Problem Set 1 (two pages)

Statistics 24510-30040 (Winter Quarter 2019)

Due Tuesday, January 15 at the beginning of class.

Requirements

- Use standard $8\frac{1}{2}" \times 11"$ papers, typed or neatly handwritten, and stapled.
- Put your name and section number (24510 or 30040) at the top of the first page.
- Show your work. Provide detailed derivations and arguments in order to receive credit.
- When using R (or other software) to solve problems, use cut-and-paste to select only the relevant part of the output to be inserted in your writing. Attach your code or output as an appendix if necessary.
- You may discuss approaches with others. However the assignment should be devised and written by yourself completely.

Problem assignments

(This problem set is mostly a review, relevant to parts of chapters 8 and 9 in the text by Rice.)

1. (*MLE exercise in genetics*)

   We suppose that there are three possible observed traits for a specimen, namely A, B, and C, and according to a theory of dominance, they should occur with probabilities $\theta, \theta(1-\theta), (1-\theta)^2$. The parameter θ is unknown, but we have n independent randomly selected specimens available to estimate θ, and they give, respectively, counts X_1, X_2, and X_3 of traits A, B, and C, where $X_1 + X_2 + X_3 = n$.

   (a) Write the likelihood function for these data.

   (b) Find the Maximum likelihood estimator (MLE) of θ. (Again, show your work.)

   (c) Find the Fisher Information of this estimator.

   (d) Use the asymptotic properties of MLE (Fisher's Theorem) to find the approximate Mean squared error (as a function of θ) of this estimator.

2. (*Bayesian inference, R plot*)

   Consider the situation as described in Problem 1 in this assignment (with the same data), but now using Bayesian inference. Specifically, suppose that you are willing to assume that θ has an a priori $Beta(40, 40)$ distribution. Given $X_1 = 100, X_2 = 80$, and $X_3 = 20$,

   (a) Find the posterior expectation of θ;

   (b) Is the Bayesian estimator in part (a) of this problem the same as the corresponding Maximum likelihood estimator (derived in Problem 1), using the observed data?

(c) In the same plot, sketch the prior, posterior, and the likelihood as functions of θ using R. Label the axes and the curves clearly. Provide your code.

(Notes: In the plot, scale the likelihood function so that the area under the curve = 1.)

3. (*Binomial probability*)

Assume the random variable $X \sim Bin(n, p)$ is of binomial distribution. We are interested in the probability $P(X = k)$ for

(a) $n = 7, p = 0.3, k = 3.$

(b) $n = 40, p = 0.4, k = 11.$

(c) $n = 400, p = 0.0025, k = 2.$

In each case of (a), (b) and (c) above, give

i. the exact Binomial probability,

ii. an approximation using the normal approximation,

iii. an approximation based on the Poisson distribution.

In each case, comment on the accuracy of the approximation and explain why.

4. (*Exponential distribution*)

Suppose $X_1, \cdots, X_n$ are iid $Exp(\lambda)$ with probability density function $f(x) = \lambda e^{-\lambda x}, x \geq 0$.

(a) Derive the MLE for λ.

(b) Derive the probability density function of $\sum_{i=1}^{n} X_i$. Start from $n = 2$ using the density for sum of independent random variables, then carry out a mathematical induction on n.

5. (*Likelihood ratio test*)

Suppose $X_1, \cdots, X_n$ are iid $Exp(\theta_x)$ independent of $Y_1, \cdots, Y_m$ which are iid $Exp(\theta_y)$.

(a) Find the likelihood ratio statistic for the null hypothesis $\theta_x = \theta_y$ versus the alternative hypothesis $\theta_x \neq \theta_y$.

(b) (Optional for all, both 24510 and 30040)

i. Show that the likelihood ratio statistic only depends on the observations through $\bar{X}/\bar{Y}$, where $\bar{X} = \sum_{i=1}^{n} X_i/n$, and $\bar{Y} = \sum_{i=1}^{m} Y_i/m$.

ii. Use the fact that for $G_k \sim Gamma(\alpha_k, \beta_k)$ independent, $\frac{\alpha_2 \beta_1 G_1}{\alpha_1 \beta_2 G_2} \sim F(2\alpha_1, 2\alpha_2)$ to derive the probability distribution of $\bar{X}/\bar{Y}$.

iii. Use these results to show that the distribution of the likelihood ratio test is the same for all (θ_x, θ_y) in the null parameter space (i.e., whenever $\theta_x = \theta_y$).