Exam Statistics and Probability

November 9, 2012, 14:00-16:00.

Include all the `R` code that you used and attach any figures in .pdf format. It should be clear which code belongs to which problem.

1. Generate a population of size $N = 300$ from the Geometric distribution with parameter $p = 0.7$ by using the code shown below

```
set.seed(12)
pop=rgeom(300,0.7)
```

   (a) From the population `pop` use the `R` command `sample` to draw with replacement 30 samples of size $n = 15$. For each sample, compute the sample mean and the sample variance.

   (b) Make a histogram of the sample means.

   (c) Estimate the population mean on the basis of the 30 sample means.

   (d) Use the 30 samples to determine the variance of your estimate of the mean.

2. Suppose that $X_1, X_2, \ldots, X_n$ are i.i.d. sample from a distribution

$$f(x, \theta) = \frac{\theta(-\log(\theta))^x}{x!}$$

where $0 < \theta < 1$.

   (a) Demonstrate mathematically (with pen and paper) that the maximum likelihood estimator of θ is given by

$$\hat{\theta} = e^{-\bar{X}}$$

Now, suppose we observe an independent sample of size n = 50

```
x=c(1, 1, 2, 3, 1, 3, 4, 2, 2, 0, 1, 0, 2, 1, 2, 1, 2, 5,
    1, 2, 4, 1, 2, 0, 1, 1, 0, 1, 3, 1, 1, 2, 1, 0, 3, 2,
    3, 0, 2, 1, 3, 2, 3, 2, 2, 3, 0, 1, 2, 2 )
```

   (b) Plot the loglikelihood function.

   (c) Maximize the loglikelihood using `R`.

3. In a clinical study of an allergy drug, 108 of the 203 subjects reported experiencing significant relief from their symptoms.

   (a) Construct a 95% confidence interval for the probability that someone who uses the drug experiences relief from the drug.

   (b) The manufacturer of the drug wants to prove that his drug is effective in more than 30% of all patients. Formulate the appropriate null hypothesis and alternative hypothesis.

   (c) Perform the test at the 5% level of significance.

   (d) Plot the power function of the test you used at 3 (c).

4. Suppose we have a sample of size $n = 5$ from the normal distribution with mean $\mu = 0$ and unknown standard deviation σ. Suppose we decide to test

$$H_0 : \sigma = 2 \quad \text{versus} \quad A : \sigma > 2$$

   by computing the sample standard deviation S, and rejecting H_0 if S is greater than 2.

   (a) Perform 10000 simulations in `R` to determine the level of significance of this test.

   (b) Modify your `R` script to determine the power of the test against the alternative $\sigma = 3$.