Oh no! Puncture!

By **Michael Fletcher**

Driving in France recently, I got a puncture in my tyre. I was 13 km from my hotel and needed to get back, but for each kilometre I drove my tyre got flatter, so I had to drive more slowly.

For the first kilometre I drove at a constant speed. For the second kilometre I drove at a slower constant speed. Similarly, for each of the remaining kilometres I drove at constant (albeit decreasing) speeds. My average speed for the first 2 km was a whole number of kilometres per hour, as were my average speeds for the first 3 km, 4 km, 5 km, …. Indeed, the average speeds for the first i km of my journey back to the hotel were whole numbers of kilometres per hour, all the way up to $i = 13$.

The speed limit on the route was 100 km/h. How many minutes did it take me to drive the last kilometre back to my hotel? Send your answers to significance@rss.org.uk. Explain how you came up with the solution, as we may publish a selection of correct entries (if received by 9 May 2022).

Solution to "How many balls in the bag?"

Last time, I was taking part in a game at a village fête where I had to guess the number of balls in a bag. Balls were numbered, and I was able to draw four balls to help me make a guess. I drew 24, 87, 14 and 35. There was a prize for guessing the correct number exactly, so I asked for advice on what estimate I should make.

The answer I was looking for: 87. Clearly, there are at least 87 balls in the bag, so we could consider, as reader Eduardo Rohde did, what size of bag is most likely to produce 87 as its largest of four draws. Rohde arrived at the answer with some R code and simulations. In essence, what he set out to do was to calculate the probability that the highest ball drawn is 87, given that there are N balls in the bag. Start by assuming there are 87 balls in the bag, so $N = 87$. We calculate this conditional probability as:

P(highest ball drawn is 87 | there are 87 balls)

$$= 4 \times \frac{1}{87} \times \frac{86}{86} \times \frac{85}{85} \times \frac{84}{84} = \frac{4}{87}$$

$$\approx 0.046$$

What this equation says is that we have four draws, there is a 1 in 87 chance of drawing an 87, and we are guaranteed to draw a number lower than 87 once we have drawn our 87. We multiply everything by 4 because there are four possibilities for which draw gives you the 87, and we need to sum over them. If we were to assume 88 balls in the bag, the conditional probability would be lower:

P(highest ball drawn is 87 | there are 88 balls)

$$= 4 \times \frac{1}{88} \times \frac{86}{87} \times \frac{85}{86} \times \frac{84}{85} = \frac{4}{87} \times \frac{84}{88}$$

$$\approx 0.044$$

Now we have a 1 in 88 chance of drawing an 87, and we are no longer guaranteed to draw a number lower than 87 once we have drawn that number. If we assume 89 balls in the bag, the conditional probability would be lower still, and so on. We can express this calculation more generally as:

P(highest ball drawn is 87 | there are N balls)

$$= 4 \times \frac{1}{N} \times \frac{86}{N-1} \times \frac{85}{N-2} \times \frac{84}{N-3}$$

$$= \frac{4}{87} \times \frac{87!(N-4)!}{N!83!}$$

For $N > 87$, the conditional probability will always be lower than when $N = 87$. So, 87 is the maximum likelihood estimate for the number of balls in the bag.

But wait! Rohde discussed this problem with a friend, who argued that the conditional probability was the wrong way round: he should be calculating the probability of there being N balls in the bag, given that the highest ball drawn is 87:

P(there are N balls | highest ball drawn is 87).

The friend also compared our problem to the famous German tank problem,[1] where Allied spies in the Second World War identified some tank numbers that they knew came from a sample space 1, 2, 3, …, N, and they wanted to estimate how many tanks the German forces had. It is true that the two problems are similar, but there is a crucial difference: for the Allies, there was no prize for guessing the exact number. It was more important that the estimate was in the right ballpark. So, a different calculation was used:

$$\hat{N} = m + \frac{m}{k} - 1$$

Here, k is the number of observations (so, 4 in our example) and m is the highest number observed (87). Plugging our numbers in gives 108 as the (nearest round number) estimate of the number of balls in the bag.

Rohde's friend calculated the same and noted a 50% chance that the actual number of balls in the bag lies between 90 and 126. But, now it was Rohde's turn to point out a problem: his friend was essentially making a bet on every number between 90 and 126. "But the rules of the fête don't allow that," says Rohde. "He has to pick one." So, the friend went away and calculated the probability of a bag containing N sequentially numbered balls, given that the largest of four successive draws is 87. The answer: "the probability of the bag having 108 balls is 1.4%, standing in the 50th percentile, consistent with the German tank formula… [but] the most likely number of balls in the bag is 87, with probability 3.4%".

Congratulations to Rohde (and friend), and also to Kurt Baldwin, Robert Elston, Kate Land, Peter Tymms and Clayton Ramsey, who all arrived at the answer of 87, albeit in different ways. ■

Reference

1. Grajalez, C. G., Magnello, E., Woods, R. and Champkin, J. (2013) Great moments in statistics. *Significance*, **10**(6), 21–28.