Probability Taster Lecture

Peter Neal

University of Nottingham

11 May 2021

Overview

This lecture will study the concepts of conditional probability and expectation through a worked example.

Example.

The 2021 Panini Premier League Football Sticker album contains 642 distinct stickers.

Assuming that I collect stickers one at a time and each sticker is equally likely to be a copy of any one of the 642 distinct stickers:

1. How many stickers do I need to collect before I obtain a duplicate sticker?
2. How many stickers do I need to collect to complete the album?

The first question is known as the **birthday problem** and the second question is known as the **coupon collector problem**.

Conditional Probability

Suppose that we have an idea of the probability of some event, but then are given some additional piece of information (a *condition*). We may be able to adjust the probability of the event given this condition.

- Suppose we think at the start of the match there is a 60% chance that Manchester City will beat Chelsea.
- At half time we are told that Chelsea are winning 4-1.
- We would adjust the probability of Manchester City winning downwards!

The conditional probability of event A, given that event B has occurred, is denoted by $\mathbb{P}(A \mid B)$ and is defined as

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \text{ and } B)}{\mathbb{P}(B)}, \qquad \mathbb{P}(B) > 0.$$

When do I get the first duplicate?

What is the probability that the first 30 stickers are all different?

When do I get the first duplicate?

Let

$$A = \{\text{First } k \text{ stickers are distinct}\}$$

and

$$B = \{\text{First } k - 1 \text{ stickers are distinct}\}.$$

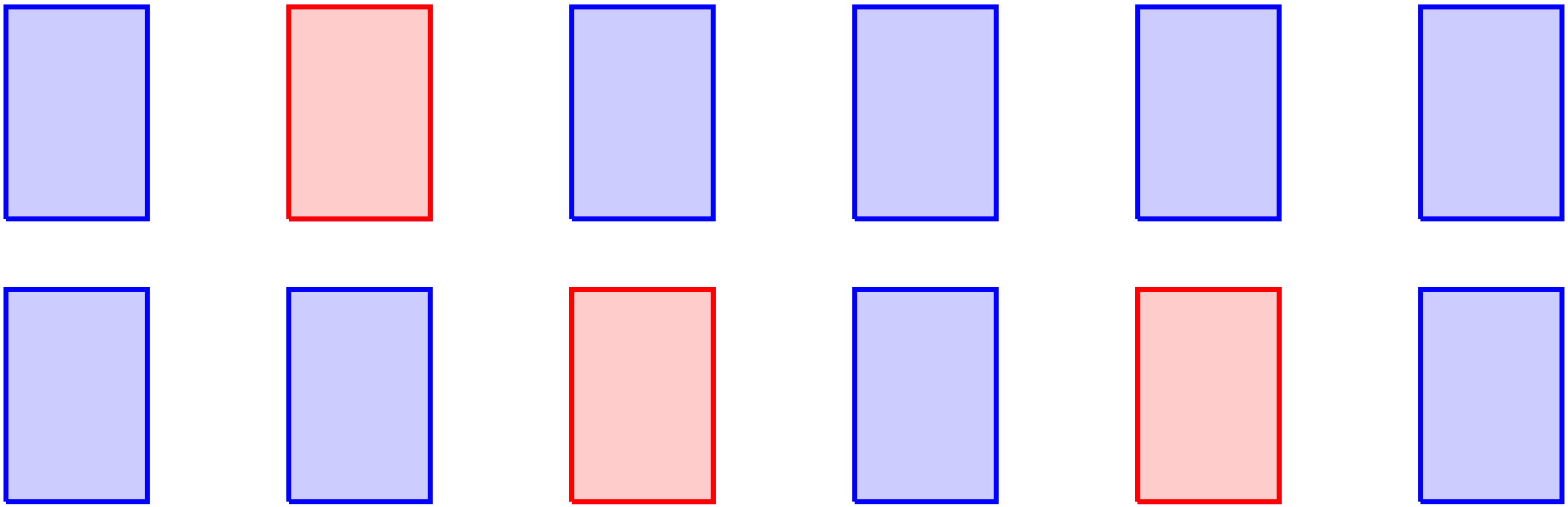
Note that A implies B and this means

$$\mathbb{P}(A \text{ and } B) = \mathbb{P}(A).$$

Therefore we can use conditional probability to write

$$\begin{aligned} \mathbb{P}(A) &= \mathbb{P}(A \text{ and } B) \\ &= \mathbb{P}(A \mid B)\mathbb{P}(B). \end{aligned}$$

$\mathbb{P}(A \mid B)$



Note that

$$\mathbb{P}(A \mid B) = \frac{642 - (k-1)}{642}.$$

Therefore

$$\begin{aligned} \mathbb{P}(A) &= \mathbb{P}(A \mid B)\mathbb{P}(B) \\ &= \frac{642 - (k-1)}{642} \times \frac{642 - (k-2)}{642} \times \ldots \times \frac{642 - 1}{642}. \end{aligned}$$

When do I get the first duplicate?

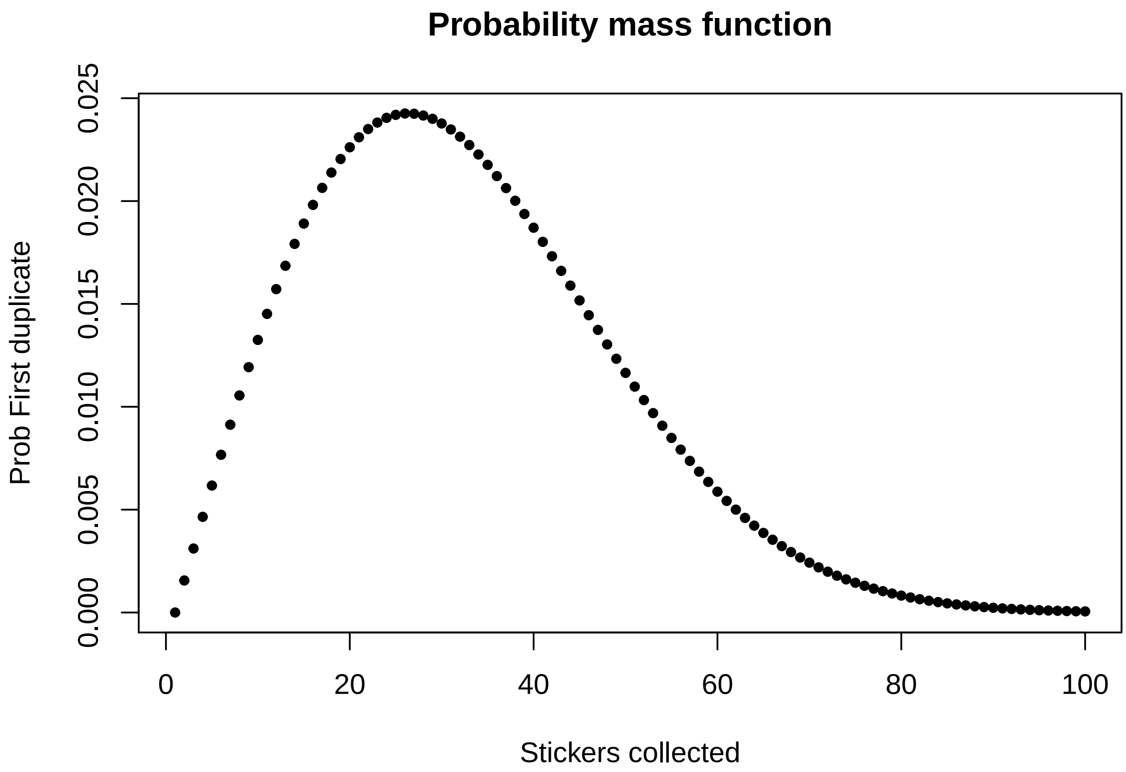


Figure: Probability of first duplicate at k^{th} sticker.

When do I get the first duplicate?

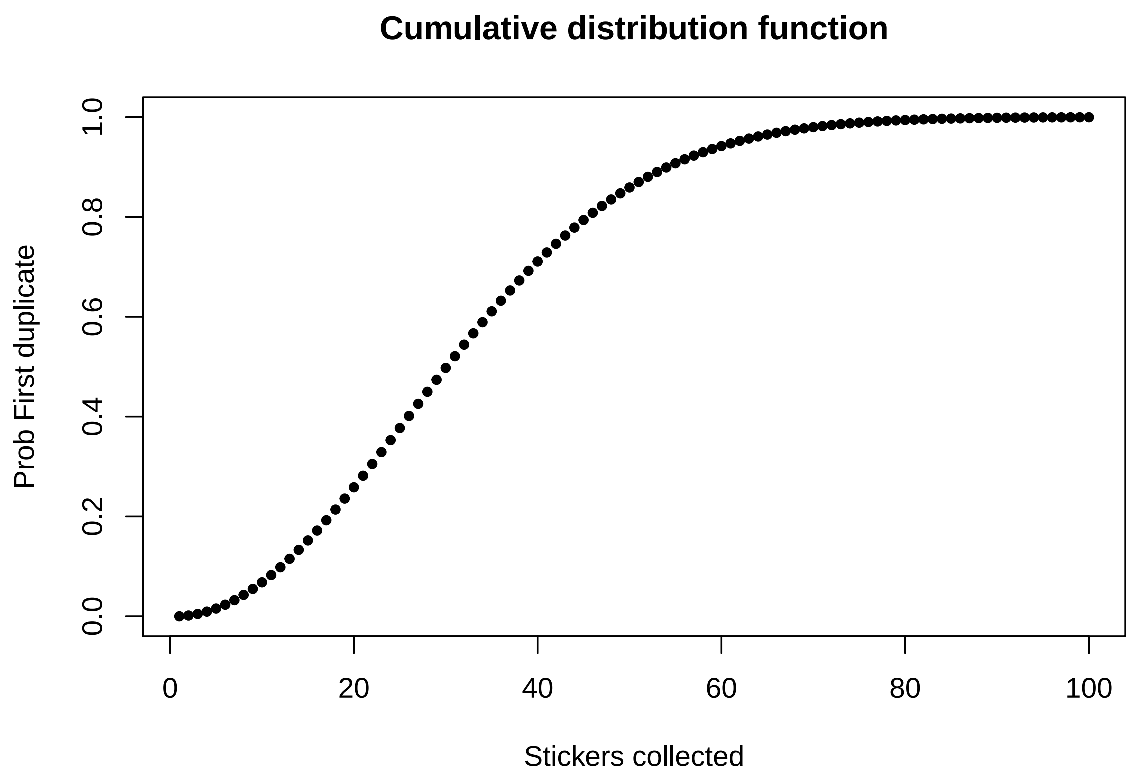


Figure: Probability of first duplicate by the k^{th} sticker.

Birthday Problem and Epidemics

This question is known as the **birthday problem**.

How likely in a group of N individuals is it that two or more people share a birthday?

For a class of 30 students there is a greater than 70% chance!

The question of when a match occurs plays a key role in modelling infectious diseases.

In the early stages of a disease (Covid-19) most infectious contacts are with susceptibles and we can approximate the epidemic process by a (simpler) birth-death process with infection = birth and removal = death.

This is effective until we start seeing individuals contacted with the disease for a second time.

How many stickers do I need to collect to complete the album?

There are 642 distinct stickers to collect.

How many stickers, on average, will I need to buy to complete the album? [Nearest 100.]

The first step to calculating the number of stickers required is to answer the following question:

Suppose that I have K distinct stickers. How many **more** stickers do I need to collect to have $K + 1$?

Geometric random variable

Suppose that I have K distinct stickers. How many **more** stickers do I need to collect to have $K + 1$?

We can define a random variable X for the number of additional stickers I need to collect.

Consider the next sticker I receive. Either:

1. The sticker is new to me. Success
2. It is a sticker I already have. Failure

The probability I am successful at each attempt is

$$p = \frac{642 - K}{642}.$$

Hence, $X = n$ if we have $n - 1$ failures followed by a success and

$$\mathbb{P}(X = n) = (1 - p)^{n-1}\, p.$$

Geometric random variable

What is the expected (average) value of X?

The expected value of a random variable $\mathbb{E}[X]$ is given by

$$\begin{aligned}\mathbb{E}[X] &= \sum_{n=1}^{\infty} n\mathbb{P}(X = n) \\ &= 1 \times p + 2 \times (1-p)p + \ldots = p \sum_{n=1}^{\infty} n(1-p)^{n-1}.\end{aligned}$$

Geometric random variable

What is $\sum_{n=1}^{\infty} n(1-p)^{n-1}$?

A geometric series with $0 \leq z < 1$ satisfies

$$\sum_{n=0}^{\infty} z^n = 1 + z + z^2 + \ldots = \frac{1}{1-z}.$$

Differentiating with respect to z gives

$$\sum_{n=0}^{\infty} nz^{n-1} = 0 + 1 + 2z + \ldots = \frac{1}{(1-z)^2}.$$

Hence, if we take $z = 1 - p$,

$$\mathbb{E}[X] = p \sum_{n=1}^{\infty} n(1-p)^{n-1} = \frac{p}{p^2} = \frac{1}{p}.$$

Collecting Stickers

How many stickers do we need in total?

Suppose that we consider the random variable how many stickers we need to collect after we have the $(k-1)^{st}$ to get the k^{th} distinct sticker.

This will be a geometric random variable:

$$Y_k \sim \text{Geom}\left(\frac{642-(k-1)}{642}\right) = \text{Geom}\left(\frac{643-k}{642}\right).$$

Then the total number of stickers required is:

$$Y_1 + Y_2 + \ldots + Y_{642},$$

and the expected number is

$$\frac{642}{643-1} + \frac{642}{643-2} + \ldots + \frac{642}{643-642}.$$

Collecting Stickers

You will require, on average, another 642 stickers after collecting the penultimate sticker to complete the set!

Note that

$$\frac{642}{643-1} + \frac{642}{643-2} + \ldots + \frac{642}{643-K}$$

is the expected number of stickers required to get K distinct stickers.

Collecting Stickers

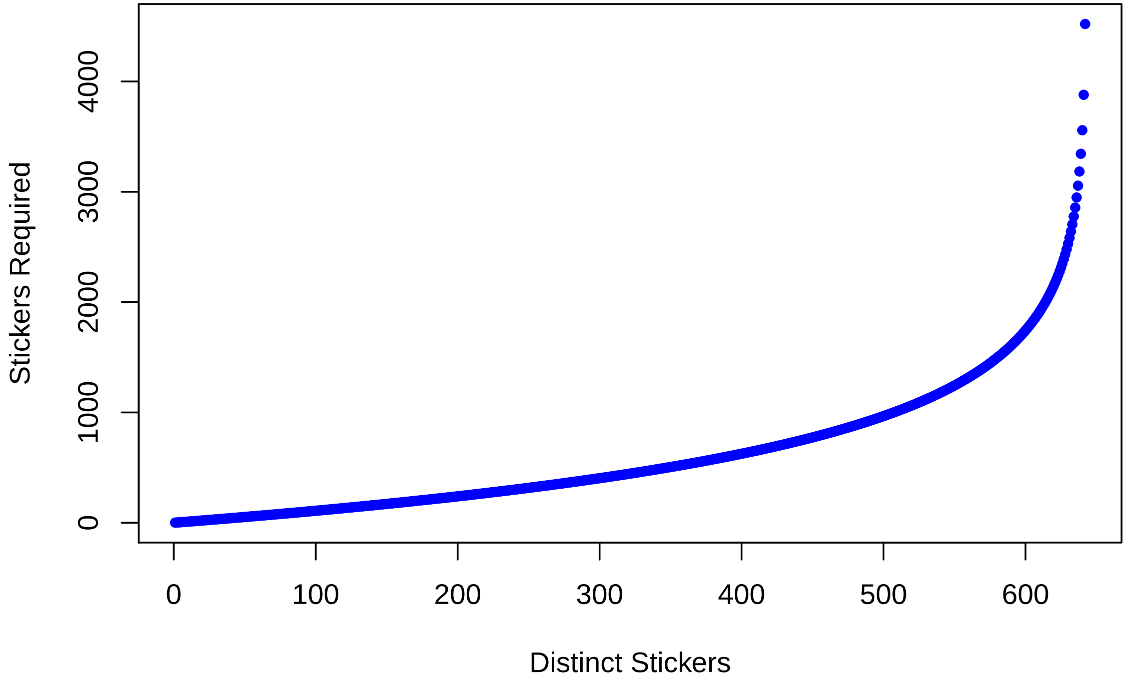


Figure: Expected number of stickers required to collect distinct stickers.

Collecting Stickers

Distinct Stickers	Average no. stickers required
100	108.6
200	239.4
300	403.9
400	625.5
500	966.9
600	1743.6
622	2211.6
642	4521.3

On average more stickers are required to collect the final 20 than the first 622 stickers!

How much will it cost?

Each sticker costs 14 p (70 p for a pack of 5 stickers).

Therefore, on average, it will cost about £633 to complete the album.

For £200, you will obtain, on average, approximately 573 distinct stickers.

How many stickers?

We have focussed on the average number but it is interesting to look at the distribution.

How likely is it we will complete the album for £500?

How likely is it that it will cost over £1000 to complete the album?

Hard to work with the actual distribution but we can use simulations (getting the computer to repeat the process of collecting the stickers many times) to estimate the probabilities.

Answers: (Based on 100,000 simulations.)
How likely is it we will complete the album for £500? ≈ 0.085

How likely is it that it will cost over £1000 to complete the album? ≈ 0.009

Collecting Stickers

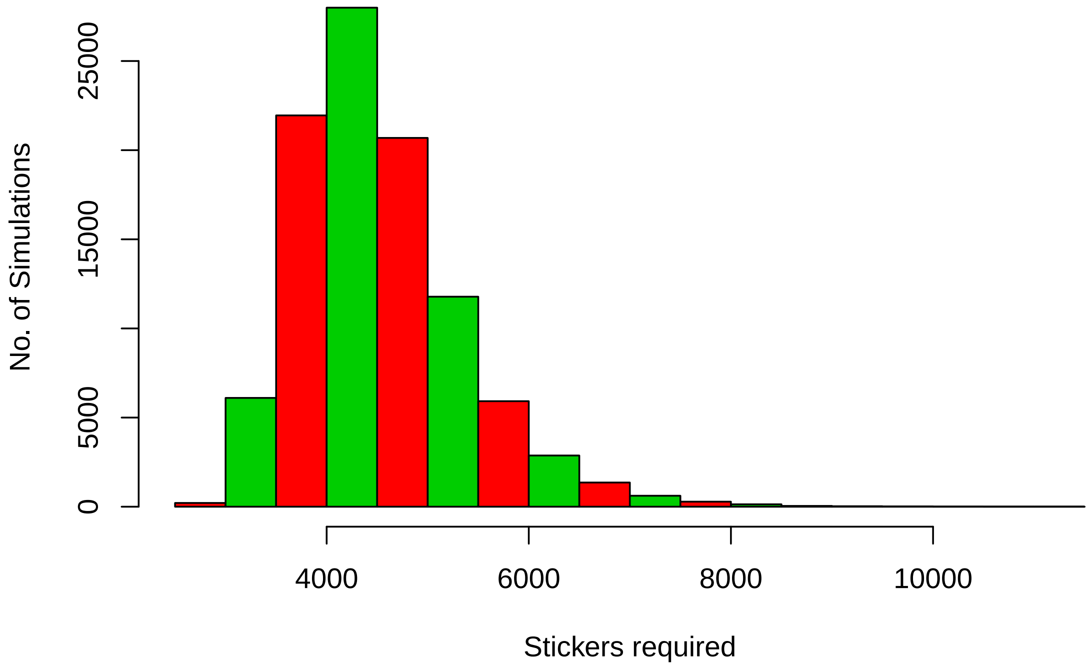


Figure: Histogram of the distribution of the number of stickers required to complete the album.

Conclusions

Introduction to the birthday and coupon collection problem.

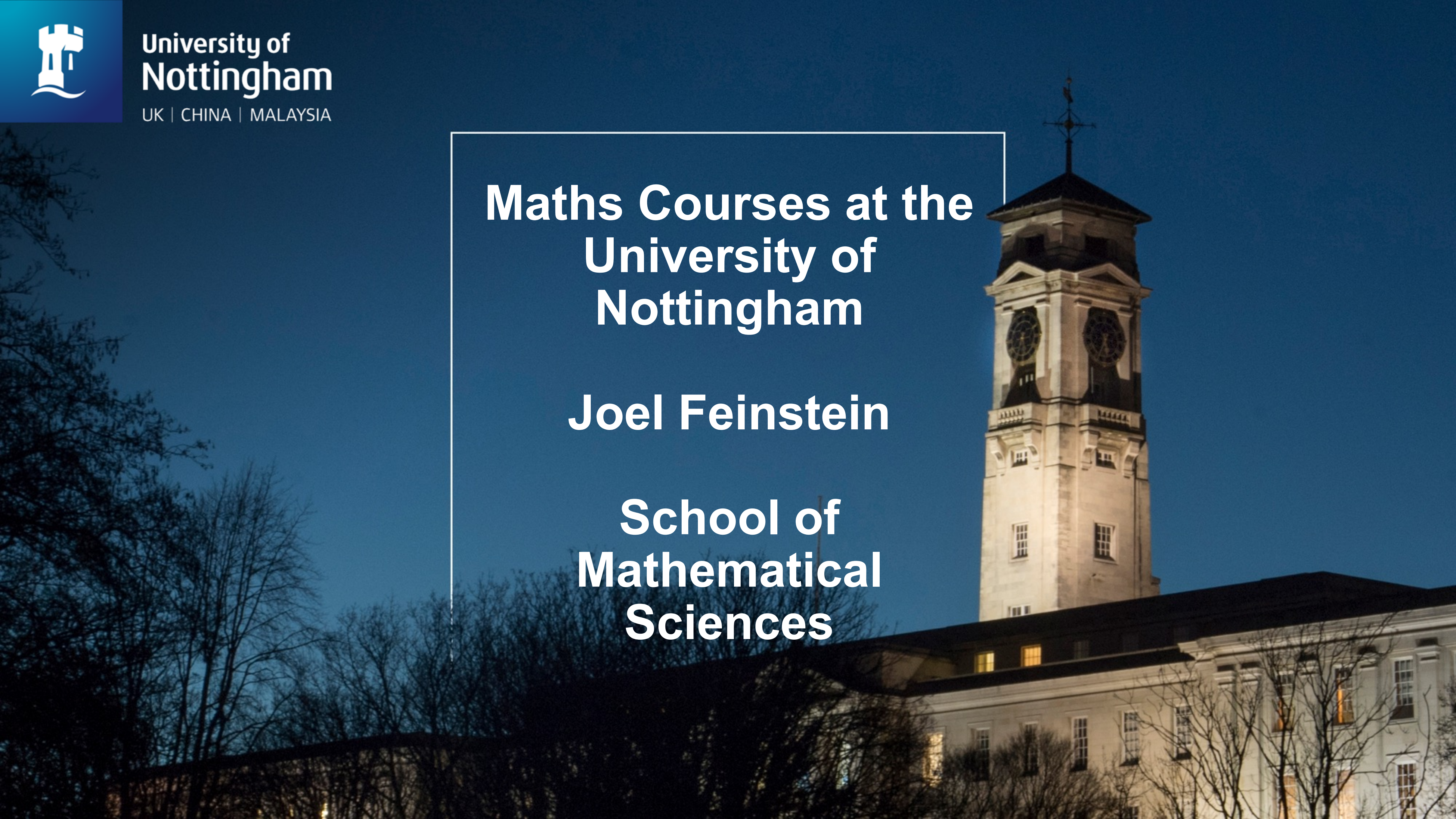
Concepts discussed:

- Conditional Probability.
- Expectation.
- Geometric Random Variables.

Extensions:

- When do we first get a sticker for the third time?
- How many stickers to collect 2 sets?
- Unequal probability of stickers.

University of Nottingham
UK | CHINA | MALAYSIA



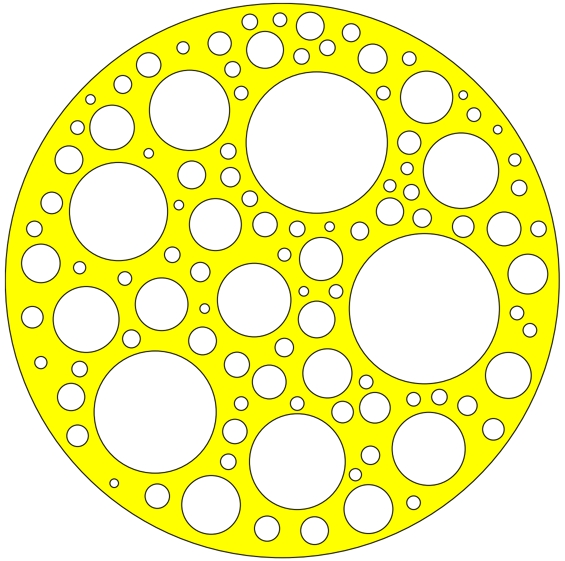
Maths Courses at the University of Nottingham

Joel Feinstein

School of Mathematical Sciences

About me – Joel Feinstein

- Associate Professor, Pure Mathematics
- Outreach Officer
- Teaching Support Officer
- I teach the first-year module **Foundations of Pure Mathematics**
- My research includes work on Swiss cheeses!



School of Mathematical Sciences

- Department of Mathematics formed in 1919
- School of Mathematical Sciences formed in 1998
- Moved to current, purpose-built, home in 2011
- Situated in a lovely campus with great facilities
- Over 70 academic staff

Maths Courses at Nottingham

Single-Subject Degrees

- **Mathematics BSc** (3 years)
- **Mathematics MMath** (4 years)
- **Mathematics (International Study) BSc** (4 years)
- **Mathematics with a Year in Industry BSc** (4 years)
- **Mathematics with a Year in Industry MMath** (5 years)
- **Statistics BSc** (3 years)

Maths Courses at Nottingham

Joint Degrees

- **Financial Mathematics BSc** (3 years)
  - with Nottingham University Business School
- **Mathematics and Economics BSc** (3 years)
  - with School of Economics
- **Mathematical Physics BSc/MSci** (3/4 years)
  - coordinated by School of Physics & Astronomy
- **Natural Sciences BSc/MSci** (3/4 years)
  - coordinated across schools involved
  - available with a year abroad

Careers with Mathematics

The most popular employment sectors nationally for maths graduates are*:

- Business, HR and finance professionals (42%)

  e.g., Consultant, Actuarial Graduate, Analyst, Strategic Consultant, Accountant

- IT professionals (12%)

  e.g., Software Engineer, Data Analyst, Cyber Security Associate, Technology Analyst

- Education professionals (9%)

  e.g., Teacher of Mathematics, Teaching Assistant

*Source: *What do graduates do?* (HECSU 2018)

Top four employers for our graduates:

- Deloitte
- PwC
- Ernst & Young
- KPMG

Some useful links

University of Nottingham, School of Mathematical Sciences and our maths courses:

https://tinyurl.com/mathsuon

https://tinyurl.com/mathscourseuon

Complete sets of videos for the first-year module **Foundations of Pure Mathematics**:

https://tinyurl.com/uonfpm

University of Nottingham
UK | CHINA | MALAYSIA

Any questions?

Please give us feedback on this session using the link in the Q&A chat

Future Maths Taster Sessions: https://tinyurl.com/uonmathstaster