THE UNIVERSITY OF AUCKLAND
DEPARTMENT OF STATISTICS

Research supported by TLRI

Visualising chance: A pilot study at the introductory level

Stephanie Budgett & Maxine Pfannkuch
The University of Auckland

THE UNIVERSITY OF AUCKLAND
DEPARTMENT OF STATISTICS

Outline

- Rationale for project
- Visual strategies
- Two pilot study tasks and students' interactions with prototype tools
- Conjectures

THE UNIVERSITY OF AUCKLAND
DEPARTMENT OF STATISTICS

Rationale

- Current teaching mainly based on a mathematical and empirical approach
    - *while the teaching and learning of statistics takes on an enquiry-based problem-solving stance, where students act as data detectives, the pedagogy of probability is ever more isolated in its strange world of coins, spinners and dice as tools for demonstrating in a rough and ready way the existence of theoretical probability* (Pratt, 2011, p. 891)
- Need to re-connect data and chance in education (Konold & Kazak, 2008)

THE UNIVERSITY OF AUCKLAND
DEPARTMENT OF STATISTICS

Rationale

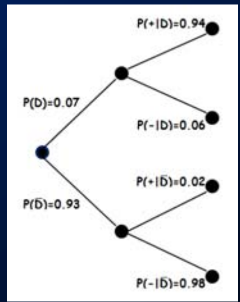
- Probability conceptions should develop around the notion of probability as a modelling tool (Pratt, 2011; Garfield et al., 2012)
    - Exploring the behaviour of models and using visual imagery may assist conceptual development
- Many misconceptions associated with probability documented
    - Our focus
        - Conditional probability (Gigerenzer, 2002; Kahnemann, 2011)
        - Independence (Watson, 2005)

Current teaching strategies

- Tables of counts/proportions

Gender	Blue	Brown	Eye Colour Green	Hazel	Other	Total
Male	37	101	15	18	12	183
Female	37	132	23	17	16	225
Total	74	233	38	35	28	408

- Probability trees



THE UNIVERSITY OF AUCKLAND
DEPARTMENT OF STATISTICS

Current teaching strategies

- Mathematics (e.g. Independence)

Statistical independence:
if A and B are independent, then
$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\,\mathbb{P}(B)$$
and
$$\mathbb{P}(A \mid B) = \mathbb{P}(A)$$
and
$$\mathbb{P}(B \mid A) = \mathbb{P}(B)\,.$$

- Procedures

$$\mathbb{P}(\overline{W} \mid D) = \frac{\mathbb{P}(D \mid \overline{W})\mathbb{P}(\overline{W})}{\mathbb{P}(D \mid \overline{W})\mathbb{P}(\overline{W}) + \mathbb{P}(D \mid W)\mathbb{P}(W)} = \frac{3/4 \times 2/3}{3/4 \times 2/3 + 1/2 \times 1/3} = \frac{3}{4}$$

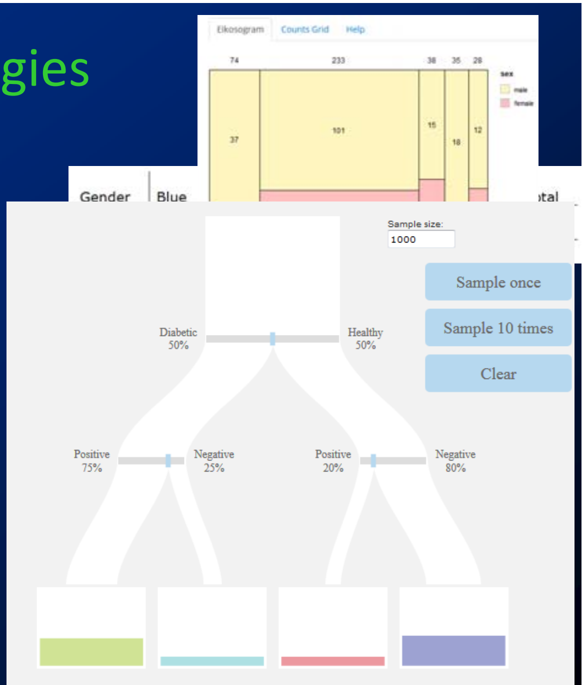
- Word problems

Assume you have a deck of cards consisting of 26 red and 26 black cards. Assume you draw a card at random from the deck and place it face down on the table without looking at it. You then draw a second card from the remaining cards and observe that it is red. What is the probability that the first card is red as well?

THE UNIVERSITY OF AUCKLAND
DEPARTMENT OF STATISTICS

Visual strategies

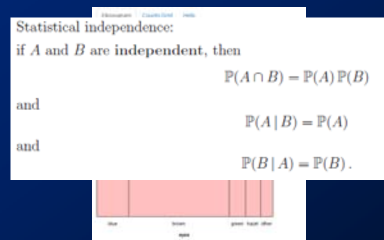
- Eikosogram
- Pachinkogram



THE UNIVERSITY OF AUCKLAND
DEPARTMENT OF STATISTICS

Visual strategies

- Visual independence
  - eikosogram



- Exploration
  – explore behaviour of model
  – simulation
- Literacy strategy
  – build intuition
  – develop a sense of the problem

Pilot study approach

- Prototype tools and tasks developed
- Participants: Four university students who had completed a first-year probability course
  - Pre-test: individual
  - Students worked in pairs – thinking captured and probed:
    - One pair did an Eye Colour task
    - One pair did a Diabetes task
  - Post-interview in pairs

THE UNIVERSITY OF AUCKLAND
DEPARTMENT OF STATISTICS

Eye Colour task- initial thinking

1. Do you think that an individual's eye colour depends on whether that individual is male or female?
2. Assume that eye colour is taken to be one of four possibilities: Brown, Blue, Green, Other

   In NZ, what proportion of the population do you estimate to have brown eyes?
3. Draw a representation of your estimated distribution of eye colour.
4. If eye colour and gender are independent, draw a representation of your estimated distribution of eye colour and gender
5. Repeat (4) under the assumption that eye colour and gender are dependent.

THE UNIVERSITY OF AUCKLAND
DEPARTMENT OF STATISTICS

Eye Colour Task

- Show eikosogram movie

THE UNIVERSITY OF AUCKLAND
DEPARTMENT OF STATISTICS

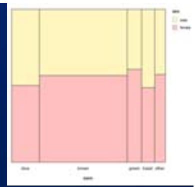
EYE COLOUR TASK

How did students interact with the task?

What story is the representation telling you?

THE UNIVERSITY OF AUCKLAND
DEPARTMENT OF STATISTICS

Eye Colour Task



- Learning to reason with an eikosogram
  - Decoding representation
    - Comparing numbers ("lot more females with green eyes than males")
  - Interpreting representation (a prompt)
    - Conditional ("60% of the people with green eyes are female")
    - Joint ("the area of an individual box would represent …brown eyes, brown-eyed girls, yeah … so one box is a proportion of the whole population")

THE UNIVERSITY OF AUCKLAND
DEPARTMENT OF STATISTICS

Eye Colour Task

- Interpreting SWAP FACTORS representation
  - Comparing Joint
    - "you can see that the proportion of males with blue eyes is higher than the proportion of females with blue eyes … the proportion is bigger. The total number might not be bigger but the proportion is definitely higher"
  - Independence representation
    - Able to describe and why (in pre-test unable to do)

THE UNIVERSITY OF AUCKLAND
DEPARTMENT OF STATISTICS

Eye colour task

- Exploring data on own
  - Articulating conditional probability questions and interpretations
  - Considering independence alongside conditional probability
    - "Of the people who have tattoos a very large number of them also have piercings … I guess that also shows that they are not independent as well"

THE UNIVERSITY OF AUCKLAND
DEPARTMENT OF STATISTICS

Eikosogram tool – student reflections

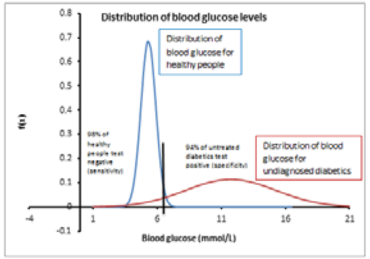
- Non-numeric visual (compared to their table)
  - "Puts mind on proportions"
  - "Comparing proportions easier"
  - "Numbers (e.g., 37% and 33%) are deceptive"
- Swap factors
  - Made a difference perceptually
  - "Easier to think with factors on the bottom"
  - Saw things they did not see in reversed visual
- Independence – "useful to visualise"
- Believed would have been "very helpful" to them when learning probability

THE UNIVERSITY OF AUCKLAND
DEPARTMENT OF STATISTICS

Diabetes task

A new housing development has been built in your neighbourhood. In order to service the needs of this new community, a new health clinic has opened. As part of the health clinic's enrolment procedure, new patients are required to undergo health check-ups which include, among other things, a series of blood tests. One such test is designed to measure the amount of glucose in an individual's blood. This measurement is recorded after the individual fasts (abstains from eating) for a prescribed period of time. Fasting blood glucose levels in excess of 6.5mmol/L are deemed to be indicative of diabetes. This threshold of 6.5mmol/L works most of the time with about 94% of people who have diabetes being correctly classified as diabetics and about 98% of those not having diabetes being correctly classified as non-diabetics.

The prevalence of diabetes in the NZ population is about 7%[1] (i.e. approximately 7% of the NZ population are estimated to have diabetes).



*Graph above taken from (Pfannkuch, Seber, & Wild, 2002)

Question 1: As part of enrolment in this health clinic, an individual has a fasting blood test. He/she is told that his/her blood glucose level is higher than 6.5mmol/L. What are the chances that he/she has diabetes?

Intuitive answer:

a) How might the prevalence of diabetes in the NZ population be estimated?

b) Out of 100 individuals, how many could be expected to have diabetes?

c) Out of 100 individuals with diabetes, approximately how many will have a fasting blood glucose level of less than 6.5mmol/L?

d) Out of 100 individuals without diabetes, approximately how many will have a fasting blood glucose level of more than 6.5mmol/L?

e) If we changed the threshold value from 6.5mmol/L to 7mmol/L
   i. how would this affect the percentage of correctly classified diabetics?
      Increase / Stay the same / Decrease
   ii. how would it affect the percentage of correctly classified non-diabetics?
      Increase / Stay the same / Decrease
   iii. What would be the implications of this change?

f) If we changed the threshold value from 6.5mmol/L to 5mmol/L
   i. how would this affect the percentage of incorrectly classified diabetics?
      Increase / Stay the same / Decrease
   ii. how would it affect the percentage of incorrectly classified diabetics?
      Increase / Stay the same / Decrease
   iii. What would be the implications of this change?

g) Who would constitute members of the sample space?

h) Do I have enough information to answer Question 1? If not, can I find the necessary information, or make certain assumptions?

i) Is there any superfluous information?

j) What pieces of information do I need in order to model this problem?
   I need to know: the reliability of the test for diabetics Yes / No
   the reliability of the test for non-diabetics Yes / No
   the prevalence of diabetes in the population Yes / No

THE UNIVERSITY OF AUCKLAND
DEPARTMENT OF STATISTICS

Diabetes task - initial thinking

- Fasting blood glucose levels for detecting diabetes
- Prevalence of diabetes is known (or can be estimated)
- If diabetic, probability of testing positive is given
- If non-diabetic, probability of testing negative is given

QUESTION-

If a person tests positive, what is the probability that they have diabetes?

THE UNIVERSITY OF AUCKLAND
DEPARTMENT OF STATISTICS

Diabetes task

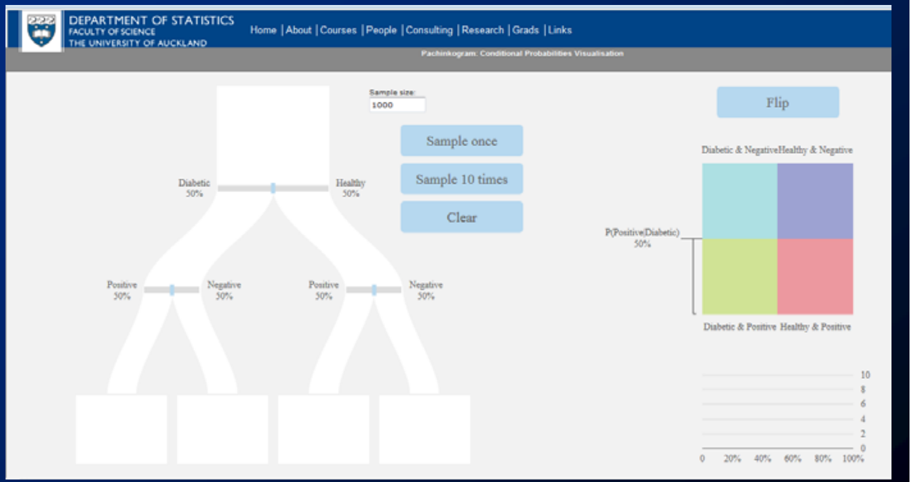
- Pachinkogram movie part 1
- Pachinkogram movie part 2

THE UNIVERSITY OF AUCKLAND
DEPARTMENT OF STATISTICS

Pachinkogram tool – student reflections

- Literacy strategy
  - "It's like a real life situation"
  - "It gives the human element"
- Non-numeric visual
  - "Really good to see a visual representation of the proportions rather than a number"
  - "Oh my gosh, if you had been able to show us those in class, oh it would have been like *I get it!*"
  - "Instead of having one learning tool you've got two. Visual might be better for some, written might be better for some… and both is surely better than one."

THE UNIVERSITY OF AUCKLAND
DEPARTMENT OF STATISTICS



THE UNIVERSITY OF AUCKLAND
DEPARTMENT OF STATISTICS

Conjectures

- Visual imagery through these tools may assist in:
  - Visualising chance as a proportion
  - Articulating stories in the data – using chance language
  - Exploring behaviour of a probability model – reconnecting chance and data – empirical versus one mathematical answer
  - Concept formation (conditioning and independence)
  - A transition to a mathematical approach

THE UNIVERSITY OF AUCKLAND
DEPARTMENT OF STATISTICS

Future research

- Develop prototype tools and tasks for
  - Markov processes (tool already developed)
  - Poisson processes (tool to be developed)
- How would beginners interact with these tools?

- Thanks to TLRI for research support

TEACHING & LEARNING RESEARCH INITIATIVE
NĀU I WHATU TE KĀKAHU, HE TĀNIKO TAKU