Prof. Alex Rogers
Department of Computer Science
University of Oxford
13th May 2023

Scientific Report Writing and Project Workshop

Probability Distributions

Counting Events

We often have settings where we count events and must estimate the underlying rate of some process.

For example, what is the rate that people visit a website if 10 people visit in one hour.

The number of events is an integer and can not be represented with a continuous distribution.

Poisson Distribution

The Poisson distribution is a positive-only discrete probability distribution described by a rate parameter:

$$P(n|\mu) = \frac{\mu^n e^{-\mu}}{n!}$$

https://en.wikipedia.org/wiki/Poisson_distribution

https://docs.pymc.io/api/distributions/discrete.html

Millikan Oil Drop Experiment

Model Building Methodology

1. Understand the setting and the scientific question being asked of the data.

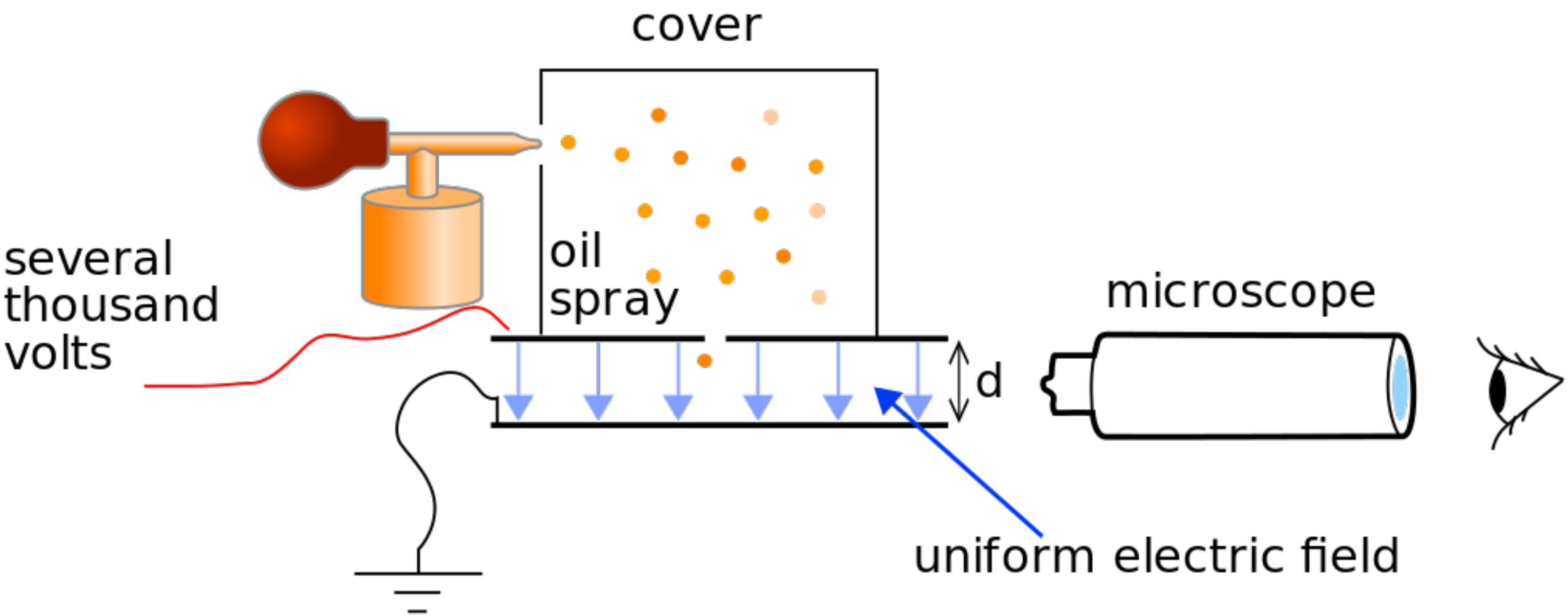
2. What are the sources of noise and imprecision in the process being considered.
    - What likelihood function is appropriate?
        - Are we making error-prone continuous measurements or are our observations discrete?
        - Do we know anything about the expected accuracy of the measurements?

3. Think about what prior information is available.
    - How should that be represented within the model?
        - What prior probability distributions should we use?
        - Are random variable discrete or continuous?

Millikan Oil Drop Experiment

https://en.wikipedia.org/wiki/Oil_drop_experiment

Millikan Oil Drop Experiment



https://en.wikipedia.org/wiki/Oil_drop_experiment

Millikan Oil Drop Experiment

1. Make noisy measurements of the charge on each oil drop.
    - The range of the measurement that we can make is around 0 to 10 x 1e-19 coulombs.
    - Our measurements should be accurate to +/- 0.1 x 1e-19 coulombs if we are careful.

2. We think each oil drop will pick up excess charge equivalent to just a few electrons.
    - The charge is due to an integer excess of electrons.

3. Previous experiments suggest that the charge on an electron is between 1 and 2 x 1e-19 coulombs.
    - What do our experimental results suggest for this value?
    - How accurate is our experimental setup?

oil_drop.ipynb

Describing Models

Describing Likelihood Models and Priors

When we define variables in PyMC3 we are really defining random variables. The random variable has a distribution from which its actual value will be derived at some future time.

We can use the same notation of random variables to describe the likelihoods and the priors in our model.

$$x_i \sim \text{Beta}(3, 5)$$

We typically separate the mathematical description of the model from an actual implementation in code. However, probabilistic programming languages, such as PyMC, try to make the two very similar.

We typically use the notation of discrete mathematics to describe experiments and algorithms formally in computer science.

Next Time

Project Presentations