MATH 308 Assignment F:
Sample Final Problem

Nakul Joshi

May 2, 2014

Benford's Law, also known as the *First-Digit Law*, describes the distribution of the first digit of numbers in many real-life sources of data. In base 4, the law states that the leading digit $d \in \{1, 2, 3\}$ is described by the probability mass function:

$$P(d) = \log_4 \left(1 + \frac{1}{d}\right)$$

A financial auditor is trying to apply Benford's Law to verify the accounts of a client. She knows that if the numbers on the account report deviate significantly from the Benford's Law distribution, the numbers are likely to be fudged. So, she decides to gather a random sample of n numbers from the report, and test them.

a)

Describe a test statistic she can use, remembering to state the null and alternative hypotheses about the statistic.
Solution. She can use a chi-squared test, by tabulating the frequencies of occurences of each of $1, 2, 3$ as the first digit. Using the pdf, she can also obtain expected values for the same frequencies. The chi-squared statistic t thus obtained is the required test statistc. The null hypothesis is that the data follows Benford's law, i.e. $t \sim \chi^2_2$. The alternative is that it does not.

b)

Assuming that she is testing at the significance level α, what is the critical region?
Solution. The critical region c is given by

$$c = [0, 2 - a) \cap (2 + a, \text{inf})$$

, where $P(2 - a \leq t \leq 2 + a) = 1 - \alpha$. a is obtained from a table of chi-square values (2 is the mean of the chi-square on 2 degrees of freedom).

c)

Write a program, using either `R` or pseudocode, to perform the test on a dataset.
Solution.

1. Read list of numbers in base 3
2. Extract observed values by reading first digits
3. Extract expected values using pmf.
4. Perform chi-squared test
5. Compare p-value from chi-squared test against α.