Spring 2006 Math 541b Exam

1. Let $X_1, \ldots, X_n$ be i.i.d. samples from a Weibull distribution with density $f(x, \lambda) = \lambda cx^{c-1}e^{-\lambda x^c}$, where $x > 0$, and $c$ is a known positive constant and $\lambda > 0$ is the scale parameter of interest. Let $\mu = 1/\lambda$.

(a) Show that $\sum_{i=1}^n X_i^c$ is an optimal test statistic or testing $H: \mu = \mu_0$ versus $K: \mu = \mu_1 > \mu_0$. That is, the most powerful test takes the form:

\[
\begin{cases} 
\text{reject } H & \text{if } \sum_{i=1}^n X_i^c > \text{critical value} \\
\text{accept } H & \text{if } \sum_{i=1}^n X_i^c \leq \text{critical value}.
\end{cases}
\]

(b) Show that $\lambda X_i^c$ follows the standard exponential distribution $Exp(1)$.

(c) Find the critical value for the size $\alpha$ most powerful test.

(d) Show that the power of the most powerful test of size $\alpha$ is given by

\[
\beta(\mu_1) = 1 - G_n\left(\frac{\mu_0}{\mu_1} g_n(1 - \alpha)\right),
\]

where $G_n$ is the distribution function of $\Gamma(n, 1)$, $g_n(1 - \alpha)$ is the $(1 - \alpha)$th quantile of $\Gamma(n, 1)$, and prove that $\beta(\mu)$ is increasing in $\mu$.

(e) Show that the most powerful test of size $\alpha$ for the simple hypotheses in (a) is uniformly most powerful, at size $\alpha$, for testing the composite hypotheses $H: \mu \leq \mu_0$ versus $K: \mu > \mu_0$.

(f) When $n$ is large, please use normal approximation to find the critical value and power.

2. Let $X_i, B_i, i = 1, \ldots, n$ be independent Bernoulli variables where $X_i$ has unknown success probability $p \in (0, 1)$, and $B_i$ has success probability $1/3$. Suppose we observes

\[Y_i = B_i X_i + (1 - B_i)(1 - X_i), \quad i = 1, \ldots, n\]

that is, we see the original $X_i$ with probability $1/3$, and $1 - X_i$ with probability $2/3$.

(a) Write the log likelihood in term so the sum $S_n = \sum_{i=1}^n Y_i$, and the equation one would solve for finding the maximum likelihood estimator.
(b) Introduce appropriate missing data for the implementation of the
EM algorithm and write out the full likelihood, and the maximum
likelihood estimator using this data.

(c) Detail the steps of the EM algorithm.
Spring 2007 Math 541b Exam

1. Let \( p = (p_1, \ldots, p_c) \) be a vector of positive numbers summing to one, and \( X \sim M(n, p) \), the multinomial distribution given by

\[
P(X = k) = \binom{n}{k} p^k,
\]

where \( k = (k_1, \ldots, k_c) \) are non-negative integers summing to \( n \),

\[
\binom{n}{k} = \frac{n!}{\prod_{i=1}^n k_i!} \quad \text{and} \quad p^k = \prod_{i=1}^n p_i^{k_i}.
\]

For a given probability vector \( p_0 \) we test \( H_0 : p = p_0 \) versus \( H_1 : p \neq p_0 \) using the chi-squared test statistic

\[
V^2 = \sum_{i=1}^c \frac{(X_i - np_{i0})^2}{np_{i0}}.
\]

(a) Calculate the mean vector and the covariance matrix of \( X \).
(b) Define a matrix \( P \) such that

\[
V^2 = n^{-1} (X - np)' P^{-1} (X - np).
\]

(c) Show that

\[
n^{-1/2} (X - np) \overset{d}{\rightarrow} Y \sim \mathcal{N}_c(0, \Sigma)
\]

(d) Find the distribution of \( U = P^{-1/2} Y \), and show that the covariance matrix of \( U \) is a projection. (Recall that \( Q \) is a projection matrix if \( Q' = Q^2 = Q \).) Hint: show

\[
P^{-1/2} \Sigma P^{-1/2} = I - P^{-1/2} PP' P^{-1/2}.
\]

(e) Show that

\[
V^2 \overset{d}{\rightarrow} \chi^2_{c-1},
\]

that is, that \( V^2 \) converges in distribution to a chi-squared distribution with \( c - 1 \) degrees of freedom.

2. Suppose \( X_1, \ldots, X_n \) are independently and identically distributed with variance \( \sigma^2 \).
(a) Show that the estimate of variance \( \hat{\theta} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n} \) has bias equal to \(-\sigma^2/n\) as an estimator of \( \sigma^2 \).

(b) Show that the bias of the jackknife estimate is \(-s^2/n\), where \( s^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n} \).
1. Let \( p_0(x) \) and \( p_1(x) \) be two distinct density functions on \( \mathbb{R}^d \).

\( \text{a) Based on } X_1, \ldots, X_n \text{ independent and identically distributed with density } p, \text{ it is desired to test } H_0 : p = p_0 \text{ versus } H_1 : p = p_1. \)

Prove the Neymann Pearson Lemma, that the test which rejects \( H_0 \) when the likelihood ratio

\[ L_n = \prod_{i=1}^{n} \frac{p_1(X_i)}{p_0(X_i)} \]

exceeds a threshold has maximum power over all tests having the same Type I error.

\( \text{b) Let } E_i, i \in \{0, 1\} \text{ be the expectation when the density of } X \text{ is } p_i(X). \) Prove that

\[ K(1, 0) < 0 \quad \text{where} \quad K(1, 0) = E_0 \log \frac{p_1(X)}{p_0(X)}. \]

\( \text{c) Let } X_1, X_2, \ldots \text{ be an infinite sequence of observations independent and identically distributed with density } p, \text{ which are observed one at a time. Let } a < 0 < b \text{ and consider the test which, after } n \text{ observations, rejects } H_0 \text{ if } \log L_n > b, \text{ accepts } H_0 \text{ if } \log L_n < a, \text{ and takes an additional observation otherwise. By considering } n^{-1} \log L_n, \text{ show that this test always terminates, whether } H_0 \text{ or } H_1 \text{ is true.} \)

2. Let \( X_1, \ldots, X_n \) be i.i.d. with mean \( \mu \) and variance \( \sigma^2 \), and \( g \) a function with continuous derivative.

Show that the jackknife estimate of variance of

\[ \hat{\sigma} = g(X) \]

is asymptotically equivalent to what is produced using the delta method.
Spring 2008 Math 541b Exam

1. Suppose that \( X_1, \ldots, X_n \) are i.i.d. samples from the uniform distribution on \((0, \theta)\).

(a) Show that the MLE of \( \theta \) is
\[
\hat{\theta} = \max(X_1, \ldots, X_n)
\]

(b) Show that \( n(\theta - \hat{\theta}) \) converges in distribution to an exponential. Please specify the parameter.

(c) Let \( \lambda_n = \sup_{\theta} \left( L_n(\theta)/L_n(\theta_0) \right) \), where \( L_n(\theta) \) is the likelihood of \( X_1, \ldots, X_n \). Show that
\[
2 \log \lambda_n = \begin{cases} 
2n \log \frac{\theta_0}{\hat{\theta}} & \hat{\theta} \leq \theta_0 \\
\infty & \hat{\theta} > \theta_0
\end{cases}
\]

(d) Show that \( 2 \log \lambda_n \rightarrow \chi^2_2 \) in distribution. Please notice that the degree of freedom is 2.

(e) What is the asymptotic distribution of the likelihood ratio test under the general regularity conditions? Is the result in the last part consistent with the general result?

2. Consider the following formulation of the EM algorithm for the estimation of the parameter \( \psi \in \Omega \) upon observing incomplete data \( y \) which is obtained through a (many to one) function as \( y = y(x) \). The incomplete data is distributed according to \( g(y; \psi) \); the full data \( x \) according to \( g_c(x; \psi) \). These two densities are are related through the mapping \( y = y(x) \) by
\[
g(y; \psi) = \int_{x : y=y(x)} g_c(x; \psi) dx.
\]

Begin with any initial value \( \psi^{(0)} \in \Omega \), then iterate the following \( E \) and \( M \) steps.

E step. Calculate
\[
Q(\psi, \psi^{(k)}) = E_{\psi^{(k)}} \left( \log L_c(\psi) \mid y \right),
\]

M step. Let \( \psi^{(k+1)} \) be any value in \( \Omega \) that maximizes \( Q(\psi, \psi^{(k)}) \), that is
\[
Q(\psi^{(k+1)}, \psi^{(k)}) \geq Q(\psi, \psi^{(k)}) \quad \text{for all } \psi \in \Omega.
\]
(a) Argue that the conditional density of \( x \) given \( y \) when \( y = y(x) \) is

\[
k(x|y, \psi) = \frac{g_c(x; \psi)}{g(y; \psi)},
\]

and zero otherwise. Letting

\[
H(\psi, \psi^{(k)}) = E_{\psi^{(k)}}(\log k(x|y, \psi)|y),
\]

prove that

\[
H(\psi, \psi^{(k)}) \leq H(\psi^{(k)}, \psi^{(k)}) \quad \text{for all } \psi \in \Omega.
\]

(b) Use (1) to write the log likelihood of the incomplete data \( \log L(\psi) = \log g(y; \psi) \) as a difference involving the functions \( Q \) and \( H \).

(c) Prove that the log likelihood sequence \( \log L(\psi^{(k)}) \) is monotone nondecreasing (Hint: consider differences).
1. Let $X_1, \ldots, X_n$ be i.i.d. from a normal distribution with unknown mean $\mu$ and known variance 1. Suppose that negative values of $X_i$ are truncated at 0, so that instead of $X_i$ we actually observe

$$Y_i = \max\{0, X_i\}, \quad i = 1, \ldots, n,$$

from which we would like to estimate $\mu$.

(a) Explain how to use the EM algorithm to estimate $\mu$ from $Y_1, \ldots, Y_n$. Specifically, give the complete log-likelihood function $\log L_c(\mu)$ (i.e., the log of the joint density of $X_1, \ldots, X_n$) and a recursive formula for the successive EM estimates $\mu^{(k+1)}$. Write these in terms of the density $\phi$ and c.d.f. $\Phi$ of the standard normal distribution. Hint: To simplify things, assume that $X_1, \ldots, X_m$ are not truncated, and $X_{m+1}, \ldots, X_n$ are.

(b) Find the partial log-likelihood function $\log L(\mu)$ (i.e., the log of the joint density of $Y_1, \ldots, Y_n$) and use it to write down a (nonlinear) equation which the MLE $\hat{\mu}$ satisfies. Use this equation to manually verify that $\hat{\mu}$ is indeed a fixed point of the recursion found in (a).

2. Let $f$ denote the true density function of $X$, and consider testing the simple hypotheses

$$H_0 : f = f_0 \quad \text{vs.} \quad H_1 : f = f_1$$

for given densities $f_0, f_1$. For a fixed value $\pi \in (0, 1)$, suppose that the probabilities $\pi_0 = \pi$ and $\pi_1 = 1 - \pi$ can be assigned to $H_0$ and $H_1$ prior to the experiment. We will describe tests of $H_0$ vs. $H_1$ by their indicator functions

$$\psi(X) = \begin{cases} 1, & \text{the test rejects } H_0 \\ 0, & \text{the test accepts } H_0. \end{cases}$$

(a) Show that the overall probability of an error resulting from using a test $\psi$ is

$$\pi E_0[\psi(X)] + (1 - \pi) E_1[1 - \psi(X)]. \quad (1)$$

(b) Call the test $\psi^*$ minimizing (1) the Bayes optimal test. By writing (1) as a single $E_0$ expectation using the “change of measure” technique

$$E_1(\cdot) = E_0 \left[ \frac{f_1(X)}{f_0(X)} \cdot \cdot \cdot \right],$$

show that the Bayes optimal test is equivalent to a simple likelihood ratio test. Also, give the value of the likelihood ratio test’s critical value.

(c) Argue that the Bayes optimal test is hence most powerful for detecting $f_1$ at a certain significance level. Write down an expression for this significance level, and also give an upper bound for it as a function of $\pi$.

(d) The posterior probability of $H_1$ is the conditional probability that $H_1$ is true, given $X = x$. Show that the posterior probability of $H_1$ is

$$\frac{\pi_1 f_1(x)}{\pi_0 f_0(x) + \pi_1 f_1(x)}. \quad (2)$$

Show that the Bayes optimal test is also equivalent to choosing which hypothesis has the larger posterior probability.
Spring 2009 Math 541b Exam

1. Let \( X_1, \ldots, X_{n_1} \) be i.i.d. \( N(\mu_1, \sigma_1^2) \) and \( Y_1, \ldots, Y_{n_2} \) i.i.d. \( N(\mu_2, \sigma_2^2) \), where \( \mu_1, \mu_2, \sigma_1, \sigma_2 \) are all unknown and \( \sigma_1, \sigma_2 \) are not necessarily assumed to be equal. This problem concerns the generalized likelihood ratio (GLR) test of

\[ H_0 : \mu_1 = \mu_2 \quad \text{vs.} \quad H_1 : \mu_1 \neq \mu_2. \]

(a) Letting \( L(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2) \) be the log of the likelihood function of \( (X_1, \ldots, X_{n_1}, Y_1, \ldots, Y_{n_2}) \), find the (unrestricted) maximum likelihood estimates (MLEs) \( \hat{\mu}_1, \hat{\mu}_2, \hat{\sigma}_1^2, \hat{\sigma}_2^2 \) and write down

\[ L(\hat{\mu}_1, \hat{\mu}_2, \hat{\sigma}_1^2, \hat{\sigma}_2^2) \]

in as simple form as possible.

(b) For the \( H_0 \)-restricted MLEs \( \hat{\mu}, \hat{\sigma}_1^2, \hat{\sigma}_2^2 \),

i. Write down formulas for \( \hat{\sigma}_1^2 \) as functions of \( \hat{\sigma}_1^2 \) and \( \hat{\mu} \).

ii. Find a cubic equation, not depending explicitly on \( \hat{\sigma}_1^2 \), that \( \hat{\mu} \) satisfies. You do not need to solve this equation.

(c) Give the asymptotic distribution of the GLR statistic under \( H_0 \), as \( n_1, n_2 \to \infty \).

2. Let \( \hat{\theta}_n \) be a parameter estimate computed from the random sample \( X_1, \ldots, X_n \), let \( \hat{\theta}_{(i)} \) be the estimate computed from

\[ X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n, \]

and let

\[ \hat{\theta}_{(\cdot)} = \frac{1}{n} \sum_{i=1}^{n} \hat{\theta}_{(i)}. \]

Recall that the Jackknife estimate of the bias of \( \hat{\theta} \) is given by

\[ b_{JACK} = (n-1)(\hat{\theta}_{(\cdot)} - \hat{\theta}_n). \]

Now let \( X_{(1)}, \ldots, X_{(n)} \) be the order statistics of the sample, which we assume are from a continuous distribution. The sample median \( \hat{m}_n \) is

\[ \hat{m}_n = \begin{cases} 
X_{(n/2)} & \text{if } n \text{ odd}, \\
(X_{(n/2)} + X_{((n/2)+1)}/2) & \text{if } n \text{ even}.
\end{cases} \]

(a) Compute the jackknife estimate of bias \( b_{JACK} \) for the sample median \( \hat{m}_n \) in both cases.

(b) Let \( b \) denote the true bias of \( \hat{m}_n \). Is \( b_{JACK} \) always unbiased for \( b \) in the even case?
Fall 2009 Math 541b Exam

1. With \( X_1, \ldots, X_n \) i.i.d. with density \( p(x; \theta), \theta \in \mathbb{R}^d \), consider the usual likelihood ratio test for \( H_0 : \theta \in \Theta_0 \) versus \( H_1 : \theta \in \Theta_1 \) based on

\[
\Lambda_n = \frac{\sup_{\theta \in \Theta_0} \prod_{i=1}^{n} p(x_i; \theta)}{\sup_{\theta \in \Theta_0 \cup \Theta_1} \prod_{i=1}^{n} p(x_i; \theta)}.
\]

Let

\[
\lambda_n = -2 \log \Lambda_n.
\]

(a) Under sufficient regularity there exists \( Y \) such that,

\[
\lambda_n \xrightarrow{d} Y \quad \text{as} \quad n \to \infty.
\]

State the distribution of \( Y \).

(b) In parts b) and c) assume \( \theta = (\theta_1, \theta_2) \), and let \( p(x; \theta) \) be the density of the bivariate normal \( X \) whose components have unknown mean \( (\theta_1, \theta_2) \), known variances \( \sigma_1^2, \sigma_2^2 \) and known correlation coefficient \( \rho \in (-1, 1). \) For testing with \( \Theta_0 = \{\theta_1 = 0, \theta_2 = 0\} \) and \( \Theta_1 \cup \Theta_2 = \mathbb{R}^2 \), verify that the distribution of \( \lambda \) is as specified in part a).

(c) For testing with \( \Theta_0 = \{\theta_1 = 0, \theta_2 = 0\} \) and \( \Theta_1 \cup \Theta_2 = \{\theta_1 > 0, \theta_2 \in \mathbb{R}\} \), determine the distribution of \( \lambda \). Compare this distribution to the one in part b), and explain.

(d) State whether we can generalize the distributional result in c) asymptotically under the regularity assumed for the convergence in a), and justify your answer.

2. Let \( X_1, \ldots, X_n \) be i.i.d. from the exponential distribution \( \mathcal{E}(a, b) \) with density

\[
b^{-1} \exp\{- (x-a)/b\} \cdot 1\{x \geq a\}, \quad -\infty < a < \infty, \quad b > 0.
\]

and let \( X_{(1)} = \min\{X_1, \ldots, X_n\} \).

(a) Determine the uniformly most powerful (UMP) test for testing

\( H_0 : a = a_0 \) vs. \( H_1 : a \neq a_0 \) when \( b \) is assumed known.
(b) Show that the most powerful level-α test of $H_0$ vs. $H'_1 : a = a_1$, for some given $a_1 < a_0$, has power equal to

$$\beta(a_1) = 1 - (1 - \alpha)e^{-a(a_0-a_1)/b}. \quad (1)$$

(c) Show that, for the problem in part (a) but with $b$ unknown, any level-α test which rejects when

$$\frac{X_{(1)} - a_0}{\sum (X_i - X_{(1)})} \notin (C_1, C_2) \quad (2)$$

is most powerful at all alternatives $(a_1, b)$ with $a_1 < a_0$ (independent of the particular choice of $C_1, C_2$).
Spring 2010 Math 541b Exam

1. Consider the multinomial model with cell probabilities $p = (p_1, p_2, \cdots, p_m) \in \Omega$, in which $\sum_{i=1}^{m} p_i = 1$, $p_i > 0$, for $i = 1, \cdots, m$. And the count data generated from a multinomial is denoted by $x = (x_1, x_2, \cdots, x_m)$, where $\sum_{i=1}^{m} x_i = n$. Under the null hypothesis $H_0$, the vector of cell probabilities $p$ is specified by $p = p(\theta) \in \omega_0$, where $\theta \in \Theta$. One example is the parameter specification in the Pearson’s chi-square test of independence. In the alternative model $H_1$, the parameter vector $p \in \Omega - \omega_0$.

(a) Denote $\hat{p}_i = x_i/n$, and the maximum likelihood estimate of $\theta$ under $H_0$ by $\hat{\theta}$. Show that the log-likelihood ratio test statistic is

$$-2 \log \Lambda = 2 \sum_{i=1}^{m} O_i \log \left( \frac{O_i}{E_i} \right),$$

where $O_i = n \hat{p}_i$ and $E_i = np(\hat{\theta})$.

(b) Assume that $\dim \omega_0 = k$, what is the asymptotic distribution of the likelihood ratio test statistic as $n \rightarrow +\infty$.

(c) What is the Pearson test statistic in this general scenario?

(d) Show that the likelihood ratio test statistic and the Pearson test statistic are approximately equivalent. (Hint: You may use a Taylor expansion.)

2. Consider the Bernoulli-Laplace model in which there are two urns, each containing $M$ balls, and of these $2M$ total balls $M$ are black and $M$ are white. Suppose at each time step, one ball is chosen from each urn at random and they are interchanged. Let $X_n \in \{0, \ldots, M\}$ be the number of black balls in the first urn just after the $n$th step.

(a) Find the transition probabilities of the Markov Chain $X_n$.

(b) Find the stationary distribution of $X_n$. Prove stationarity.
Fall 2010 Math 541b Exam

1. Let \( p(x) \) and \( q(x) \) be two distinct density functions that are positive on \( \mathbb{R} \).

   (a) We define the Kullback Leibler divergence as
   \[
   D(p||q) = E_p \log \frac{p(x)}{q(x)},
   \]
   where \( E_p \) means taking the expectation under density \( p(x) \). Prove that \( D(p||q) \) is strictly positive.

   (b) Let \( X_1, \ldots, X_n \) be i.i.d. with density
   \[
   p(x; \theta) = \begin{cases} 
   p(x) & \theta = 0 \\
   q(x) & \theta = 1.
   \end{cases}
   \]
   For some fixed \( \delta > 0 \), consider the test \( H_0 : \theta = 0 \) versus \( H_1 : \theta = 1 \) which has rejection region \( A_n \) given by
   \[
   A_n = \{ (X_1, \ldots, X_n) : e^{n(D(p||q) - \delta)} \leq \prod_{i=1}^{n} \frac{p(X_i)}{q(X_i)} \leq e^{n(D(p||q) + \delta)} \}.
   \]
   Prove that the sequence of Type I errors \( P(A_n) \) tend to zero as \( n \to \infty \), where \( P(\cdot) \) is the probability under \( p(x) \).

   (c) With \( Q(\cdot) \) the probability under \( q(x) \), prove that the Type II errors \( Q(A_n^c) \) satisfy
   \[
   \lim_{\delta \to 0} \lim_{n \to \infty} \frac{1}{n} \log Q(A_n^c) = -D(p||q).
   \]

2. Suppose that \( X \) follows a Poisson distribution \( P(\theta) \), \( \theta > 0 \). Let \( \theta \) have a Gamma \( \Gamma(p, \lambda) \) distribution.

   (a) Show that the posterior distribution is \( \Gamma(p+x, 1+\lambda) \)

   (b) Show that if we take the loss function as \( l(\theta, a) = (a - \theta)^2 \), then the Bayes estimate of \( \theta \) is \( (p+x)/(1+\lambda) \).

   (c) Find the Bayes estimate if the loss function is \( l(\theta, a) = (a - \theta)^2 / \theta \)
1. Let \((X_1, Y_1), \ldots, (X_n, Y_n)\) be independent pairs of independent random variables \(X_i\) and \(Y_i\), where \(X\) and \(Y\) have continuous distribution functions \(F(x)\) and \(G(x)\), respectively. We want to test the hypothesis \(H_0 : F = G\) versus the alternative \(H_1 : F \neq G\). Let \(Z_i = X_i - Y_i\) and rank the numbers \(|Z_1|, |Z_2|, \ldots, |Z_n|\) in increasing order, and let \(r_i\) be the rank of \(|Z_i|\), so that the smallest absolute value receives the rank of 1.

Define the Wilcoxon signed-rank statistic \(W\) as

\[
W = \sum_{i=1}^{n} \text{sign}(Z_i)r_i,
\]

where \(\text{sign}(z) = 1\) if \(z > 0\), \(\text{sign}(z) = -1\) if \(z < 0\), and \(\text{sign}(z) = 0\) if \(z = 0\).

(a) Calculate \(P(\text{sign}(Z_i) = 1)\) and \(P(\text{sign}(Z_i) = 0)\) under the null hypotheses \(H_0\).

(b) Calculate the mean and the variance of \(W\) under the null hypothesis \(H_0\).

(c) Propose a test for \(H_0\) versus \(H_1\) that has approximate Type I error level \(\alpha\) when the sample size is large.

2. Let \(X_1, X_2, \cdots, X_n\) be i.i.d observations drawn from a mixture of two normal densities \(\mathcal{N}(\mu_1, 1)\) and \(\mathcal{N}(\mu_2, 1)\), where \(\alpha\) and \(1 - \alpha\) are the probabilities that a given observation is taken from the first and second normal distribution, respectively. We suppose that \((\mu_1, \mu_2, \alpha)\) are unknown.

(a) Write down the likelihood function of the observed data \((X_1, X_2, \cdots, X_n)\) as a function of \(\theta = (\mu_1, \mu_2, \alpha)\).

(b) In order to design an EM algorithm to estimate \(\theta\), define missing data \((Z_1, Z_2, \cdots, Z_n)\) where \(Z_i = 1\) if \(X_i\) is drawn from the first population \(\mathcal{N}(\mu_1, 1)\), and \(Z_i = 0\) if it comes from the second population. Write the likelihood function of the complete data \(((X_1, Z_1), (X_2, Z_2), \cdots, (X_n, Z_n))\).

(c) Design an EM algorithm to estimate the parameter vector \(\theta\).
1. Let $Y_1, \ldots, Y_n$ be a random sample from the uniform density on $[0, \theta]$, where $\theta$ is an unknown parameter.
   
   (a) Calculate the Likelihood function $L(\theta; Y_1, \ldots, Y_n) = p(Y_1, \ldots, Y_n|\theta)$ for $\theta$.
   
   (b) Let $\Omega = \{ \theta : 0 < \theta \leq \theta_0 \}$. Show that $\max\{L(\theta; Y_1, \ldots, Y_n) : \theta \in \Omega\} = (Y_{\text{max}})^{-\theta}$, where $Y_{\text{max}} = \max\{Y_1, \ldots, Y_n\}$.
   
   (c) Write out the form of the Generalized Likelihood Ratio $\lambda$ for the test of $H_0 : \theta = \theta_0$ versus $H_1 : \theta < \theta_0$.
   
   (d) Show that the Generalized Likelihood Ratio Test calls for $H_0$ to be rejected at level $\alpha$ if $Y_{\text{max}} \leq \theta_0 \sqrt{\alpha}$.

2. Let $(I_1, Y_1), \ldots, (I_n, Y_n)$ be i.i.d. from distribution $P_\theta$, where $\theta = (\lambda, \mu) \in (0, 1) \times \mathbb{R}$,
   
   $$P_\theta(I_i = 1) = \lambda = 1 - P_\theta(I_i = 0),$$
   
   and, given $I_i = j$, $Y_i \sim N(\mu, \sigma_j^2)$, where $\sigma_0 \neq \sigma_1$ are known positive values.
   
   (a) Write down the complete likelihood function $L_c(\lambda, \mu)$ assuming that all of $(I_1, Y_1), \ldots, (I_n, Y_n)$
   
   (b) Give explicitly the maximum likelihood estimates of $\lambda$ and $\mu$.
   
   (c) Now suppose that the $I_i$ are not observed. Give as explicitly as possible the $E$- and $M$-steps
   
   of the $EM$ algorithm, including recursive formulae for the $EM$ iterates $\lambda^{(k)}$ and $\mu^{(k)}$. 