

8. For a continuous random variable X , its cumulative distribution function (CDF) $F(x)$ follows a uniform distribution the interval zero to one, i.e. $F(x) \sim Uniform(0, 1)$. The order statistics of the continuous random variable $X_i, i = 1 \dots, n$ are denoted by $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$. Then

$$F(X_{(1)}) < F(X_{(2)}) < \dots < F(X_{(n)})$$

since CDF $F(\cdot)$ is a nondecreasing function and the probability of an equality is zero. Let $U_i = F(X_{(i)})$ for $i = 1, \dots, n$.

- (a) (5 points) Find the pdf of the r th order statistic $U_{(r)}$ of U .
- (b) (7 points) Find the mean $E(U_{(r)})$ and variance $Var(U_{(r)})$ of the r th order statistic $U_{(r)}$.

7. For n independent observations $(x_1, y_1), \dots, (x_n, y_n)$, the simple linear regression model says that the value of y_i is generated by a random variable

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

where β_0, β_1, x_i are fixed (not random) values, and ε_i are random variable with mean 0 and variance σ^2 .

- (a) (6 points) Assume that each $\varepsilon_i \sim N(0, \sigma^2)$. Show that $Y_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2)$.

Assume part (a) holds, do the following parts:

- (b) (5 points) Find the p.d.f. $f_{Y_i}(y)$ of Y_i .
- (c) (5 points) Write down the likelihood based on the data $(x_1, y_1), \dots, (x_n, y_n)$ as a function of β_0, β_1 and σ .
- (d) (6 points) If we further assume that σ^2 is known, find the maximum likelihood estimation (MLE) of β_0 and β_1 .

(15 points) Let $X_1, \dots, X_n$ be i.i.d. copies of a Bernoulli random variable $X \in \{0, 1\}$ with success parameter $\theta \in (0, 1)$:

$$P_\theta(X = 1) = 1 - P_\theta(X = 0) = \theta$$

We consider the testing problem with the chosen level $\alpha = 0.05$. We want to test, at level α , the hypothesis

$$H_0 : \theta = 1/2 := \theta_0$$

against the alternative

$$H_a : \theta = 1/4 := \theta_a$$

By the Neyman Pearson Lemma, we can find the most powerful test ϕ (at level α). Specifically, consider the test

$$\phi(T) := \begin{cases} 1 & \text{if } T < t_0 \\ q & \text{if } T = t_0 \\ 0 & \text{if } T > t_0 \end{cases}$$

where T is a sufficient statistic. The constants q and $t_0 \in \{0, \dots, n\}$ are chosen in such a way that the probability of rejecting H_0 when it is in fact true, is equal to α :

$$P_{\theta_0}(H_0 \text{ rejected}) = P_{\theta_0}(T \leq t_0 - 1) + qP_{\theta_0}(T = t_0) := \alpha$$

Thus, we take t_0 in such a way that

$$P_{\theta_0}(T \leq t_0 - 1) \leq \alpha, P_{\theta_0}(T \leq t_0) > \alpha$$

and

$$q = \frac{\alpha - P_{\theta_0}(T \leq t_0 - 1)}{P_{\theta_0}(T = t_0)}.$$

Now, suppose that $n = 7$. Find the t_0 , q and the power of the test ϕ .

5. Suppose that $X_1, X_2, \dots, X_{n_1}$, $Y_1, Y_2, \dots, Y_{n_2}$, and $W_1, W_2, \dots, W_{n_3}$ are independent random samples from normal distributions with respective unknown means μ_1 , μ_2 , and μ_3 and variances σ_1^2 , σ_2^2 , and σ_3^2 .
- (a) (6 points) Find the likelihood ratio test for $H_0 : \sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma^2$ against the alternative of at least one inequality.
 - (b) (6 points) Find an approximate critical region for the test in part (a) if n_1 , n_2 , and n_3 are large and $\alpha = .05$.

6. (8 points) Suppose that $Y_1, Y_2, \dots, Y_{40}$ denote a random sample of measurements on the proportion of impurities in iron ore samples. Let each variable Y_i have a probability density function given by

$$f(y) = \begin{cases} 3y^2, & 0 \leq y \leq 1, \\ 0, & \text{elsewhere.} \end{cases}$$

The ore is to be rejected by the potential buyer if $\bar{Y}$ exceeds 0.7. Find $P(\bar{Y} > 0.7)$ for the sample of size 40.

Suppose we have n data points $x_1, x_2, \dots, x_n$.

(a) (9 points) First show that

$$\sum_{i=1}^n (x_i - A)^2 = \sum_{i=1}^n (x_i - \bar{x})^2 + n(\bar{x} - A)^2$$

where $\bar{x}$ denotes the mean of $x_1, x_2, \dots, x_n$. Explain your answer. Show all steps to receive the full credit.

(b) (6 points) From equation above, show that

$$\sum_{i=1}^n (x_i - A)^2$$

is minimized by $A = \bar{x}$.

Suppose that the historical median sales prices for new single-family houses in Orlando area are as follows

x : Year	y : median sales price in thousands
1991	199
1992	207
1993	223
1994	235
1995	249
1996	264
1997	278
1998	293

- (a) (5 points) Given this data, find the least squares line $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$
- (b) (5 points) Test the significance of the linear relationship between year and median sales price using a significance level of $\alpha = .05$. (Write down the hypothesis, test statistics, p-value and conclusion clearly).

1. (6 points) Do you believe that an exceptionally high percentage of the executives of large corporations are right-handed? Although 85% of the general public is right-handed, a survey of 300 chief executive officers of large corporations found that 96% were right-handed. Is this difference in percentages statistically significant? Find the p -value for the test and explain your conclusion with $\alpha = .01$.