Exercise 4: Statistical inference (I)

Benjamin Smith

June 24, 2026

Part 1: Probability distributions

1. A contestant on a game show needs to answer 10 questions correctly to win the jackpot. However, if they get 4 incorrect answers, they are kicked off the show. Suppose one contestant consistently has a 80% chance of correctly responding to any question.
   (a) What is the probability distribution?
   (b) What is the probability that she will correctly answer 10 questions before 4 incorrect responses?
   (c) Write out the R code to calculate (b).
2. A small town's police department issues 5 speeding tickets per month on average.
   (a) Using a Poisson random variable, what is the likelihood that the police department issues 3 or fewer tickets in one month?
   (b) What is the probability that 10 days or fewer elapse between two tickets being issued?
   (c) Write out the R code to calculate (a), (b).

Part 2: Statistical inference

1. (AoS 6.6.2) Let $X_1, \ldots, X_n \sim \text{Uniform}(0, \theta)$ and let $\hat{\theta} = \max\{X_1, \ldots, X_n\}$. Find the bias, se and MSE of this estimator.
2. (AoS 6.6.3) Let $X_1, \ldots, X_n \sim \text{Uniform}(0, \theta)$ and let $\hat{\theta} = 2\bar{X}_n$. Find the bias, se and MSE of this estimator.
3. Let $X_1, \ldots, X_n \sim \text{Uniform}(0, 1)$. Let $Y_n = \bar{X}_n^2$. Find the limiting distribution of Y_n. (Hint: CLT)

Part 3: Newton-Raphson

The ABO-gene or ABO-locus is on chromosome 9. It has 3 alleles (antigens) (A, B, O) and it determines 4 blood type (A, B, AB, O).

```
 genotype   phenotype
  AA  AO       A
  BB  BO       B
    AB         AB
    OO         O

A, B  are dominant to O.
O     is recessive to A, B.
A, B  are co-dominant.
```

We have a large random sample obtained from Berlin (Bernstein 1925, Sham's book page 44):

- $n_A = 9123$ blood type A
- $n_B = 2987$ blood type B
- $n_{AB} = 1269$ blood type AB
- $n_O = 7725$ blood type O

For instance, $n_A = 9123 = n_{AA} + n_{AO}$: Among 9123 blood type A individuals, some have genotype AA and the others have genotype AO.

Our interest is to estimate the allele frequencies of alleles A, B, and O. i.e. $p =$ freq (allele A), $q =$ freq (allele B), $1 - p - q =$ freq (allele O).

1. Write out the log-likelihood $L(p, q)$.
2. Is there a closed-form solution of this log-likelihood function?
3. Formulate the problem as a missing data problem and use the Newton-Raphson algorithm to find the MLEs, $\hat{p}$ and $\hat{q}$, that maximize the log-likelihood, $\ln L(p, q)$.