McGill University

Part A Examination in Statistics

Theory Paper

Department of Mathematics & Statistics

Date: 19 August 2002
Time: 13:00

Instructions

- Answer two questions out of Section P.
- Answer four questions out of Section S.
- All questions are weighted equally.
- Each question will be assessed independently by at least two members of the statistics group, and the final result determined after discussion within the Part A Exam Subcommittee.
- Your best 2 answers from Section P and best 4 answers from Section S will be used for the purpose of grading.
- Your answers to other questions may be used as assessment aids.
- Good luck!

This exam comprises the cover and 9 questions on 4 pages.
Section P: Answer two questions out of questions P1 to P3.

P1. (a) State, but do not prove Fubini's Theorem
(b) Prove, with careful justification of each step, that for any distribution function \( F(x) = P(X \leq x) \) and any \( a \geq 0 \), we have
\[
\int_{-\infty}^{\infty} [F(x + a) - F(x)] \, dx = a
\]
(c) Show that if \( X \) and \( Y \) are independent random variables, then
\[
E(|X + Y|) < \infty \quad \text{implies that} \quad E(|X|) < \infty \quad \text{and} \quad E(|Y|) < \infty.
\]

P2. Let \( \{X_n\} \) be a sequence of identically distributed random variable with finite mean.

(a) Prove that
\[
\frac{1}{n} E\left[ \max_{1 \leq j \leq n} |X_j| \right] \rightarrow 0 \quad \text{as} \quad n \rightarrow \infty.
\]
(b) Prove that for any \( y > 0 \)
\[
P\left( \cup_{j=1}^{n} \{ \omega : \frac{X_j(\omega)}{n} > y \} \right) \rightarrow 0 \quad \text{as} \quad n \rightarrow \infty.
\]

P3. Let \( Y \) be an integrable random variable on the probability space \( (\Omega, \mathcal{F}, P) \).
(Integrable means \( E(|Y|) < \infty \).

(a) Show that
\[
\lim_{n \rightarrow \infty} \int_{\{\omega : |Y(\omega)| > n\}} |Y(\omega)| \, P(d\omega) = 0.
\]
(b) Prove that for any \( \varepsilon > 0 \) there is a \( \delta > 0 \) such that
\[
\int_{A} |Y(\omega)| \, P(d\omega) < \varepsilon \quad \text{if} \quad P(A) < \delta.
\]
(Hint: Integrate \( I_A |Y| \) over \( \{|Y| > n\} \) and \( \{|Y| \leq n\} \).)
Section S: Answer four questions out of questions S1 to S6.

S1. Certain machines each produce a batch of $n$ items in a day. Machines with a certain defect will produce flawed items with probability $p$. Flawed items are produced independently by a given machine. Machines without the defect will not produce flawed items. A machine can only be known to be defective if flawed items are observed in a batch it has produced.

(a) Suppose that the probability of observing that a batch contains flawed items is 1 when it contains at least one flawed item and 0 when it does not. One such batch is observed. Find the expectation and variance of the number of flawed items in this batch in terms of $n$ and $p$.

(b) Suppose now that the probability of observing that a batch contains flawed items is $x/n$, where $x$ is the number of flawed items in the batch. One such batch is observed. Find the expectation and variance of the number of defective items in this batch in terms of $n$ and $p$.

S2. Let $X_1, \ldots, X_n$ be a random sample from density
\[ f_\mu(x) = \exp (\mu - x) \mathbb{1}[x \geq \mu] \]
with $-\infty < \mu < \infty$. Let $X_{(1)} = \min_i X_i$, $\overline{X} = n^{-1} \sum_{i=1}^n X_i$ and $S^2 = (n-1)^{-1} \sum_{i=1}^n (X_i - \overline{X})^2$.

(a) Show that $X_{(1)}$ is a sufficient statistic for $\mu$.

(b) Show that $X_{(1)}$ is a complete statistic for $\mu$.

(c) Show that $S^2$ is an ancillary statistic for $\mu$.

(d) State Basu’s Theorem (without proof) and apply its conclusion to $X_{(1)}$ and $S^2$.

S3. Let $X = (X_1, \ldots, X_n)$ be a random sample from a $N(\mu, \sigma^2)$ distribution, $\mu$ and $\sigma^2$ unknown. The density of $X_i, i = 1, \ldots, n$, is given by
\[ f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[ -\frac{1}{2\sigma^2} (x - \mu)^2 \right] \]
for $-\infty < x < +\infty, -\infty < \mu < +\infty$ and $\sigma^2 > 0$. We wish to test $H_0 : \mu \leq \mu_0$ against $H_1 : \mu > \mu_0$.

Write down the Likelihood ratio statistic (LRS) and show that an $\alpha$-level test based on this LRS is given by a $t$-test.
You may assume it known that the maximum likelihood estimators of \( \mu \) and \( \sigma^2 \) are respectively
\[
\hat{\mu} = \bar{X} \quad \text{and} \quad \hat{\sigma}^2 = \frac{n-1}{n} S^2
\]
where \( \bar{X} = n^{-1} \sum_{i=1}^{n} X_i \) and \( S^2 = \frac{(n-1)}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2 \).

S4. Let \( X_1, \ldots, X_m \) be independent and identically distributed (iid) Exponential(\( \beta \)) random variables, with density
\[
f_\beta(x) = \frac{1}{\beta} \exp\left(-\frac{x}{\beta}\right) \mathbb{1}[x > 0].
\]

(a) Find a pivotal \( 1 - \alpha \) confidence interval \([\beta_L^{\text{pivot}}, \beta_U^{\text{pivot}}]\) for \( \beta \).

(b) Provide and justify conditions for the interval in (a) to have shortest expected length amongst such pivotal intervals. Be specific.

(c) Find a pivotal \( 1 - \alpha \) confidence interval \([\theta_L^{\text{pivot}}, \theta_U^{\text{pivot}}]\) for \( \theta = \beta^{-1} \).

(d) Provide and justify conditions for the interval in (c) to have shortest expected length amongst such pivotal intervals. Be specific.

S5. Let \( X \) be a single observation taken from the density
\[
f(x) = \begin{cases} \frac{2x}{m} & \text{if } 0 \leq x \leq m \\ \frac{2(1-x)}{1-m} & \text{if } m < x \leq 1 \\ 0 & \text{otherwise.} \end{cases}
\]

where \( 0 < m < 1 \) is an unknown parameter. (The density \( f(x) \) looks like a triangle of height 2 on the unit interval, with \( m \) indicating the projection of its apex on the \( x \)-axis.) Fix \( 0 < \alpha < 1 \).

(a) Show that the cdf of \( X \) is given by
\[
F_m(x) = \begin{cases} x^2/m & \text{if } 0 \leq x \leq m \\ (2x - x^2 - m)/(1-m) & \text{if } m < x \leq 1. \end{cases}
\]

(b) State the condition(s) that must be satisfied in order for the statistical method (i.e., use of the CDF as a pivot) to be applicable to produce a confidence interval.

(c) Based on the single observation \( X \), use the statistical method to find an equal-tailed \( 1 - \alpha \) confidence interval for \( m \) for any \( \alpha \in (0, 1/2) \). Justify the use of the statistical method in this case.
S6. Let $Y_n$ be a sequence of random variables that satisfies

$$\sqrt{n}(Y_n - \theta) \xrightarrow{d} Z \sim N(0, \sigma^2).$$

where $\xrightarrow{d}$ indicates convergence in distribution (weak convergence). For a given function $g$ and a specific value of $\theta$, suppose that $g'(\theta)$ exists and is different from 0.

Show that

$$\sqrt{n}[g(Y_n) - g(\theta)] \xrightarrow{d} Z^* \sim N(0, g'(\theta)\sigma^2).$$

End of Theory Paper.