(1) (a) Define “The family of random variables \{X_1, X_2, \ldots\} is uniformly integrable.”
(b) Give an example of a sequence of random variables \(X_n, n = 1, 2, \ldots\) where \(X_n \geq 0, \mathbb{E}X_n = 1\) and the family of random variables \{X_1, X_2, \ldots\} is not uniformly integrable.

For parts (c) and (d) assume that \(X_n \to X\) a.s. as \(n \to \infty\) and that \(X_n \geq 0\) for all \(n, \omega\).
(c) Assume that \{\(X_1, X_2, \ldots\)\} is uniformly integrable. Prove that \(\mathbb{E}X_n \to \mathbb{E}X\) as \(n \to \infty\).
(d) Assume that \(\mathbb{E}X_n \to \mathbb{E}X\) as \(n \to \infty\). Prove that \{\(X_1, X_2, \ldots\)\} is uniformly integrable.
(c) Let \(f: [0, \infty) \to [0, \infty)\) with
\[
\lim_{x \to \infty} \frac{f(x)}{x} = \infty
\]
If \(X_n \geq 0, \mathbb{E}f(X_n) \geq c < \infty\), show that \{\(X_1, X_2, \ldots\)\} is uniformly integrable.

(2) (a) Let \(Y_n, n \geq 1\) be random variables and \(Y_n\) independent of \(Y_n\) with the same distribution as \(Y_1\). If \(Y_n\) converges in distribution show that \(Y_n \to Y_n\) also converges in distribution.
(b) Let \(X_1, X_2, \ldots\) be independent and identically distributed with characteristic function \(f(i) = \exp(-c|t|^\alpha)\) where \(c > 0, \alpha > 0\) are constants, and let \(S_n = X_1 + X_2 + \cdots + X_n\). Find constants \(a_n\) so \(a_nS_n\) converges in distribution to some random variable \(Z\). How is \(Z\) related to \(X_1\)?

(3) Let \(\theta \in \Theta = (-\infty, \infty)\) and \(X_1, X_2, \ldots, X_n\) be independent and identically distributed with density
\[
f(x, \theta) = I(|x - \theta| \leq 1/2)
\]
as usual, denote the order statistics by \(X_{(1)} < X_{(2)} < \cdots < X_{(n)}\).
(a) Show that \((X_{(1)}, X_{(n)})\) is sufficient for \(\theta\).
(b) Show that \(T_n = \frac{1}{2}(X_{(n)} + X_{(1)})\) is unbiased for \(\theta\).
(c) Compute the variance of \(T_n\).
(d) Find a maximum likelihood estimate for \(\theta\) based on \((X_{(1)}, X_{(n)})\). Is it unique?

(4) Let \(\theta \in \Theta = (0, \infty)\) and \(X_1, X_2, \ldots, X_n\) be independent and identically distributed with density
\[
f(x; \theta) = I(x \in [0, \theta])
\]
Construct uniformly minimum variance unbiased estimators \(q(\theta)\) for the following choices of \(q(\theta)\), or prove they do not exist.
(a) \(q(\theta) = \theta^k\) for \(k \in \{1, 2, \ldots\}\).
(b) \(q(\theta) = e^\theta\).
MATHEMATICAL PROBABILITY AND STATISTICS (II) QUALIFYING EXAM (MATH 541A AND 507A)

SPRING 1995

(1) Recall that we say the distribution functions \( F_n \) converge in distribution to the distribution function \( F \), written \( F_n \Rightarrow F \), if

\[
\lim_{n \to \infty} F_n(x) = F(x)
\]

at all continuity points of \( F \). Show that \( F_n \Rightarrow F \) if and only if for all bounded continuous functions \( h \)

\[
\int h \, dF_n \to \int h \, dF \quad \text{as} \quad n \to \infty
\]

(2) Suppose \( X_1, X_2, \ldots \) are iid with values in \((1, \infty)\). State a condition on \( X \) which is necessary and sufficient for \( \lim_{n \to \infty} (X_1 X_2 \cdots X_n)^{1/n} \) to exist almost surely and be finite. Demonstrate why your condition is necessary.

(3) (a) Let \( X_1, \ldots, X_n \) be independent normal variates with common variance \( \sigma^2 \). With \( \mu_1, \ldots, \mu_n \) not all zero, derive the level \( \alpha \) most powerful test for the hypotheses

\[
H_0 : \mathbb{E}X_1 = \cdots = \mathbb{E}X_n = 0 \quad \text{versus} \quad H_1 : \mathbb{E}X_1 = \mu_1, \ldots, \mathbb{E}X_n = \mu_n
\]

(b) Find the level \( \alpha \) most powerful test for the above hypotheses when \( X = (X_1, \ldots, X_n) \) is multivariate normal with known covariance matrix \( \Sigma \).

(4) (a) Complete the following statement of the factorization theorem. In a regular model, a statistic \( T(X) \) is sufficient for \( \theta \) if and only if there exists functions \( g \) and \( h \) such that:

(b) Let \( X_1, X_2, \ldots, X_n \) be independent Cauchy (\( \theta \)) random variables each with density

\[
p(x; \theta) = \frac{1}{\pi} \left( \frac{1}{1 + (x - \theta)^2} \right)
\]

Show that the order statistics \( (X_{(1)}, \ldots, X_{(n)}) \) are minimal sufficient for \( \theta \).
4. Let \( X = (X_1, X_2) \) be bivariate normal, with common unknown mean \( \mu \), and known variances \( \sigma_1^2, \sigma_2^2 \), and correlation \( \rho \).

a) What is the Fisher information \( I(\mu) \) based on one observation of the pair \( (X_1, X_2) \)?

b) Express \( I(\mu) \) as a function of \( \rho \) in the special case \( \sigma_1^2 = \sigma_2^2 = 1 \).

c) Explain why the expression found in part b takes on the values it does, in the (further) special cases when \( \rho = 0 \) and \( \rho = 1 \).

d) What is the Fisher information \( I(\mu) \) based on one observation \( X \) of a multivariate \( p \)-dimensional normal vector whose components have common mean \( \mu \) and known covariance matrix \( \Sigma \)?

5. Recall for \( \mu > 0 \) the exponential density with parameter \( \mu \) is \( \mu \exp(-\mu x) \) when \( x \) is positive. Let \( \mu, \nu \) be positive and suppose that \( X_1, \ldots, X_n \) are exponential with parameter \( \mu \), and \( Y_1, \ldots, Y_m \) are exponential with parameter \( \nu \) and that all variables are independent.

a) Construct the generalized likelihood ratio test for the hypotheses \( H_0 : \mu = \nu \) versus \( H_1 : \mu \neq \nu \) and show that it can be based on the statistic

\[
T(X_1, \ldots, X_n, Y_1, \ldots, Y_m) = \frac{X_1 + \cdots + X_n}{X_1 + \cdots + X_n + Y_1 + \cdots + Y_m}.
\]

In particular, express the critical region of this test in terms of \( T \).

b) What is the distribution of \( T \) under the null hypotheses?

c) For the case \( n = m \), show that the rejection region can be written as \( \{ T : |T - a| > b \} \) and find \( a \). Can the type I error probability in this case be written entirely in terms of the cumulative distribution function of \( T \)?
(b) Justify the claim: if $S_n/n \to a$ in probability, then $\phi'(0)$ exists and equals $ia$. You may assume without proof that for characteristic functions, if $\phi_n \to \phi$ pointwise, then the convergence is uniform on compact sets.

Q4 Assume $X, X_1, X_2, \ldots$ are i.i.d. with $X \geq 0$ always.
(a) Prove that if $\text{IE}X < \infty$ then $X_n/n \to 0$ almost surely.
(b) Prove that if $\text{IE}X = \infty$, then $\limsup X_n/n = \infty$ almost surely, and hence $(X_1 + \cdots + X_n)/n \to \infty$ almost surely.

Section II. Statistics

DO ANY TWO OF THE FOLLOWING THREE PROBLEMS.

Q1 Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed random variables having exponential distribution with mean $1/\theta$.
(a) Let $X_{(1)} \leq \cdots \leq X_{(n)}$ denote the order statistics, and write $T_1 = X_{(1)}$, $T_2 = X_{(2)} - X_{(1)}$, $\ldots$, $T_k = X_{(k)} - X_{(k-1)}$ for $2 \leq k \leq n$. Find the joint distribution of $T_1, T_2, \ldots, T_k$.
(b) Light bulbs are known to have this exponential distribution. Suppose that a random sample of $n$ bulbs is put under observation, and observation is stopped when the $k^{th}$ bulb burns out. Find the maximum likelihood estimator of the mean lifetime of a bulb.
(c) Find a 95% confidence interval for this mean.

Q2 (a) Give the definition of a sufficient statistic.
(b) Let $X_1, \ldots, X_n$ be a random sample from a Poisson distribution with parameter $\lambda$. Find a sufficient statistic for $\lambda$.
(c) Show that $T = \frac{1}{n} \sum_{j=1}^{n} I(X_j = 0)$ is an unbiased estimator of $e^{-\lambda}$, and find its variance. Here, $I(A) = 1$ if $A$ is true, $= 0$ if false.
(d) Use the Rao-Blackwell Theorem and (c) to find a better estimator of $e^{-\lambda}$.
(e) What optimality properties does the estimator in (d) have?
Q3  (a) State the Cramér-Rao Inequality.
(b) Let $X_1, \ldots, X_n$ be independent and identically distributed Bernoulli random variables with mean $\theta$. Find the Cramér-Rao lower bound for the variance of unbiased estimators of $\tau(\theta) = \theta(1 - \theta)$.
(c) State and prove the Neyman-Pearson Lemma.
(d) 1000 individuals were classified according to sex, and according to whether or not they were color-blind as follows:

<table>
<thead>
<tr>
<th></th>
<th>Male</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>442</td>
<td>514</td>
</tr>
<tr>
<td>Color-blind</td>
<td>38</td>
<td>6</td>
</tr>
</tbody>
</table>

According to a genetic model, these numbers should have relative frequencies given by

\[
\begin{array}{c|cc}
  & p/2 & p^2/2 + pq \\
  & q/2 & q^2/2
\end{array}
\]

where $q = 1 - p$ is the proportion of color-blind individuals in the population. Are the data consistent with this model?
MATH 507a/541b QUALIFYING EXAM - FALL 1998

To pass you must do well enough on both the Probability and the Statistics parts – high performance on one portion does not compensate for low performance on the other.

STATISTICS

1. There are $g$ categories, $i = 1, 2, \ldots, g$ with probabilities $\pi_1, \pi_2, \ldots, \pi_g$. The random variable $X$ is defined by

$$P(X = i) = \pi_i, \quad 1 \leq i \leq g.$$ 

Let $X_1, X_2, \ldots, X_n$ be iid as $X$. Define $Z_{ij}$ by

$$Z_{i,j} = I(X_j = i).$$

The number of times the variables $X_1, \ldots, X_n$ fall in category $i$ is given by

$$Y_i = \sum_{j=1}^{n} Z_{ij}.$$ 

a) Find the mean and variance of $Y_1$.

b) Find $P((Y_1, \ldots, Y_g) = (n_1, \ldots, n_g))$.

c) Find the mean vector $\mu$ and covariance matrix $\Sigma$ of the vector $(Y_1, \ldots, Y_g)$.

d) Give a large sample test for the hypothesis $H_0 : \pi_1 = \pi_2$. 

1
2. Let \( X_1, \ldots, X_n \) be iid \( \mathcal{N}(\mu, \sigma^2) \) with known mean \( \mu \).

a) Find the UMVE of the parameter \( \sigma^2 \) and prove it to be such.

b) Of all estimators of \( \sigma^2 \) of the form

\[
\hat{\sigma} = a \sum_{i=1}^{n} (X_i - \mu)^2, \quad a \in \mathbb{R},
\]

find the one which achieves the smallest mean square error. (Hint: If \( Z \) is \( \mathcal{N}(0,1) \) then \( EZ^4 = 3 \).)
1. Let \( a_n \) and \( \mu_n \) be deterministic sequences tending to \( \infty \) and \( \mu \) respectively, and assume that the random variables \( X_n \), properly scaled, converge in distribution to \( X \); in particular, that

\[
a_n(X_n - \mu_n) \xrightarrow{d} X.
\]

a) Prove that if \( g \) is a function having a continuous derivative at \( \mu \), then

\[
a_n(g(X_n) - g(\mu_n)) \xrightarrow{d} g'(\mu)X.
\]

Now let \( Y_1, \ldots, Y_n \) be a sample of independent exponential variables ('failure times') with density \( f(t; \lambda) = \lambda e^{-\lambda t} \) for \( \lambda, t \) positive.

b) Calculate the Fisher information for \( \lambda \) in the sample.

c) Find, and justify, the limiting distribution of the maximum likelihood estimator for \( \lambda \).

d) Suppose it is desired to estimate the probability that an exponential from the same distribution will not fail before time \( x \); that is, we wish to estimate

\[
q(\lambda) = P(Y > x) = e^{-\lambda x}.
\]

What is the limiting distribution of the maximum likelihood estimator of \( q(\lambda) \)? (Hint: Use part a)

2. a) Prove the following form of the Neyman Pearson Lemma: If \( X \in \mathbb{R}^n \) is a random vector with density \( f(x; \theta) \), where \( \theta \in \{\theta_0, \theta_1\} \), then the test for \( H_0 : \theta = \theta_0 \) versus \( H_1 : \theta = \theta_1 \) which rejects \( H_0 \) when \( L(X) = f(x; \theta_1)/f(x; \theta_0) \) exceeds a level \( k \) achieves the maximum power over all tests of size \( P_0(L(X) \geq k) \).

b) Let \( X_1, \ldots, X_n \) be independent exponential variables with parameters either \( \mu_1, \ldots, \mu_n \), or \( \nu_1, \ldots, \nu_n \), known values. Find a simple test and test statistic for the Neyman Pearson tests that distinguish between the two hypotheses.
Math 541 Exam Portion

1.a) Let \( a_n \) and \( \mu_n \) be deterministic sequences tending to \( \infty \) and \( \mu \) respectively, and assume that the random variables \( X_n \), properly scaled, converge in distribution to \( X \); in particular, that
\[
a_n(X_n - \mu_n) \xrightarrow{d} X.
\]
Prove that if \( g \) is a function having a continuous derivative at \( \mu \), then
\[
a_n(g(X_n) - g(\mu_n)) \xrightarrow{d} g'(\mu)X.
\]

b) State a multidimensional version of this fact.

Now let \( X_1, \ldots, X_n \) be iid with mean \( \mu \) and variance \( \sigma^2 \).

c) Find a method of moments estimator for the coefficient of variation
\[
CV = \frac{\sigma}{\mu}.
\]

d) Find the asymptotic distribution of the estimator in c). What moments of the \( X \) distribution need to exist?

2) Let \( X_1, \ldots, X_n \) be iid normal with unknown mean \( \mu \) and known variance \( \sigma^2 \).

a) Find the critical region for the Neyman Pearson test at level \( \alpha \in (0, 1) \) for \( H_0 : \mu = \mu_0 \) versus \( H_1 : \mu = \mu_1 \) with \( \mu_0 < \mu_1 \).

b) Determine the power function \( \beta(\mu) \) of this test.
PLEASE NOTE: To pass you must do well enough on both the Probability and the Statistics sections. High performance in one portion does not compensate for insufficient performance on the other.

STATISTICS SECTION

Q1. Let \( X_1, X_2, \ldots, X_n \) be independent and identically distributed random variables having density

\[
f(x) = \begin{cases} 
  e^{-(x-\theta)} & \text{if } x > \theta \\
  0 & \text{if } x \leq \theta
\end{cases}
\]

for some \( \theta \in (-\infty, \infty) \).

(a) State the Factorization Theorem for sufficient statistics.

(b) Find a one-dimensional sufficient statistic for \( \theta \).

(c) Find a 95% confidence interval for \( \theta \).

(d) Derive the likelihood ratio test of the null hypothesis that \( \theta \geq 0 \) against the alternative that \( \theta < 0 \).

Q2. Outputs \( X_1, X_2, \ldots, X_n \) from a physical device are independent and identically distributed random variables having exponential distribution with (unknown) mean \( \lambda^{-1} \). A measuring device records the values of the \( X_j \) as long as \( X_j < c \), for some known threshold \( c > 0 \). If \( X_j \geq c \) then the device returns the value \( c \). Define

\[
S_n = \sum_{j=1}^{n} X_j I(X_j < c), \quad T_n = \sum_{j=1}^{n} I(X_j \geq c),
\]

where \( I(A) \) denotes the indicator of the event \( A \).

(a) Write down the likelihood function of the observed values in terms of \( S_n \) and \( T_n \).
(b) Show that the Maximum Likelihood Estimator of $\lambda$ is

$$\hat{\lambda} = \frac{n - T_n}{S_n + cT_n}.$$ 

(c) Find the joint asymptotic distribution of $(S_n, T_n)$. 

Hint:

$$\int_0^c x\lambda e^{-\lambda x} dx = \lambda^{-1} \left(1 - (1 + c\lambda)e^{-c\lambda}\right)$$

and

$$\int_0^c x^2\lambda e^{-\lambda x} dx = \lambda^{-1} \left(2 - (2 + 2c\lambda + c^2\lambda^2)e^{-c\lambda}\right).$$

(d) Using the result of the previous part, or otherwise, find the asymptotic distribution of $\hat{\lambda}$. 