Large Sample Theory

Homework 5: Asymptotic, Maximum Likelihood

Due Date: January 9th, 2004

1. Let X_n be a random variable having the Poisson distribution $P(n\theta)$, where $\theta > 0$, $n = 1, 2, \ldots$. Show that $(X_n - n\theta)/\sqrt{n\theta} \xrightarrow{d} N(0, 1)$.
2. 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 $\sqrt{n}(Y_n - e) \xrightarrow{d} N(0, e^2)$.
3. Set $\hat{\sigma} = \sqrt{n^{-1}\sum_{i=1}^n (X_i - \bar{X})^2}$. Show that $\sqrt{n}(\hat{\sigma} - \sigma) \xrightarrow{d} N(0, \sigma^2/2)$.
4. 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 $\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{L} N(0, 1)$.
5. 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 $f^{(2)}(\theta) \neq 0$.
6. 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) and (b). (Hint: You may need to consider $\theta \neq 0$ and $\theta = 0$ separately.)
7. 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.
8. 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 $\sqrt{n}(Y_n - \log p) \xrightarrow{d} N(0, (1-p)/p)$.
9. Let $X_1, \ldots, X_n$ be iid random variables with $Var(X_1) < \infty$. Show that
   $$\frac{2}{n(n+1)}\sum_{j=1}^n jX_j \xrightarrow{P} EX_1.$$
10. Let $X_1, \ldots, X_n$ be iid random variables having a finite $E|X_1|^4$ and let $\bar{X}$ and S^2 be the sample mean and variance. Express $E(\bar{X}^3)$, $Cov(\bar{X}, S^2)$, and $Var(S^2)$ in terms of $\alpha_k = EX_1^k$, $k = 1, 2, 3, 4$. Find a condition under which $\bar{X}$ and S^2 are uncorrelated.

11. Let $X_1, \ldots, X_n$ be iid with mean μ, variance σ^2, and finite $\mu_j = EX_1^j$, $j = 2, 3, 4$. The sample coefficient of variation is defined to be $S/\bar{X}$, where S is the squared root of the sample variance S^2.
(a) If $\mu \neq 0$, show that $\sqrt{n}(S/\bar{X} - \sigma/\mu) \xrightarrow{d} N(0, \tau)$ and obtain an explicit formula of τ in terms of μ, σ^2, and μ_j.
(b) If $\mu = 0$, how would you describe the asymptotic distribution of $S/\bar{X}$ with suitable normalization?

12. 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).

13. Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be i.i.d. from a two-dimensional normal distribution with $E(X_1) = E(Y_1) = 0$, $Var(X_1) = Var(Y_1) = 1$, and an unknown correlation coefficient $\rho \in (-1, 1)$. Show that the likelihood equation is a cubic in ρ and the probability that it has a unique root tends to 1.

14. et $Z_1, \ldots, Z_n$ be *survival times* that are iid nonnegative random variables from a cdf F, and $Y_1, \ldots, Y_n$ be iid nonnegative random variables independent of Z_i's. Suppose that we are only able to observe the smaller of Z_i and Y_i and an indicator of which variables is smaller:

$$X_i = \min(Z_i, Y_i), \quad \delta_i = I_{(0,Y_i)}(Z_i), \quad i = 1, \ldots, n.$$

An *maximum empirical likelihood estimator* (MELE) of F is defined by the maximizer of

$$\ell(F) = \prod_{i=1}^{n} p_i^{\delta_{(i)}} \left(\sum_{j=i+1}^{n+1} p_j \right)^{1-\delta_{(i)}}$$

subject to $p_i = P_F(\{x_{(i)}\}) \geq 0$, $1 \leq i \leq n$, $p_{n+1} = 1 - F(x_{(n)}) \geq 0$, and $\sum_{i=1}^{n+1} p_i = 1$.
(a) Show the above maximization problem is equivalent to the maximization of

$$\prod_{i=1}^{n} q_i^{\delta_{[i]}}(1 - q_i)^{n-i+1-\delta_{[i]}}$$

where $q_i = p_i / \sum_{j=i}^{n+1} p_j$, $i = 1, \ldots, n$.
(b) Show that the MELE is

$$\hat{F}(t) = \sum_{i=1}^{n+1} \hat{p}_i I_{[X_{(i-1)}, X_{(i)})}(t),$$

where $X_{(0)} = 0$, $X_{(n+1)} = \infty$, $X_{(i)}$ are order statistics, and

$$\hat{p}_i = \frac{\delta_{[i]}}{n-i+1} \prod_{j=1}^{i-1} \left(1 - \frac{\delta_{[j]}}{n-j+1}\right), i = 2, \ldots, n, \hat{p}_{n+1} = 1 - \sum_{j=1}^{n} \hat{p}_j.$$

(c) Show that $\hat{F}(t)$ can also be expressed as

$$1 - \prod_{X_{(i)} \le t} \left(1 - \frac{\delta_{[i]}}{n - i + 1}\right),$$

which is the Kaplan-Meier product-limit estimator. (Hint: Use the fact that $p_i = q_i \prod_{j=1}^{i-1}(1 - q_j)$.)

15. Consider a 2×2 contingency table from a propective study in which people in a particular industry who were or were not exposed to a pollutant are followed up and, after several tears, categorised according to the presence or absence of a disease. The table below gives the probabilities for each cell

	Exposed	
	Yes	No
Diseased	p_1	p_2
Not diseased	$1 - p_1$	$1 - p_2$

Define the odds ratio ψ of the disease for the *exposed* and *not exposed* groups. Consider the simple logistic model

$$p_i = \frac{\exp(\beta_i)}{1 + \exp(\beta_i)}$$

(a) Show that $\psi = 1$ corresponds to no difference between the exposed and unexposed groups.

(b) Now consider n 2×2 tables, one for each level x_j of a factor, with $j = 1, \ldots, n$. For the logistic model

$$p_{ij} = \frac{\exp(\alpha_i + \beta_i x_j)}{1 + \exp(\alpha_i + \beta_i x_j)}$$

where $i = 1, 2$ and $j = 1, \ldots, n$. Show that $\log \psi$ is constant over all tables if $\beta_1 = \beta_2$.