

Section #3: Random Variables

1 Gender Composition of Sections

A massive online Stanford class has sections with 10 students each. Each student in our population has a 50% chance of identifying as female, 47% chance of identifying as male and 3% chance of identifying as non-binary. Even though students are assigned randomly to sections, a few sections end up having a very uneven distribution just by chance. You should assume that the population of students is so large that the percentages of students who identify as male / female / non-binary are unchanged, even if you select students without replacement.

- Define a random variable for the number of people in a section who identify as female.
- What is the expectation and standard deviation of number of students who identify as female in a single section?
- Write an expression for the exact probability that a section is skewed. We defined skewed to be that the section has 0, 1, 9 or 10 people who identify as female.
- The course has 1,200 sections. Approximate the probability that at 5 or more sections will be skewed.

2 Better Evaluation of Eye Disease

When a patient has eye inflammation, eye doctors "grade" the inflammation. When "grading" inflammation they randomly look at a single 1 millimeter by 1 millimeter square in the patient's eye and count how many "cells" they see.

There is uncertainty in these counts. If the true average number of cells for a given patient's eye is 6, the doctor could get a different count (say 4, or 5, or 7) just by chance. As of 2021, modern eye medicine does not have a sense of uncertainty for their inflammation grades! In this problem we are going to change that. At the same time we are going to learn about poisson distributions over space.

- Explain, as if teaching, why the number of cells observed in a 1x1 square is governed by a poisson process. Make sure to explain how a binomial distribution could approximate the count of cells. Explain what λ means in this context. Note: for a given person's eye, the presence of a cell in a location is independent of the presence of a cell in another location.
- For a given patient the true average rate of cells is 5 cells per 1x1 sample. What is the probability that in a single 1x1 sample the doctor counts 4 cells?

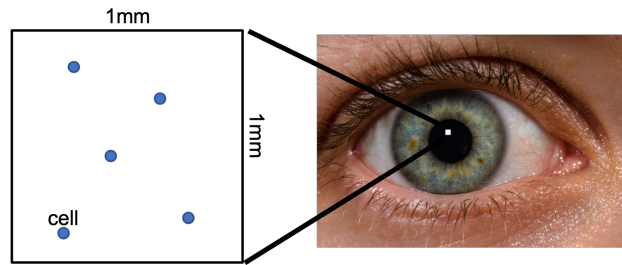


Figure 1: A $1 \times 1\text{mm}$ sample used for inflammation grading. Inflammation is graded by counting cells in a randomly chosen 1mm by 1mm square. This sample has 5 cells.

3 Website Visits

You have a website where only one visitor can be on the site at a time, but there is an infinite queue of visitors, so that immediately after a visitor leaves, a new visitor will come onto the website. On average, visitors leave your website after 5 minutes. Assume that the length of stay is exponentially distributed. What is the probability that a user stays more than 10 minutes, if we calculate this probability:

- Using the random variable X , defined as the length of stay of the user?
- Using the random variable Y , defined as the number of users who leave your website over a 10-minute interval?

4 Continuous Random Variables

Let X be a continuous random variable with the following probability density function:

$$f_X(x) = \begin{cases} c(e^{x-1} + e^{-x}) & \text{if } 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

- Find the value of c that makes f_X a valid probability distribution.
- What is $P(X < 0.75)$? What is $P(X < x)$?