

Math 4397/6397, Fall 2009
Problem Set 5, due Thursday, Oct 1

Problem 1. Suppose we study the number of times a student sits in a classroom with a TB infected, coughing neighbor until the student contracts the disease. Assume that each classroom encounter is an independent Bernoulli trial with probability p that the student becomes infected. This leads to the so-called geometric distribution $P(\text{Person is infected on encounter } x) = p(1 - p)^{x-1}$ for $x = 1, 2, \dots$

- Suppose that one subject's number of classroom encounters until infection is recorded, say x . Derive the maximum likelihood estimate of p given the observed number x .
- Suppose that the subject's value was $x = 2$. Use R to plot the likelihood function for p and interpret the result.
- Suppose that it is often assumed that the probability of transmission, p , is .01. We think that it is perhaps strange to have a subject get infected after only 2 encounters if the probability of transmission is really on 1%. According to the geometric mass function, what is the probability of a person getting infected in 3 or fewer encounters if p truly is .01?
- Suppose that we follow n subjects (in different classrooms, each with a coughing, TB infected neighbor) and record the number of classroom encounters until infection (assume all subjects became infected) x_1, \dots, x_n . Derive the maximum likelihood estimate of p . State your assumptions.
- Suppose that we record values $x_1 = 3$, $x_2 = 5$, $x_3 = 2$. Plot and interpret the likelihood for p .

Problem 2. In a study of aquaporins, 6 frog eggs received a protein treatment. If the treatment of the protein is effective, the frog eggs implode. The experiment results in 5 frog eggs imploding. Historically, ten percent of eggs implode without the treatment. Assuming that the results for each egg are independent and identically distributed:

- What is the probability of getting 5 or more eggs imploding in this experiment if the true probability of implosion is 10%? Interpret this number.
- What is the maximum likelihood estimate for the probability of implosion?
- Plot and interpret the likelihood for the probability of implosion.

Problem 3. Suppose that IQs in a particular population are normally distributed with a mean of 110 and a standard deviation of 10.

- What is the probability that a randomly selected person from this population has an IQ between 95 and 115?
- What is the 65th percentile from this distribution?

- c. Suppose that 5 people are sampled from this distribution. What is the probability of 4 (80%) or more having IQs above 130?
- d. Suppose that 500 people are sampled from this distribution. What is the probability of 400 (80%) or more having IQs above 130?
- e. Consider the average of 100 people drawn from this distribution. What is the probability that this sample mean is larger than 112.5?

Problem 4. Note that R's function `rexp` generates random exponential variables. The exponential distribution with rate 1 (the default) has a theoretical mean of 1 and variance of 1.

- a. Sample 1,000 observations from this distribution. Take the sample mean and sample variance. What numbers should these estimate and why?
- b. Retain the same 1,000 observations from part a. Plot the sequential sample means by observation number. Explain the resulting plot.

Hint: If x is a vector containing the simulated uniforms, then the code

```
y <- cumsum(x) / (1 : length(x))
```

will create a vector of the sequential sample means.

- c. Plot a histogram of the 1,000 numbers. Does it look like an exponential density?
- d. Now sample 1,000 *sample means* from this distribution, each comprised of 100 observations. What numbers should the average and variance of these 1,000 numbers be equal to and why?
- e. Plot a histogram of the 1,000 sample means. What does it look like and why?