The background

As described in Part 1 [2], the concepts, structures and thinking of probability and distributions underpin all statistics, but the development of these in statistics education has received far less attention over the past decade than the topics of what has become known as the “statistics education reform movement”. This reform has been oriented to data and data investigations, with emphasis on statistical literacy, reasoning and thinking. Many statisticians think it is time to bring the principles of data-driven, holistic, authentic and active learning into the development of probabilistic and distributional understanding, reasoning and modelling.

Probability is not just a sub-section of mathematics with limited applications, and distributions are not just about “lumps” of data. Just as the importance of data, statistical literacy and statistical thinking has been increasing across disciplines and society, so too has the importance of probabilistic and distributional thinking and modelling in themselves, as well as underlying all sound statistical data analysis techniques and the statistical thinking and practice of professional statisticians. Computing power and areas such as risk analysis are increasing the demand for stochastic and distributional thinking in applications and problem-solving in statistics, mathematics, business, information technology, engineering and health, physical and social sciences.

This article continues the selection of examples and problems specifically developed for a course in introductory probability and distributional modelling, which is one of the two first year statistics courses at the Queensland University of Technology that are compulsory for all mathematics and statistics majors, including mathematics education students. This course [4, 5] builds skills and foundations in basic concepts, thinking and methods in probability, conditional arguments, distributional and stochastic modelling for applications in a wide range of areas, from communication systems and networks to traffic to law to biology to financial analysis. The course helps students to unpack, analyse and extend their prior knowledge, understanding and misunderstandings. The learning and assessment package of the course is built around problem-solving, with activities that link data, everyday processes, student experiences, modelling and simulation, and consolidate core mathematical skills.

These two articles give a number of the problems developed for this course, with comments on their aims. The problems given in these articles tend to be the most challenging, with prior examples and exercises focusing on developing components of the reasoning in a wide variety of contexts. The problems selected...
for these articles are also in the sections that benefit most from the tutorial group exercise strategy outlined in [3] that creates the problem-solving learning environment envisioned in [1]. No problems are included from sections requiring standard mathematical techniques such as obtaining expected values and variances. Below are selected problems from introductory Markov chains, special distributions, introductory queues, finding means through conditioning, and linear combinations of normal random variables.

Examples from introductory Markov chains
The focus in this section is on the setting up of the matrix of transition probabilities. Apart from finding 2-step and 3-step transition probabilities, all theory is left to second and third year courses, with the stationary result being justified by demonstration and simulation. As with conditional probabilities, the activities and learning experiences embody all aspects of quantitative problem-solving, and include either modelling of probabilities or estimation from data.

1. A type of switch has a probability of 0.1 of failing by the beginning of the next day if it is working at the beginning of a day. Two of these switches are placed in parallel in a component in a system, so that only one of them needs to be working for the component to work within the system. The switches fail or not independently of each other. The component does not receive any maintenance unless both switches have failed, in which case both switches are replaced for the beginning of the next day. The number of switches that are working in the component at the beginning of each day is a Markov chain.

   (i) Find the matrix of transition probabilities for this Markov chain.
   (ii) Over a long period of time, for what proportion of time is the component not working?

2. The number of machine failures, Y, in a day at a certain plant has the following probabilities.

   The plant can repair one machine per day, but a whole day is required for the repairs. That is, a machine’s repair will not be completed on the same day it breaks down. However, the plant cannot cope with having more than 3 machines down. Hence if the plant has more than 3 machines down, the excess machines are sent out for repairs.

   The number of machines to be repaired within the plant at the beginning of each day is a Markov chain.

   (i) Obtain the matrix of transition probabilities for this Markov chain.
   (ii) Assuming steady-state conditions, find the probability that there are no machines waiting to be repaired at the beginning of a day.

3. A binary signal (takes the values 0 or 1) starts as a 0 and passes through a number of stages where its state is noted each time, giving the following sequence.

   000111110111000000

   (This sequence includes the first 0).

   Assuming that the state of the signal at stage n is a Markov chain, use the above observed sequence to estimate the transition probabilities in the 2x2 transition matrix.

   “Many statisticians think it is time to bring the principles of data-driven, holistic, authentic and active learning into the development of probabilistic and distributional understanding, reasoning and modelling.”

Examples from special distributions
This section separates the three aspects of special distributions to help students develop their skills in recognising and applying them. These are: (a) the circumstances and assumptions of special distributions; (b) the mathematics of special distributions; and (c) the use of special distributions in context. It was with great interest to discover in 1994, that staff in the Statistics Department of University of Glasgow were developing the same kinds of questions as I was in question 1 below to facilitate students’ understanding of (a) above without the distractions of (b) so that they could develop better skills in (c). Ongoing collaboration is now resulting in development of a program as outlined in [6]. Questions such as question 2 below focus on identification of random variables and their distributions.

1. In each of the following situations, suggest a distribution that could be suitable for the variable:

   (i) The number of trolleys with defective wheels at a supermarket;
   (ii) The number of supermarket trolleys a shopper has to try before finding one without defective wheels;
   (iii) The number of tins of tomatoes a shopper has to check before obtaining 3 without any rust spots;
   (iv) The number of shoppers arriving in 5 minutes;
   (v) The number of checkouts open when the shopper is choosing a queue to join;
   (vi) The service time per customer in the express queue.

   (This sequence includes the first 0).

   Assuming that the state of the signal at stage n is a Markov chain, use the above observed sequence to estimate the transition probabilities in the 2x2 transition matrix.

   “Many statisticians think it is time to bring the principles of data-driven, holistic, authentic and active learning into the development of probabilistic and distributional understanding, reasoning and modelling.”

Examples from an introductory course in developing probabilistic statistical thinking: Part 2 – Helen MacGillivray
The time taken to complete a computer job is exponentially distributed, independently of other jobs, with, on average, 2 completed per minute. Find the probability that:

(i) completing a job takes less than a minute;
(ii) more than 15 are completed in 5 minutes;
(iii) out of 10 (independent) jobs, at least 2 take less than a minute to complete.

"It is hoped that this small selection of problems will whet statisticians’ appetites for enriching and revolutionising the teaching of introductory probabilistic and distributional thinking and modelling…"
(ii) Use the correlation found in (i) in the following: If the arrival time of the bus at this stop is normally distributed around 7.48 am with a standard deviation of 2 minutes, what is the chance that Fred misses the bus? (Assume the bus’s arrival time is independent of Fred’s.)

2. The application rate for a pesticide in a building is supposed to be 17.5 gms per sq m, but is actually normally distributed around this value with a standard deviation of 0.9 gms per sq m. The safe re-entry level of humans into a sprayed building is 6 gms per sq m. Suppose that the amount of pesticide dissipated per day (by absorption etc.) in a sprayed building is normally distributed with a mean of 1.5 gms per sq m and a standard deviation of 0.5 gms per sq m.

(i) Assuming that the amounts dissipated per day are independent of each other and of the amount applied, show that the distribution of the pesticide residue in the building ten days after spraying is normal with mean 2.5 gms per sq m, and variance 3.31 gms² per sq m.

(ii) Find the probability that it is safe to re-enter the building ten days after spraying.

(iii) The only parameter that can be controlled in the above is the standard deviation of the application rate (which is 0.9 above). To what should this be reduced, if we wish the probability that it is safe to re-enter the building after ten days, to be at least 0.98?

(iv) Assume now that the amount dissipated in the first day has a correlation of 0.1 with the amount applied, and that subsequently the amount dissipated each day has a correlation of 0.05 with the amount dissipated the previous day but no correlation with days other than those immediately preceding or following. Show that the pesticide residue in the building ten days after spraying is now distributed as a normal with mean 2.5 gms per sq m, and variance 3.445 gms² per sq m.

Conclusion

It is hoped that the above small selection of problems will whet statisticians’ appetites for enriching and revolutionising the teaching of introductory probabilistic and distributional thinking and modelling, not by denying the mathematics that is so important, but by building a problem-solving environment that connects with the real, the everyday, data and students’ prior learning, and helps them develop a sound foundation for their future learning and problem-solving in a range of disciplines and workplaces. It is also hoped that this selection will bring forth contributions from many others with a love of probability, distributions and modelling for real processes.

References


