

ST2132 Mathematical Statistics

Assignment 1

Due date: 2 February 2024 23:59 on Canvas

Please submit your writeup in pdf format. Submit your code in zip format.

Please answer all questions. Please work on the assignment by yourself only. We follow a strict rule on academic honesty.

1. Let X_1, X_2, \dots, X_n be independent chi-square distribution with degrees of freedom r_1, r_2, \dots, r_n , that is, $X_i \sim \chi^2(r_i)$, where $r_i > 0$, for $i = 1, \dots, n$. Identify the distribution of $\sum_{i=1}^n X_i$ using the moment generating function technique.
2. The number of floods that occur in a certain region over a given year is a random variable having a Poisson distribution with mean 1.8, independently from one year to the other. Moreover, the time period (in days) during which the ground is flooded, at the time of an arbitrary flood, is an exponential random variable with mean 3. We assume that the durations of the floods are independent. Using the central limit theorem, calculate (approximately)
 - (a) the probability that over the course of the next 20 years, there will be at least 19 floods in this region. Assume that we do not need to apply half-unit correction for this question.
 - (b) the probability that the total time during which the ground will be flooded over the course of the next 120 floods will be smaller than 365 days.
3. Suppose X_1, X_2, \dots, X_n are i.i.d. exponential random variables with pdf given by

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x > 0, \\ 0 & \text{otherwise.} \end{cases}$$

We are interested in estimating the parameter λ using a number of methods.

- (a) What is the method of moment estimator $\hat{\lambda}_{mom}$ of λ ?
- (b) What is the maximum likelihood estimator $\hat{\lambda}_{mle}$ of λ ?
- (c) Suppose the following $n = 6$ samples was generated from a exponential distribution with parameter λ as described in this question:

3.8 3.24 1.4 1.22 4.5 4.6

Compute $\hat{\lambda}_{mom}$ and $\hat{\lambda}_{mle}$.

- (d) Suppose that $n > 1$. Is $\hat{\lambda}_{mle}$ an unbiased estimator of λ ? Why?
4. Consider random samples X_1, \dots, X_n drawn i.i.d. from the probability density function given by

$$f(x; \theta) = \begin{cases} \frac{1}{2}(1 + \theta x), & \text{if } -1 \leq x \leq 1, \\ 0, & \text{otherwise.} \end{cases}$$

where $-1 \leq \theta \leq 1$. Find the method of moment estimator of θ . Is it an unbiased estimator of θ ?

5. Find the maximum likelihood estimator of the unknown parameter θ where X_1, X_2, \dots, X_n is a sample from the distribution whose density function is

$$f_X(x) = \begin{cases} e^{-(x-\theta)} & \text{if } x \geq \theta \\ 0 & \text{otherwise.} \end{cases}$$

6. Let X_1, \dots, X_n be i.i.d. random samples from $\text{Uniform}[0, \theta]$. We saw in class that the mle $\hat{\theta}$ of θ is $\hat{\theta} = \max\{X_1, \dots, X_n\}$. Show that this is a biased estimator of θ . Use the result to propose an unbiased estimator of θ . Hint: First find the cdf and then the pdf of $\max\{X_1, \dots, X_n\}$. Then use it to calculate its mean.
7. (Gaussian Mixture Model (GMM)) This question is about (a simplified version of) the Gaussian Mixture Model (GMM), which is a popular model in statistics, data science and machine learning. For example, it is used in image processing and various clustering algorithms. Suppose that K is a discrete random variable that can either be 0 or 1 with probability π_0 and π_1 respectively, that is,

$$K = \begin{cases} 0, & \text{with probability } \pi_0 = P(K = 0), \\ 1, & \text{with probability } \pi_1 = P(K = 1) = 1 - \pi_0. \end{cases}$$

Conditional on $K = k$ with $k \in \{0, 1\}$, the distribution of X is $N(\mu_k, \sigma_k^2)$, a normal distribution with mean μ_k and variance σ_k^2 . That is,

$$X|K = 0 \sim N(\mu_0, \sigma_0^2),$$

$$X|K = 1 \sim N(\mu_1, \sigma_1^2).$$

- (a) Derive the joint density of (X, K) . State clearly the support of (X, K) in the joint density. Hint: consider conditional distribution and the law of total probability.
- (b) Denote the distribution of $(X, K) \sim \text{GMM}(\pi_0, \mu_0, \sigma_0^2, \mu_1, \sigma_1^2)$. Note that π_1 can be omitted as a parameter since $\pi_1 = 1 - \pi_0$. Suppose that we have an i.i.d. random sample of size n of these n pairs $(X_1, K_1), (X_2, K_2), \dots, (X_n, K_n)$. Each X_i belongs to either group 0 or group 1 depending on K_i . Using part (a), derive the maximum likelihood estimator for all the five parameters $\pi_0, \mu_0, \sigma_0^2, \mu_1, \sigma_1^2$. Hint: Let $n_0 = \sum_{i=1}^n \mathbf{1}_{\{K_i=0\}}$ and $n_1 = \sum_{i=1}^n \mathbf{1}_{\{K_i=1\}}$ be the number of X_i that belongs to group 0 and group 1 respectively. You may find expressing the likelihood function in terms of n_0 and n_1 useful.
- (c) On Canvas, there is a dataset called “GMM.csv”, which is a comma separated file. This dataset contains $n = 100,000$ rows and 2 columns. The first column is the realizations of K , and the second column is the realizations of X . Using any computing language of your choice, e.g. Python/R/Julia/MATLAB/Excel, compute the maximum likelihood estimators derived in part (b) using this dataset. Attach your code at the end of your assignment.