



Northeastern University

College of Science

Module 5 Homework

1. (20 points)

(a) Find the analytic MLE formula for exponential distribution $\exp(\lambda)$. Show that MLE is the same as MoM estimator here.

(b) A random sample of size 6 from the $\exp(\lambda)$ distribution results in observations: 1.636, 0.374, 0.534, 3.015, 0.932, 0.179. Find the MLE on this data set in two ways: by numerical optimization of the likelihood and by the analytic formula.

For (b): please give both values from the analytic MLE formula and numerical MLE solution on this data set. Also, please submit the R code for numerically finding the MLE.



Northeastern University

College of Science

2. (15 points)

A random sample of X_1, \dots, X_{53} , from the chi-square distribution with \mathbf{m} degree of freedom, has sample mean $\bar{X} = 100.8$ and sample standard deviation $s = 12.4$.

(a) Find the point estimator of \mathbf{m} using the method of moments.

(b) Find a one-sided 90% lower confidence interval of \mathbf{m} .

Please provide the formulas and the derivations together with your numerical answer.



Northeastern University

College of Science

3. (35 points)

On the Golub et al. (1999) data set, analyze the Zyxin gene expression data separately for the ALL and AML groups.

- (a) Find the bootstrap 95% CIs for the mean and for the variance of the gene expression in each group separately.
- (b) Find the parametric 95% CIs for the mean and for the variance of the gene expression in each group separately. (You need to choose the appropriate approximate formula to use: z-interval, t-interval or chi-square interval.)
- (c) Find the bootstrap 95% CI for the median gene expression in both groups separately.
- (d) Considering the CIs in parts (a)-(c), does the Zyxin gene express differently in ALL and AML patients?

Please provide numerical answers for each part. Please also submit your R codes used for the calculations (the R code should be clearly labeled and separated for each part).



Northeastern University

College of Science

4. (30 points)

For a random sample of 50 observations from Poisson distribution, we have two ways to construct a 90% CI for the parameter λ .

(1) Since the Poisson mean is λ , we can use the interval for the sample mean

$$(\bar{X} + t_{0.05, 49} \sqrt{\frac{\bar{X}}{50}}, \bar{X} + t_{0.95, 49} \sqrt{\frac{\bar{X}}{50}}).$$

(2) Since the Poisson variance is also λ , we can use the interval for the sample

variance directly: $(\frac{49s^2}{\chi^2_{0.95, 49}}, \frac{49s^2}{\chi^2_{0.05, 49}}).$

(a) Write a R-script to conduct a Monte Carlo study for the coverage probabilities of the two CIs. That is, to generate $\text{nsim}=1000$ such data sets from the Poisson distribution. Check the proportion of the CIs that contains the true parameter λ .

(b) Run the Monte Carlo simulation for $\text{nsim}=1000$ runs, at three different parameter values: $\lambda=0.1$, $\lambda=1$ and $\lambda=10$. Report the coverage probabilities of these two CIs at each of the three parameter values.

(c) Considering your result in part (b), which one of these two CI formulas should you use in practice? Can you explain the pattern observed in (b)?