THE LOST BOARDING PASS, AND OTHER PRACTICAL PROBLEMS

GEOFFREY R. GRIMMETT, DAVID R. STIRZAKER

1. Puzzles in probability

The world is amply supplied with probability teasers. These can mystify and intoxicate. Here are some classics.

1. (Family planning) A family has three children, two of whom are boys. What is the probability the third is a boy?
2. (Bertrand's paradox) A chord of the unit circle is picked at random. What is the probability that an equilateral triangle with the chord as base can fit inside the circle?
3. (Monty Hall problem) Monty Hall shows you three doors concealing, respectively, two goats and one Cadillac, and he asks you to choose a door and take whatever is behind it. Without opening your door, he opens another and displays a goat. Is it to your advantage to accept his invitation to shift your choice to the third door?

These three puzzles have the common feature of being 'ill posed': insufficient information is given for the answer to be properly determined. It is left to readers to discuss formulations of and answers to the above teasers[1].

There are further teasers that tempt the reader to indulge in sometimes over-complex calculations. Here is one such.

4. (Red Now) This game is played by a single player with a well shuffled conventional pack of 52 playing cards. At times $n = 1, 2, \ldots, 52$ the player turns over a new card and observes its colour. Exactly once in the game s/he must say, just before exposing a card, "Red Now". The game is won if the next exposed card is red. Is there a strategy for the player that results in a probability of winning different from $\frac{1}{2}$?

Date: October 6, 2019, revised November 15, 2019.

[1]All problems in this article are considered in some detail in Grimmett and Stirzaker [1, 2, Chaps 1, 3, 4, 12].

It will be tempting to embark on a calculation that analyses the success probability of a general strategy, possibly using some fairly heavyweight machinery such as optional stopping. One would need firstly to decide what exactly makes a 'strategy', and secondly which strategies to prioritise (such as waiting until there is a given preponderance of reds in the unseen cards). There is however a beautiful 'one-line' solution[2] that finesses such complexities as follows, at the price of employing a touch of what might be called 'probabilistic intuition'. It may be seen that the chance of winning is the same for a player who, after calling "Red Now", picks the card placed at the bottom of the pack rather than that at the top. At the start of the game, the bottom card is red with probability $\frac{1}{2}$ irrespective of the strategy of the player.

We amplify this discussion with a more detailed analysis in Sections 2–3 of the well known 'lost boarding pass problem'[3]. This provides a surprising illustration of the probabilistic property of 'independence' of events, and reveals the full structure of the random mapping of passengers to seats.

2. One lost boarding pass

Question 1. *(Lost boarding pass problem) A plane has $n \geq 2$ seats, and is fully booked. A line of n people board one by one. Each has an assigned seat, but the first passenger (doubtless a mathematician thinking of other things) has lost his or her boarding pass and sits in a seat chosen at random. On boarding, subsequent passengers sit in their assigned seats if free, and in randomly chosen empty seats if not. What is the probability that the last passenger sits in the assigned seat?*

This problem has generated a good deal of discussion. It is well known that the answer is $\frac{1}{2}$, and several arguments may be found for this in puzzle books and online (see, for example, [3]). The neatest is probably the following (attributed by Peter Winkler [4, p. 35] in 2004 to Alexander Holroyd, but perhaps known earlier to others). Let the passengers be labelled $1, 2, \ldots, n$ in order of boarding, and label their assigned seats similarly (so that passenger k has been assigned the seat labelled k). The seating process is illustrated in Figure 2.1.

Solution 1. *When the nth passenger chooses a seat, there is only one available, and this must be either seat* 1 *or seat n (since seat*

[2] As explained to one of the authors by an anonymous undergraduate following a lecture, and probably the outcome of student debate.

[3] See [1, 2, Prob. 1.8.39].

$r \in \{2, 3, \ldots, n-1\}$, if free, would have been claimed earlier by passenger r). Each of these two possibilities has probability $\frac{1}{2}$, since no earlier decision has distinguished between them.

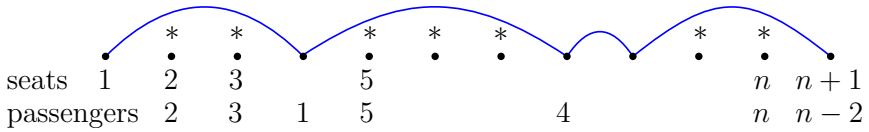

FIGURE 2.1. Seats are labelled $1, 2, \ldots, n$, and the extra label $n+1$ represents a copy of seat 1. In this example, passenger 1 sits in seat 4, passenger 4 is displaced into a later seat, and the chain of displacement continues until some displaced passenger chooses seat $n+1$ ($\equiv 1$). An asterisk denotes a correctly seated passenger.

Some will be uncomfortable with this solution, in that it uses a symmetry rather than an equation. Such is not atypical of arguments considered by mathematicians to be 'neat', though they frequently have the potential to mystify the uninitiated (see, for example, [5, p. 176]).

The solution presented in [1] is more prosaic than the above, and utilizes conditional probability to derive a recurrence relation. Such a method was used by Dermot Roaf [6] to obtain an answer to the following elaboration of Question 1.

Question 2. *What is the probability that passenger m finds seat m to be already occupied?*

We do not include Roaf's solution (which uses an equation). Here however is the 'one-liner'[4] that extends Solution 1.

Solution 2. *Let $2 \le m \le n$. When passenger m chooses a seat, seats $2, 3, \ldots, m-1$ are already taken (since seat $r \in \{2, 3, \ldots, m-1\}$, if free, would have been claimed earlier by passenger r), and one further seat also. No information is available about the label of this further seat. By symmetry, it is equally likely to be any of the $n-m+2$ seats labelled $1, m, m+1, \ldots, n$. Therefore, it is seat m with probability $1/(n-m+2)$.*

Let D_m be the event that passenger m finds seat m already occupied. Henze and Last [8] have observed that the D_m are independent events. We explain this using the simple argument of Solution 2.

Theorem 2.1. *The events $D_2, D_3, \ldots, D_n$ are independent.*

[4]Possibly noted first by Nigussie [7].

Proof. This is by iteration of the argument of Solution 2. Let $2 \le m_1 < m_2 \le n$, so that

$$\mathbb{P}(D_{m_1} \cap D_{m_2}) = \mathbb{P}(D_{m_1})\mathbb{P}(D_{m_2} \mid D_{m_1}). \tag{2.1}$$

When passenger m_1 chooses a seat, seats $2, 3, \dots, m_1 - 1$ have already been taken, together with one further seat denoted S_1. Conditional on D_{m_1}, we have $S_1 = m_1$, so that passenger m_1 is displaced to some later seat. When, subsequently, passenger m_2 chooses a seat, the seats $m_1+1, m_1+2, \dots, m_2-1$ are occupied and in addition one further seat denoted S_2. The choices made so far contain no information about S_2 other than it is one of $1, m_2, m_2+1, \dots, n$. Thus, $\mathbb{P}(D_{m_2} \mid D_{m_1})$ equals the (conditional) probability that $S_2 = m_2$, namely $1/(n-m_2+2)$. By (2.1),

$$\mathbb{P}(D_{m_1} \cap D_{m_2}) = \frac{1}{n-m_1+2} \cdot \frac{1}{n-m_2+2} = \mathbb{P}(D_{m_1})\mathbb{P}(D_{m_2}), \tag{2.2}$$

so that D_{m_1} and D_{m_2} are independent. By the same argument iterated, any finite subset of $\{D_m : 2 \le m \le n\}$ is independent. □

We summarise the above as follows. By Solution 2 and Theorem 2.1, the D_m are independent wth probabilities $p_m := \mathbb{P}(D_m)$ given by

$$p_m = \frac{1}{n-m+2}, \qquad m = 2, 3, \dots, n. \tag{2.3}$$

We may construct the seating arrangement 'backwards' as follows, by first choosing the disputed seats and then seating the passengers. For $s = 1, 2, \dots, n$, we declare s to be *red* with probability $1/s$, and *black* otherwise, with different integers having independent colours. We relabel seat 1 as seat $n+1$, so that the seat-set is now $S = \{2, 3, \dots, n+1\}$, and we declare seat m to be red if the integer $s = n-m+2$ is red, which occurs with probability p_m. The passengers may now be seated. Passenger 1 sits in the earliest red seat, R_1 say; passenger R_1 sits in the next red seat, R_2 say, and so on; undisplaced passengers sit in their assigned seats.

Remark 2.2. *The harmonic progression of* (2.3) *provokes a coupling of seat colours inspired by the theory of record values. Let* $\{U_s : 1 \le s \le n\}$ *be independent random variables with the uniform distribution on* $[0, 1]$. *The integer* s *is called a* record value *if* $U_s > U_r$ *for* $r < s$. *It is standard*[5] *that the events* $V_s = \{s \text{ is a record value}\}$, *for* $1 \le s \le n$, *are independent with* $\mathbb{P}(V_s) = 1/s$. *The seats may be coloured as follows. Seat* $m \in S$ *is coloured red if and only if* $n-m+2$ *is a record value.*

[5]See, for example, [1, 2, Prob. 7.11.36].

Remark 2.3. *What happens in the limit as $n \to \infty$, and in particular how does Figure 2.1 evolve when scaled by the factor $1/n$? Let S_i be the seat occupied by passenger i, with the convention that the seat labelled 1 is relabelled $n+1$ (see Figure 2.1). Note that $S_i \geq i$ for all i, and $S_i = i$ if and only if $i \neq 1$ and i is correctly seated. The displacements $D_i := S_i - i$ are a measure of the degree of disorder in the plane. It turns out that the normalised vector $D = (D_1/n, D_2/n, \ldots, D_n/n)$, when re-ordered in decreasing order (so that the first entry is the maximum displacement), has a limiting distribution in the limit as $n \to \infty$. The limit is the so-called Poisson–Dirichlet distribution, which crops up in a variety of settings including probabilistic number theory, the theory of random permutations, and population genetics*[6]*. The key reason for this is illustrated in Figure 2.1: S_1 is uniformly distributed on the set of seats; passenger S_1 sits in a seat labelled S_2 that is uniform on the seats to the right of S_1; and so on.*

Much mathematics proceeds by identification and solution of a sequence of smaller problems, which are obtained from the original by a process of distillation of intrinsic difficulties. In this way, even teasers may play significant roles in solving challenging problems. The lost boarding pass problem (more specifically, the argument of Solutions 1 and 2) has contributed to work by Johan Wästlund [10] on the random assignment problem and the so-called Buck–Chan–Robbins urn process.

3. Many lost boarding passes

A more general version of Questions 1 and 2 has been studied by Henze and Last [8], and is illustrated in Figure 3.1.

Question 3. *Let $k \geq 1$. A plane has $n \geq k+1$ seats, and is fully booked. A line of n people board one by one. Each has an assigned seat, but the first k passengers have lost their boarding passes and sit in seats chosen at random from those available. Subsequent passengers sit in their assigned seats if free, and in randomly chosen empty seats if not. Let $k+1 \leq m \leq n$. What is the probability that passenger m finds seat m already occupied?*

The solution using symmetry follows the route of Solution 2.

Solution 3. *Let $k+1 \leq m \leq n$. When passenger m chooses a seat, the seats $k+1, k+2, \ldots, m-1$ are already taken (since seat $r \in \{k+1, k+2, \ldots, m-1\}$, if free, would have been claimed earlier by*

[6] See Terry Tao's blog [9].

passenger r), and k further seats also. By symmetry, these further seats are equally likely to be any k-subset of the $n-m+k+1$ seats labelled $1, 2, \dots, k, m, m+1, \dots, n$. Therefore, seat m is occupied with probability $k/(n-m+k+1)$.

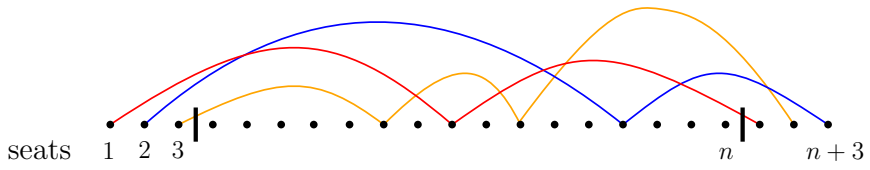

FIGURE 3.1. With the first 3 boarding passes lost, each of the first 3 passengers initiates a chain of displacements.

In particular, the probability that passenger n sits in the assigned seat is

$$1 - \frac{k}{k+1} = \frac{1}{k+1}.$$

We turn finally to independence. As in Section 2, let D_m be the event that passenger m finds seat m already occupied.

Theorem 3.1. *The events $D_{k+1}, D_{k+2}, \dots, D_n$ are independent.*

Proof. This follows that of Theorem 2.1. For example, by the same argument, (2.2) becomes

$$\begin{aligned}\mathbb{P}(D_{m_1} \cap D_{m_2}) &= \frac{k}{n-m_1+k+1} \cdot \frac{k}{n-m_2+k+1} \\ &= \mathbb{P}(D_{m_1})\mathbb{P}(D_{m_2}),\end{aligned}$$

and similarly for intersections of three or more of the events D_m. □

The overall seating program is now as follows, using the language of the end of Section 2. We have $S = \{k+1, k+2, \dots, n, n+1, \dots, n+k\}$, and seat $m \in S$ is declared *red* with probability

$$p_m = \begin{cases} \dfrac{k}{n-m+k+1} & \text{if } m \le n, \\ 1 & \text{if } m > n, \end{cases}$$

and *black* otherwise, different seats being coloured independently of one another. There are k distinct shades $\rho_1, \rho_2, \dots, \rho_k$ of red. Each red seat m satisfying $m \le n$ is shaded with a uniformly random shade, chosen independently of the shades of other red seats. As for the seats labelled $n+1, n+2, \dots, n+k$, for each i exactly one of these is shaded ρ_i, with the distribution of shades being given by a random permutation of the labels.

The passengers may now be seated according to the following algorithm. Passenger 1 sits in the earliest seat shaded ρ_1; the displaced passenger sits in the next seat shaded ρ_1, and so on. The process is repeated for passengers $2, 3, \ldots, k$, and finally the black seats are occupied by their correct passengers.

Remark 3.2. *In Remark 2.3 we identified the limit of the seating process as $n \to \infty$, with displacements scaled by $1/n$. With $k \geq 1$ lost boarding passes, the diagram of Figure 3.1 converges, when rescaled, to k independent copies of the corresponding limit of Figure 2.1.*

One may also consider the more general situation when the lost boarding passes may not be consecutive. The associated probabilities can be calculated using the methods and conclusions given above.

Acknowledgement

We are grateful to David Bell for having drawn our attention to Question 1 in 2000 or earlier. We suspect he was then aware of the quick argument of Solution 1.

References

[1] G. R. Grimmett and D. R. Stirzaker. *One Thousand Exercises in Probability*. Oxford University Press, Oxford, 4rd edition, 2020.
[2] G. R. Grimmett and D. R. Stirzaker. *Probability and Random Processes*. Oxford University Press, Oxford, 4th edition, 2020.
[3] Mathematics Stack Exchange. https://math.stackexchange.com/.
[4] P. Winkler. *Mathematical Puzzles*. A. K. Peters, Wellesley MA, 2004.
[5] B. Bollobás. *The Art of Mathematics*. Cambridge University Press, Cambridge, 2006.
[6] D. Roaf. Letter to the editor. *Mathematical Spectrum*, 43:89, 2010/2011.
[7] Y. Nigussie. Finding your seat versus tossing a coin. *The American Mathematical Monthly*, 121:545, 2014.
[8] N. Henze and G. Last. Absent-minded passengers. *The American Mathematical Monthly*, 126:867–875, 2019.
[9] T. Tao. The Poisson–Dirichlet process, and large prime factors of a random number. 2013. https://tinyurl.com/sz2uur7.
[10] J. Wästlund. The blind passenger and the assignment problem. *Combinatorics, Probability and Computing*, 20:467–480, 2011.

Statistical Laboratory, Centre for Mathematical Sciences, Cambridge University, Wilberforce Road, Cambridge CB3 0WB, UK
Email address: g.r.grimmett@statslab.cam.ac.uk
URL: http://www.statslab.cam.ac.uk/~grg/

St John's College, Oxford OX1 3JP, UK
Email address: david.stirzaker@sjc.ox.ac.uk
URL: https://www.sjc.ox.ac.uk/discover/people/professor-david-stirzaker/