutorials. Discussion and time to attempt problems are built into lectures. The use of the eikosogram and pachinkogram both include discussion and opportunity for the students to frame their own solutions. A whole

lecture is devoted to the pachinkogram task, with students using their own devices to interact with the tool.

THE VISUAL REPRESENTATION TOOLS AND TASKS
The Eikosogram Tool

The eikosogram (see http://docker.stat.auckland.ac.nz/spawn/?application=probability) is a visual representation of a contingency table, based on the unit square, with areas matching probabilities. Vertical strips are scaled so that their width is equal to the probability of each distinct value of the conditioning variable. The rectangles are scaled to be proportional to the underlying joint probability, so that the height of each rectangle is equal to the corresponding conditional probability (Figure 1).
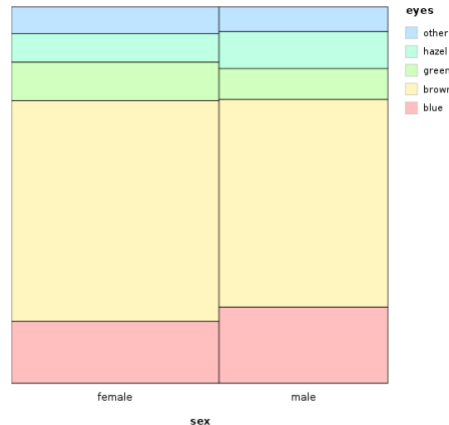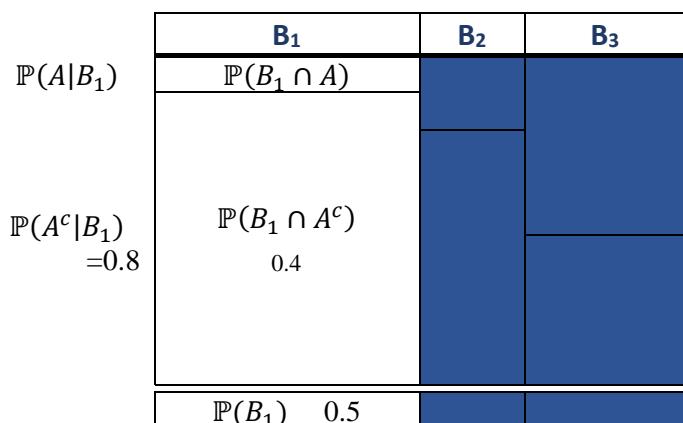

Figure 1. The eikosogram

The Eikosogram Task

In class a two-way table of counts is used to introduce the formal mathematical definition of conditional probability in an intuitive manner. Students are given a two-way table of counts (Figure 2(a)), which is used to calculate joint and marginal probabilities. This is then progressively scaled, by hand, into an eikosogram by the lecturer. (Figure 2(b) shows step 1). They are then asked to give the probability statement (e.g. $\mathbb{P}(A|B_1)$ or $\mathbb{P}(A \cap B_1)$) that is represented by the length of each side and the area of each rectangle. Figure 2(c) shows some of the probabilities written in. Students are challenged to consider whether the same rectangles would be drawn if we 'conditioned' on the rows rather than on the columns. A statement of the multiplication rule follows.

	B_1	B_2	B_3	Total
A	1	1	2	4
A^c	4	1	1	6
Total	5	2	3	10

(a) Two-way table of counts

	B_1	B_2	B_3
A	0.1	0.1	0.2
A^c	0.4	0.1	0.1
Total	0.5	0.2	0.3

(b) Contingency table scaled by B_i

(c) Eikosogram displaying some probabilities

Figure 2. Creating an eikosogram from a contingency table

The Pachinkogram Tool

Although most students have already met and worked with probability trees many are quite unaware of what probabilities are represented on the branches. The pachinkogram (see www.stat.auckland.ac.nz/~vt/) is a dynamic probability tree with an additional property that the thickness of each branch is scaled to be proportional to the associated probability. When first opened all probabilities are set to 50% (see Figure 3(a)). Sliders are used to adjust probabilities. When the sample once button is pressed small black balls representing individual cases roll through the pachinkogram and 'fill' the buckets at the bottom of each branch, Figure 3(b).

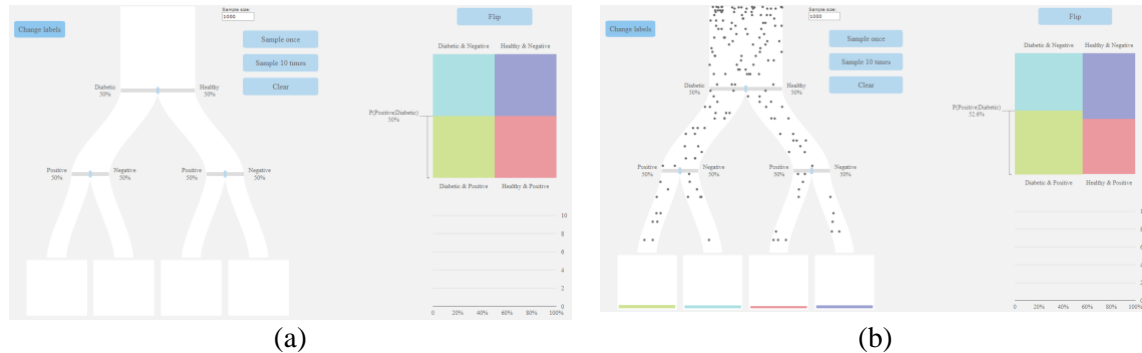(a)   (b)

Figure 3. Pachinkogram tool: default (a), showing a sample (b)

The Pachinkogram Task

The pachinkogram is used in the lecture following the introduction of the eikosogram and immediately prior to the formal introduction of Bayes theorem. At this point, as well as having been introduced to the eikosogram and also to probability trees in an earlier session, students have met the idea of conditional probability and performed calculations using both the definition of conditional probability and the multiplication rule.

Students are given a worksheet in which they use the pachinkogram to explore the following questions:

- Does $\mathbb{P}(A|B) = \mathbb{P}(B|A)$ always / never / sometimes (when)? *Circle one*
- What affects the values of $\mathbb{P}(A|B)$ and $\mathbb{P}(B|A)$?
- Do sample estimates of conditional probabilities have the same value as the theoretical calculations?

Before beginning they are asked to note what their current answers are to the questions.

A class discussion around the incidence of diabetes in the New Zealand population and the accuracy of screening tests sets the scene for the scenario used. The context of diabetes is used as most students are familiar with it and are aware of the variability of its incidence in different population groups. We also spend some time discussing positive and negative test results and the fact that screening tests are not 100% accurate. Students, most of whom have brought a web-enabled device to class, are encouraged to work in groups of two to four and to discuss their findings as they go rather than just working in isolation.

Students begin by adjusting the sliders on the pachinkogram to reflect the given base rate and test accuracies (see Figure 4(a)). They are then able to simulate the described scenario where, with each simulation, the black balls corresponding to members of the sample flow through the branches of the pachinkogram and arrive in one of four possible buckets. The effect of the base rate on the probability of someone who receives a positive screening test result being diabetic, $\mathbb{P}(\text{Diabetic}|\text{Positive})$, is explored through changing the base rate on the pachinkogram and noticing what else changes. The eikosogram in Figure 4(b) displays the value of $\mathbb{P}(\text{Positive}|\text{Diabetic})$ resulting from one simulation. The flip button is used to change the conditioning variable on the eikosogram so that students can compare $\mathbb{P}(\text{Diabetic}|\text{Positive})$ with the probability of someone who is diabetic receiving a positive screening test result, $\mathbb{P}(\text{Positive}|\text{Diabetic})$. Note that students are also able to visualize these two conditional probabilities, $\mathbb{P}(\text{Positive}|\text{Diabetic})$ and $\mathbb{P}(\text{Diabetic}|\text{Positive})$, in the pachinkogram by considering the volumes in the buckets at the bottom of the appropriate branches. $\mathbb{P}(\text{Positive}|\text{Diabetic})$ involves assessing the volume in the green bucket as a proportion of the sum of

the volumes in the green and blue buckets, while $\mathbb{P}(\text{Diabetic}|\text{Positive})$ involves assessing the volume in the green bucket as a proportion of the sum of the volumes in the green and pink buckets (see Figure 4(a)).
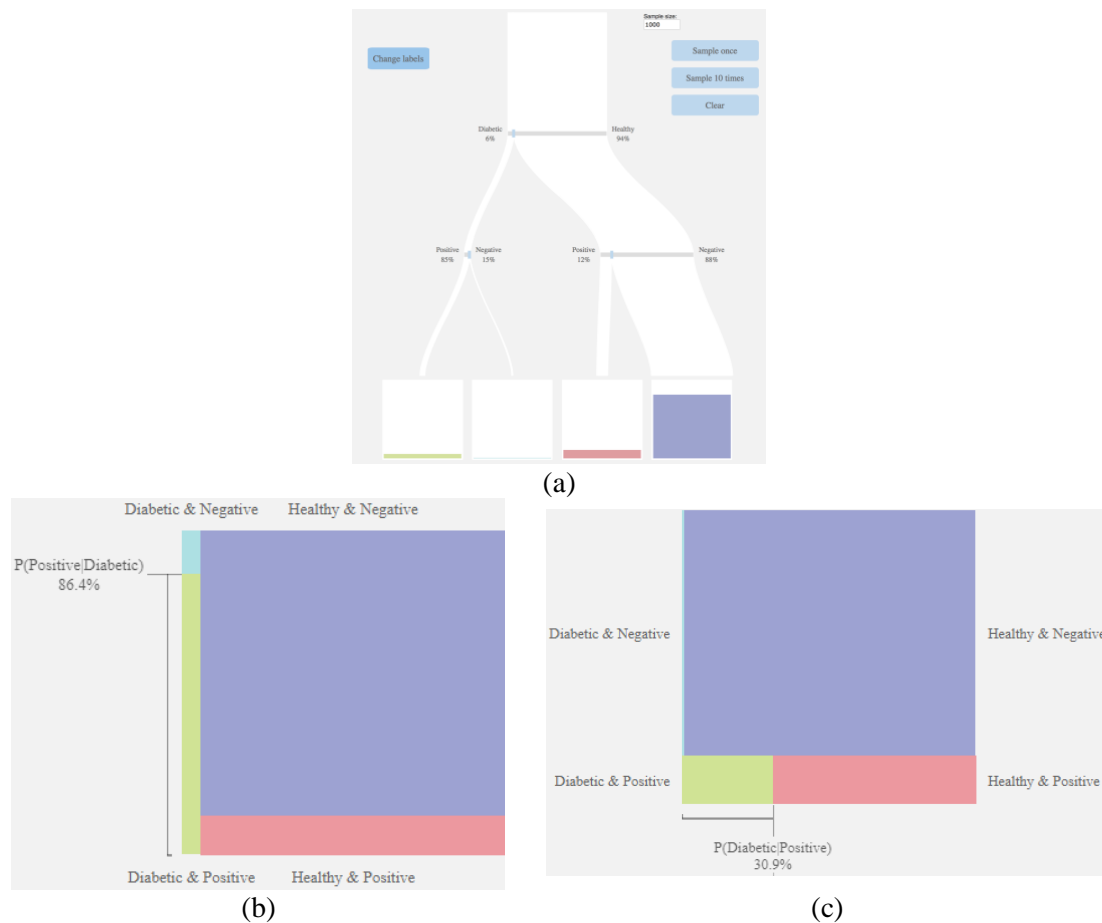

(a)

(b)

(c)

Figure 4. Pachinkogram denoting screening scenario (a); eikosogram conditioned on: diabetic status (b); screening test status (c)

The students are encouraged to further explore the relationship between the base rate and $\mathbb{P}(\text{Diabetic}|\text{Positive})$ by considering what they would consider an acceptable probability of having diabetes if they had received a positive test result. They adjust the base rate to try and achieve their preferred probability. This section concludes with looking at a screening test with different accuracies and considering what, if any, difference that makes to the probabilities. The final section asks students to use the tool to explore whether or not it is ever possible for $\mathbb{P}(A|B) = \mathbb{P}(B|A)$.

Class feedback and further discussion occurs at several points throughout the class. The class concludes with students being directed back to the three questions that began the class, to answer them again and reflect on how and why any of their answers have changed as a result of the activities they have just worked through.

STUDENT RESPONSES

In August 2017, immediately after the class in which the pachinkogram was used, students were invited to give online anonymous feedback about their experiences. Just under half of the class (49 students) responded to the survey. Of those who had attended the lecture (40) seven had taken time since the lecture to further explore using the tool. A further four students who had missed the lecture had taken time to use the tool. The students were also given a free response question which asked "How useful (or not) did you find using the pachinkogram? Please be specific about what was and was not helpful and how it aided (or not) your understanding".

Two of those who had not attended the lecture commented.

"*...it is a very interesting and fun way to show probability. It made me interested...*"

"It is a really good visual tool to aid in understanding. I found this concept (with all the symbols) really hard to grasp and, as a visual learner, the pachinkogram really aided my learning."

Most of the responses from those who had attended the lecture were very positive. There were only six negative responses, three of which were about the difficulty of using the tool on a mobile phone or tablet with the remainder saying it was hard, confusing or that it did not further their understanding. A further seven students did not respond to the free response question.

Many of the 27 who students who rated the tool positively commented specifically on how the tool had increased their understanding of conditional probability in general as illustrated in the following comment:

"(It) provided a very useful visual representation of the process of answering the problem. It also was useful in more general understanding of conditional probabilities and extended off the concept of tree diagrams. It was a great way to incorporate hands-on [interactive] and visual learning into the course material."

For some the tool was much more intuitive than mathematical formulae. For these students, the tool enabled them to develop an understanding of concepts which they found complex. Several students commented on the usefulness of being able to explore the effects of changing the base rate without getting bogged down in complex calculations. This gave them the confidence to make conjectures and then to check them.

"It's helpful since it shows a visible diagram which can be altered for different results which made it easier to think about the outcome and how it occurred."

"... it can let us the predict some of the answer as we can check the answer from pachinkogram and it can save the time of calculating the answer, so that we can check the concept or thinking is correct or not."

For at least one student the benefit was broader and they felt its worth went beyond their understanding of conditional probability to probabilistic reasoning in general.

"It gets the calculating process visualized while solving questions about conditional possibilities and it was way more obvious to find out all possibilities involved in questions. Moreover, it seems to be able to strengthen my probabilistic reasoning ability."

CONCLUSION

Our experiences suggest that there are three keys to student engagement with the pachinkogram tool. Firstly, informing students of the key learning outcomes and encouraging them to make a conjecture before beginning and then to reflect upon what they have learnt, not only gives students a clear purpose for engaging in the activity but also brings their learnings to the fore as they leave class. Secondly, a key part to using the pachinkogram tool is to use a relatable context and to ensure that adequate time is given to setting the scene and explaining the tool. In our initial forays into using the tool this step was rushed and we found many more students were confused about what they were doing and why. Finally, ease of use is vital and to this end further development of the tool to ensure that it is fully usable on mobile phones and tablets would also help to limit student frustrations.

While we are heartened by student perception of increased understandings we are yet to see that fully borne out in assignment and examination answers. Their answers suggest that the root of their remaining issues with conditional probability appear to be centred on a failure to recognise the difference in wording between a conditional probability and a joint probability, rather than issues with the base rate fallacy or confusion of the inverse. It may be worth exploring whether setting up ongoing problems where students can use the tool to quickly check if they have understood word-based scenario descriptions is helpful.

Overall visual tools show much promise in aiding and increasing understanding of conditional probabilities.

REFERENCES

Arcavi, A. (2003). The role of visual representations in the learning of mathematics. *Educational Studies in Mathematics, 52*, 215-241.

Arnold, P., Pfannkuch, M., Wild, C., Regan, M., & Budgett, S. (2011). Enhancing students' inferential reasoning: From hands-on to "movies". *Journal of Statistics Education, 19*(2), 1-32. Retrieved from http://www.amstat.org/publications/jse/v19n2/pfannkuch.pdf

Binder, K., Krauss, S., & Bruckmaier, G. (2015). Effects of visualizing statistical information - an empirical study on tree diagrams and 2 x 2 tables. *Frontiers in Psychology, 6*(1186). doi:10.3389/fpsyg.2015.01186

Budgett, S. & Pfannkuch, M. (2016). Visualizing chance: tackling conditional probability misconceptions. *13th International Congress on Mathematical Education.* 1-8.

Budgett, S., Pfannkuch, M., Regan, M., & Wild, C. J. (2013). Dynamic visualizations and the randomization test. *Technology Innovations in Statistics Education, 7*(2), 1-21. Retrieved from http://escholarship.org/uc/item/9dg6h7wb

Chernoff, E. J., & Sriraman, B. (Eds.). (2014). *Probabilistic thinking: Presenting plural perspectives.* Dordrecht, The Netherlands: Springer. doi:10.1007/978-94-007-7155-0

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., and Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences*, 111(23), 8410-8415.

Garcia-Retamero, R., & Hoffrage, U. (2013). Visual representation of statistical information improves diagnostic inferences in doctors and patients. *Social Science and Medicine, 83*, 27-33.

Gigerenzer, G. (2014). *Risk savvy: How to make good decisions*. New York, NY: Viking.

Konold, C., & Kazak, S. (2008). Reconnecting data and chance. *Technology Innovations in Statistics Education, 2*(1). Retrieved from http://escholarship.org/uc/item/38p7c94v

Oldford, R.W., & Cherry, W.H. (2006). Picturing probability: the poverty of Venn diagrams, the richness of eikosograms. Retrieved from http://www.stats.uwaterloo.ca/~rwoldfor/papers/venn/.

Pfannkuch, M., & Budgett, S. (2016). Reasoning from an Eikosogram: An exploratory study. *International Journal of Research in Undergraduate Mathematics Education*, 1-28. doi:10.1007/s40753-016-0043-0

Pfannkuch, M., Budgett, S., & Arnold, P. (2015). Experiment-to-causation inference: Understanding causality in a probabilistic setting. In A. Zieffler, & E. Fry (Eds.), *Reasoning about uncertainty: Learning and teaching Informal Inferential Reasoning* (pp. 95-127). Minneapolis, MN: Catalyst Press.

Villejoubert, G., & Mandel, D. R. (2002). The inverse fallacy: An account of deviations from Bayes theorem and the additivity principle. *Memory & Cognition, 30*, 171-178. doi:10.3758/BF03195278