Problem Set 8 (two pages)

Statistics 24510-30040 (W19)

Due Tuesday, March 12, at the beginning of class.

Requirements Provide detailed derivations. Select only the relevant part of the output to be inserted. Attach your code or output as an appendix if necessary. Discussions allowed, the assignment should be devised and written by yourself completely.

Problem assignments (Relevant reading: Review all, including chapters 11, 12, 14 of the text.)

1. (*Approximate confidence intervals for Poisson random variables*)

   Let $X_1, \cdots, X_n$ be i.i.d. $\sim Pois(\lambda)$. Then the MLE of λ is $\hat{\lambda} = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. The two random variables

   $$T_1 = \frac{\hat{\lambda} - \lambda}{\sqrt{\lambda/n}} \qquad \text{and} \qquad T_2 = \frac{\hat{\lambda} - \lambda}{\sqrt{\hat{\lambda}/n}}$$

   both have approximately a standard normal distribution $N(0,1)$ for large n.

   (a) Using the "pivotal method" to derive two approximate confidence intervals for λ.

   (b) Find the midpoint of the two confidence intervals.

   (c) Are the intervals guaranteed to comprise on nonnegative numbers? Explain.

   (d) For 95% confidence intervals, conduct a simulation in R to find which interval gives a more accurate coverage probability. You may use $\lambda = 1$ and, say, $n = 30$.

2. (*Poisson process parameter estimation*)

   A detector counts the number of particles emitted from a radioactive source over the course of 10-second intervals. For 180 such 10-second intervals, the following counts were observed:

Count	Number of intervals
0	23
1	77
2	34
3	26
4	13
5	7

   The above table states, for example, that in 34 of the 10-second intervals, a count of 2 was recorded. Sometimes, however, the detector did not function properly and recorded counts over intervals of length 20 seconds. This happened 20 times and the recorded counts are

Count	Number of intervals
0	2
1	4
2	9
3	5

   Assume a Poisson process model for the particle emission process. Let $\lambda > 0$ (time unit = 1 sec.) be the unknown rate of the Poisson process.

   (a) Formulate an appropriate likelihood function for the described scenario and derive the maximum likelihood estimator $\hat{\lambda}$ of the rate λ. Compute $\hat{\lambda}$ for the above data.

   (b) What approximation to the distribution of $\hat{\lambda}$ does the Central Limit Theorem suggest? (Note that the sum of all 200 counts has a Poisson distribution. What is its parameter?)

3. (*Bayesian inference for exponential distributions*)

Suppose, for given θ, that $X_1, \cdots, X_n$ are i.i.d. $\sim Exp(\theta)$ with density $f_\theta(x) = \theta e^{-\theta x}$ for $x > 0$, and we use the (improper) prior density $1/\theta$ for $\theta > 0$.

(a) Find an expression for the posterior density fo θ.
Identify the parametric family and the parameter values that this density belongs.

(b) There were 65 major earthquakes during the years 2001-2005. The gap between the first and last major earthquakes during these years was 1769.61 days. Let us assume that this gap can be modeled as the sum of 64 i.i.d. $Exp(\theta)$ random variables, where θ represents the rate (in 1/days) of major earthquakes. Find the 95% highest posterior density credible interval (HPD, see, for example, Rice, p.288) for θ. You may have to do some numerical exploration to find this interval.

(c) Find the 95% highest posterior density credible interval for the mean $\mu = 1/\theta$.
Translate this interval back into an interval for θ.

(d) Compare the intervals in (b) and (c). Explain any differences you find between the two intervals or, if they are the same, explain why they should be the same.

(e) If the times of occurrences of major earthquakes follow a Poisson process, the waiting time for the next major earthquakes should follow an $Exp(\theta)$ distribution. Using your posterior density for θ from (b), find the (marginal) probability that the waiting time for the next major quake will be at least 60 days.

4. (*Two samples of exponential distributions*)

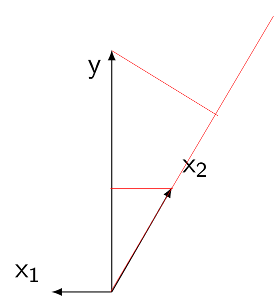
This is a two-sample version of the previous problem. Suppose, given (θ_X, θ_Y), $X_1, \cdots, X_n$ are i.i.d. $Exp(\theta_X)$ and, independent of the X_i's, $Y_1, \cdots, Y_n$ are i.i.d. $Exp(\theta_Y)$.

(a) If the (improper) prior density for (θ_X, θ_Y) is $1/(\theta_X \theta_Y)$, for $\theta_X, \theta_Y > 0$, find the posterior density for θ_Y/θ_X.

(b) What is a good reason to consider the inference of θ_Y/θ_X instead of the inference of $\theta_Y - \theta_X$? You may refer to your work in earlier assignments.

5. (*Projections of least squares method*)

Vectors x_1, x_2 and Y are in $\mathbb{R}^d, d \geq 20$. Suppose that they fall in the same two-dimensional plane given by the surface of this page just as pictured (exactly, angles and all). Also assume the length $\|Y\| = 1$.



(a) If we regressed Y on just x_1 (no intercept), what would be the least squares estimate of the regression coefficient and the resulting residual sum of squares?

(b) Do the same as in (a) but using x_2 as the regressor.

(c) For the model in part (b), work out a good approximation to the t-statistic for the hypothesis that the regression coefficient is 0 and give your best guess for the associated p-value.

(d) When Y is regressed on both x_1 and x_2 (again no intercept), give reasonable approximations to the regression coefficients and the residual sum of squares. Explain how you obtained your answers.