Large Sample Theory

Homework 5: Maximum Likelihood Estimate, Testing, Asymptotic Distribution

Due Date: January 12th

1. Consider the classical Gaussian linear model $Y_i = \mu_i + \epsilon_i$, $1 \le i \le n$, where $\mu_i = \mathbf{z}_i^T \boldsymbol{\beta}$ and ϵ_i are i.i.d. Gaussian with mean 0 and variance σ^2. Here $\mathbf{z}_i$ are d-dimensional vectors for covariate values. Suppose that the covariates are ranked in order of importance. (It means that the first covariate is the most important and etc.)
   To entertain the possibility that the last $d-p$ don't matter, $\beta_{p+1} = \cdots = \beta_d = 0$. Let $\hat{\boldsymbol{\beta}}^{(p)}$ be the least-squares estimate with $\beta_{p+1} = \cdots = \beta_d = 0$ and $\hat{Y}_i^{(p)}$ the corresponding fitted value.
   In this fashion, we end up d possible regression models. Now the problem is which one to use. A natural goal to entertain is to obtain new values $Y_1^*, \dots, Y_n^*$ at $\mathbf{z}_1, \dots, \mathbf{z}_n$ and evaluate the performance of $\hat{Y}_1^{(p)}, \dots, \hat{Y}_n^{(p)}$ as estimates of $Y_1^*, \dots, Y_n^*$ and, hence, the model with $\beta_{d+1} = \cdots = \beta_p = 0$ by the (average) expected prediction error

   $$EPE(p) = n^{-1} E \sum_{i=1}^{n} (Y_i^* - \hat{Y}_i^{(p)})^2.$$

   Here $Y_1^*, \dots, Y_n^*$ are independent of $Y_1, \dots, Y_n$ and Y_i^* is distributed as Y_i, $i = 1, \dots, n$. Let $RSS(p) = \sum_{i=1}^n (Y_i - \hat{Y}_i^{(p)})^2$ be the residual sum of squares. Suppose that σ^2 is known.

   (a.) Show that

   $$EPE(p) = \sigma^2 \left(1 + \frac{p}{n}\right) + \frac{1}{n} \sum_{i=1}^{n} (\mu_i - \mu_i^{(p)})^2$$

   where $\mu_i^{(p)} = \mathbf{z}_i^T \hat{\boldsymbol{\beta}}^{(p)}$ and $\hat{\boldsymbol{\beta}}^{(p)} = (\beta_1, \dots, \beta_p, 0, \dots, 0)^T$.

   (b). Show that

   $$E[RSS(p)] = \sigma^2 \left(1 - \frac{p}{n}\right) + \frac{1}{n} \sum_{i=1}^{n} (\mu_i - \mu_i^{(p)})^2.$$

   (c). Show that $RSS(p) + (2p/n)\sigma^2$ is an unbiased estimate of $EPE(p)$.

   (d). Mallow (1973, *Technometrics*) suggested a model selection rule in which p is selected to be the one minimizes $RSS(p) + (2p/n)\sigma^2$ and then using $\hat{\mathbf{Y}}(\hat{p})$ as a predictor. Suppose $p = 2$ and $d = 3$. Find the probability that $P(\hat{p} = 3)$ and $P(\hat{p} \le 1)$ when n goes to infinity. (You can assume that those covariates are realized values of 3 independent $UNIF(0,1)$ random variables. For example, $\mu_i = \beta_1 z_{i1} + \beta_2 z_{i2} + \beta_3 z_{i3}$ where z_{i1}, z_{i2}, and z_{i3} are independent $UNIF(0,1)$ random variables.

2. Consider model $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ where $E(\boldsymbol{\epsilon}) = \mathbf{0}$ and $Var(\boldsymbol{\epsilon}) = \sigma^2 \mathbf{J}_n$. Let $\hat{Y}_i = \mathbf{X}_i \hat{\boldsymbol{\beta}}$ and $h_{ii} = \mathbf{X}_i (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}_i^T$.
   (a) Show that for any $\epsilon > 0$,

   $$P(|\hat{Y}_i - E(\hat{Y}_i)| \ge \epsilon) \ge \min[P(\epsilon_i \ge \epsilon / h_{ii}), [P(\epsilon_i \le -\epsilon / h_{ii})].$$

   (Hint: for independent random variables X and Y, $P(|X+Y| \ge \epsilon) \ge P(X \ge \epsilon)P(Y \ge 0) + P(X \le -\epsilon)P(Y < 0)$.)
   (b) Show that $\hat{Y}_i - E(\hat{Y}_i) \xrightarrow{P} 0$ if and only if $h_{ii} \to 0$.

3. Let (X_i, Y_i), $1 \le i \le n$, be iid with X_i and Y_i independent, $N(\theta_1, 1)$, $N(\theta_2, 1)$, respectively. Suppose $\theta_1 \ge 0$ and $\theta_1 \ge \theta_2 \ge 0$. Consider testing $H_0 : \theta_1 = \theta_2 = 0$ versus $H_1 : \theta_1 > 0$ or $\theta_2 > 0$. Show that whatever be n, under H_0, λ_n is distributed as a mixture of point mass at 0, χ_1^2 and χ_2^2 with probabilities $3/8$, $1/2$, $1/8$, respectively.

4. Let $(X_{11}, X_{12}), \ldots, (X_{n1}, X_{n2})$ be i.i.d. from a bivariate normal distribution with unknown mean and covariance matrix. For testing $H_0 : \rho = 0$ versus $H_1 : \rho \ne 0$, where ρ is the correlation coefficient, show that the test rejecting H_0 when $|W| > 0$ is an LR test, where

$$W = \sum_{i=1}^{n}(X_{i1} - \bar{X}_1)(X_{i2} - \bar{X}_2) \Big/ \left[\sum_{i=1}^{n}(X_{i1} - \bar{X}_1)^2 + \sum_{i=1}^{n}(X_{i2} - \bar{X}_2)^2\right].$$

   Find the distribution of W under H_0.

5. Suppose you are studying the number of visitations of a pollinator to a flower. Your hypothesis is that yellow flowers are better than red flowers (in terms of pollinator attraction). Previous studies have found that the number of visitors to red flowers follows a normal distribution with a mean of 200 visits per flower and a variance of 50. Suppose in a sample of 20 yellow flowers that the mean number of visits is 202 with a known variance (of visits per flower) of 50. Again, assume the number of visitors is normally distributed.

   a. What is the probability of this data under the null hypothesis (yellow and red flowers are equivalent)?
   b. What is the critical value for a (one-sided) test of the null hypothesis at the $\alpha = 0.05$ level?
   c. What are the values for (a) and (b) when the variance for yellow flowers (50) is instead a SAMPLE variance (i.e., an estimate of the true variance)? Hint: Would you now use a normal or a t distribution?
   d. Suppose that yellow flowers are indeed better. Given the sample size (20) and assuming the variance (50) is the true value, how small an effect can we detect using a (one-sided) test of significance of $\alpha = 0.05$ with 80% power?
   e. Repeat the calculation in (d) assuming that the variance (50) is now an estimated value, not necessarily the true value.
   f. Suppose the true mean and variance for yellow flowers are 201 and 10. How large a sample size is required to have a power of 80 percent of detecting a difference between red and yellow using a test of significance with level $\alpha = 0.05$? Compute this for both the normal (variance assumed know) and t (variance estimated) settings.
   g. If the true variance for yellow is 35, what is the probability that we observe a sample variance of 50 (or larger) given our sample size of 20.

6. Let $X_1, X_2, \ldots, X_n$ be a random sample from the $unif(0, \theta)$ distribution for some $\theta > 0$. Suppose we wish to test

$$H_0 : \theta = \theta_0 \quad \text{versus} \quad H_a : \theta < \theta_0$$

   at level (size) α. Suppose that we use test statistic $X_{(n)}$.

   a. Derive the test with the probability of a Type I error α.

b. What is the probability of a type Type II error for any particular $\theta = \theta_1$ where θ_1 is some fixed number less than θ_0?

c. What is the power function of this test?

d. What sample size is necessary in order to get $\beta(\theta_1) = \beta$ where β is a fixed number between 0 and 1 and θ_1 is a fixed value between 0 and θ_0?

7. Let $X_1, \ldots, X_n$ be the times in months until failure of n similar pieces of equipment. Since the equipment is subject to wear, we often model $X_1, \ldots, X_n$ as a random sample of size n from a Weibull distribution with density $f(x, \lambda) = \lambda c x^{c-1} \exp(-\lambda x^c)$, $x > 0$. Here c is a known positive constant and $\lambda > 0$.

   a. Find an optimal test for testing $H_0 : 1/\lambda \leq 1/\lambda_0$ versus $H_a : 1/\lambda > 1/\lambda_0$.

   b. Suppose that the only table you have is a normal probability table. Can you use this table to carry out the test derived in (a)? Give reasons to justify your answer.

8. Let X_n be a random variable having the Poisson distribution $P(n\theta)$, where $\theta > 0$, $n = 1, 2, \ldots$. Show that $(X_n - n\theta)/\sqrt{n\theta} \xrightarrow{d} N(0, 1)$.

9. Let $U_1, \ldots, U_n$ be i.i.d. random variables having the uniform distribution on $[0, 1]$ and $Y_n = (\prod_{i=1}^n U_i)^{-1/n}$. Show that $\sqrt{n}(Y_n - e) \xrightarrow{d} N(0, e^2)$.

10. Set $\hat{\sigma} = \sqrt{n^{-1} \sum_{i=1}^n (X_i - \bar{X})^2}$. Show that $\sqrt{n}(\hat{\sigma} - \sigma) \xrightarrow{d} N(0, \sigma^2/2)$.

11. Let $X_1, \ldots, X_n$ be i.i.d. $N(\theta, 1)$ with $\theta \geq 0$.
    (a) Show that the MLE of θ, $\hat{\theta}_n$, is $\bar{X}$ if $\bar{X} > 0$ and 0 otherwise.
    (b) If $\theta > 0$, show that $\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{L} N(0, 1)$.
    (c) If $\theta = 0$, the probability is $1/2$ that $\hat{\theta}_n = 0$ and $1/2$ that $\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{L} N(0, 1)$.

12. If $X_1, \ldots, X_n$ are i.i.d. according to $U(0, \theta)$ and $T_n = X_{(n)}$, the limiting distribution of $n(\theta - T_n)$ is exponential with density $\theta^{-1} \exp(-x/\theta)$. Use this result to determine the limit distribution of
    (a) $n[f(\theta) - f(T_n)]$, where f is any function with $f^{(1)}(\theta) \neq 0$;
    (b) $[f(\theta) - f(T_n)]$ is suitably normalized if $f^{(1)}(\theta) = 0$ but $f^{(2)}(\theta) \neq 0$.

13. Let $X_1, \ldots, X_n$ be i.i.d. $N(\theta, \sigma^2)$ and consider the estimation of θ^2.
    (a) Find the maximum likelihood estimator.
    (b) Obtain the limit distribution of the estimators obtained in (a). (Hint: You may need to consider $\theta \neq 0$ and $\theta = 0$ separately.)

14. Let $X_1, \ldots, X_n$ be i.i.d. with $E(X_i) = \theta$, $Var(X_i) = \sigma^2 < \infty$, and let $\delta_n = \bar{X}$ with probability $1 - \epsilon_n$ and $\delta_n = A_n$ with probability ϵ_n. If ϵ_n and A_n are constants satisfying

$$\epsilon_n \to 0 \quad \text{and} \quad \epsilon_n A_n \to \infty,$$

    then δ_n is consistent for estimating θ, but $E(\delta_n - \theta)^2$ does not tend to zero.

15. Suppose that X_n is a random variable having the binomial distribution $Bin(n, p)$, where $0 < p < 1$, $n = 1, 2, \ldots$. Define

$$Y_n = \begin{cases} \log(X_n/n) & X_n \geq 1 \\ 1 & X_n = 0. \end{cases}$$

    Show that $Y_n \xrightarrow{a.s.} \log p$ and $\sqrt{n}(Y_n - \log p) \xrightarrow{d} N(0, (1-p)/p)$.

16. Let $X_1, \ldots, X_n$ be iid random variables with $Var(X_1) < \infty$. Show that

$$\frac{2}{n(n+1)} \sum_{j=1}^{n} jX_j \xrightarrow{P} EX_1.$$

17. Let $(Y_1, Z_1), \ldots, (Y_n, Z_n)$ be i.i.d. with the Lebesgue pdf

$$\lambda^{-1}\mu^{-1}e^{-y/\lambda}e^{-z/\mu}I_{(0,\infty)}(y)I_{(0,\infty)}(z),$$

where $\lambda > 0$ and $\mu > 0$.
(a) Find the MLE of (λ, μ).
(b) Suppose that we only observe $X_i = \min(Y_i, Z_i)$ and $\delta_i = 1$ if $X_i = Y_i$ and $\delta_i = 0$ if $X_i = Z_i$. Find the MLE of (λ, μ).

18. Let X be $N(0, \theta)$, $0 < \theta < \infty$. Find the Fisher information $I(\theta)$.