Homework 1 (Part II) – 30 Points

Problem 1: Suppose that $X_1, X_2, \cdots, X_n$ are iid $N(\mu, \sigma^2)$. Then, the corresponding log-likelihood for estimating μ and σ is given by

$$\ell(\mu,\sigma) = -\frac{n}{2}\log 2\pi - n\log\sigma - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$$

a) **[4 pts]** Derive the observed information matrix for μ and σ.

b) **[2 pts]** Obtain the Fisher-information matrix for μ and σ.

c) **[3 pts]** Using the transformation theorem, provided in the lecture, obtain the Fisher-information matrix for μ and σ^2.

d) **[1 pts]** What are the standard errors for the maximum likelihood estimates of μ, σ, and σ^2?

Problem 2: Our aim in this problem is to use simulation to verify the result of the theorem stated in class about the asymptotic distribution of the maximum likelihood estimate, using the multinomial example given in the lecture.

Suppose that there are three possible outcomes in each trial of an experiment, with the first, second, and third outcomes having respective probabilities `p1 and p2`, and `p3 = 1 - p1 - p2`. Let `T1, T2, and T3` respectively denote the number of occurrences of each of the outcomes 1, 2, and 3 in n trials of the experiment. Then the random vector (`T1, T2, T3`) is said to have a trinomial distribution with parameters n, `p1, p2`, and `p3.`

(a) **[3 pts]** For the case where $n = 200, p_1 = \frac{1}{4}, p_2 = \frac{1}{4}, p_3 = \frac{1}{2}$, compute the theoretical values for the inverse of the Fisher Information matrix for this problem, using the formulas obtained in class, and obtain the standard errors for maximum likelihood estimates of p_1, p_2, and p_3.

(b) **[7 pts]** Write a general R function to simulate `n_sim` observations from a trinomial distribution with n trials and given probabilities p1, p2, and p3 (`n_sim` is the number of simulated values). The input variables for your function should be n_sim, n, p1, p2 (Note p3 = 1 – p2 – p1), and the output should be a matrix of size n_sim by 3 with each row representing an observation (t_1, t_2, t_3) from the trinomial, where t_i is the number observed, out of the n trials, for group $i = 1,2,3, respectively$ (Note: $\sum_{i=1}^{3} t_i = n$.) Call the simulated matrix T. For example, if $n = 200$ and $p_1 = \frac{1}{4}, p_2 = \frac{1}{4}, p_3 = \frac{1}{2}$, then a typical row of T (simulated values) may be $(t_1 = 57, t_2 = 48, t_3 = 95)$. Note: there are various ways to simulate the values, but you must use the `sample()` function in R for your simulation. Use your program to simulate n_sim = 1000 trinomial observation with $n = 200, p_1 = \frac{1}{4}, p_2 = \frac{1}{4}, p_3 = \frac{1}{2}$, and print the first five rows of T.

(c) **[2 pts]** Write another function that inputs a matrix T (1000 by 3) that you obtain from the function in (b), and outputs a matrix of maximum-likelihood estimates corresponding to each row of T; call the output M. For example, for $(t_1 = 57, t_2 = 48, t_3 = 95)$ the corresponding maximum likelihood estimates of (p_1, p_2, p_3) are $\left(\widehat{p_1} = \frac{57}{200}, \widehat{p_2} = \frac{48}{200}, \widehat{p_2} = \frac{95}{200}\right)$. Your output M should be a matrix of size 1000 by 3. Print the first five rows of M. [Note that each row of Mrepresents MLE corresponding to a sample that is simulated.]

(d) **[4 pts]** Obtain the covariance of your 1000 simulated values, using the R's `cov()` function. This approximates the variance-covariance matrix of the maximum likelihood estimates by simulation. Form a table that displays the following values in each column: The name of the parameter, theoretical values that you calculated for the standard errors in part (a), their corresponding simulated approximation obtained in part (c), and the absolute difference between the true and approximate values. Only list the diagonal and lower diagonal values (there are 6 values).

(f) **[4 pts]** Use R to draw a histogram of the values in each of the three columns of M. Explain how your graph relates to the theorem stated in class about the asymptotic distribution of MLE.