GX 2112

OLLSCOIL NA hÉIREANN, GAILLIMH
NATIONAL UNIVERSITY OF IRELAND, GALWAY

SEMESTER II EXAMINATIONS 2002-2003

MODULE CODE: MA 236
MODULE: STATISTICAL INFERENCE

External Examiner: Dr. D. Harrington
Internal Examiner: Dr. J.N. Sheahan

INSTRUCTIONS: **Answer the ten questions in PART A (30 marks)**
and
two of the questions in PART B (35 marks each).

DURATION: Two hours

PART A

[Multiple choice. 30 marks] In each of questions **A1.** through **A10.** below, **write down one** choice of answer. For example, if in **A1.** below you think **A)** is the answer, you would write in your answer book **A1. A)** .

A1. If $\hat{\theta}$ is an estimator of a parameter θ satisfying $E(\hat{\theta}) = \theta$, **then** $\hat{\theta}$ is said to be
A) an unbiased estimator of θ
B) the minimum variance estimator of θ
C) the maximum likelihood estimator of θ
D) a sufficient statistic for θ
E) a consistent estimator of θ.

A2. **Which** of the following is/are always true?
A) $\sum_{i=1}^n (X_i - \mu)^2 = \sum_{i=1}^n (X_i - \overline{X})^2 + n(\overline{X} - \mu)^2$
B) $\sum_{i=1}^n (X_i - \mu)^2 = \sum_{i=1}^n X_i^2 - n\mu^2$
C) $\sum_{i=1}^n (X_i - \overline{X})^2 = \sum_{i=1}^n X_i^2 - n\overline{X}^2$
D) If $X_1, X_2, \ldots, X_n$ is a random sample from a $N(\mu, 1)$ distribution, then $\overline{X}$ is the minimum variance unbiased estimator of μ
E) exactly two of A), B), C) and D) are always true
F) exactly three of A), B), C) and D) are always true
G) all four of A), B), C) and D) are always true.

A3. Let $X_1, X_2, \ldots, X_9$ be a random sample of size 9 from $N(4, 9)$. **What** is the distribution of $(\overline{X} - 4)^2$?
A) $N(2,9)$ **B)** $N(2,1)$ **C)** $N(2,3)$ **D)** χ_1^2 **E)** χ_8^2 **F)** χ_9^2 .

A4. Assume that the population of peoples' weights has mean $\mu = 60$ kg. and standard deviation $\sigma = 10$ kg.. Suppose that a footbridge will break down if the total weight on it exceeds $6,700$ kg.. **What** is the approximate probability that it will break down if $n = 100$ people get on it all at once and each of them is carrying a case of beer that weights 5 kg.?
Note: If $Z \sim N(0,1)$, then $P(Z > 1) = 0.1587$, $P(Z > 2) = 0.0228$, $P(Z > 3) = 0.0013$.

A) 0 **B)** 0.1587 **C)** 0.0228 **D)** 0.0013 **E)** 0.05 **F)** 0.95.

A5. Let $\overline{X}_1$ and $\overline{X}_2$ be the means of independent random samples of sizes $n_1 > 1$ and $n_2 > 1$ from a population that has mean μ and (finite) variance σ^2. If we desire that the statistic $w\overline{X}_1 + (1-w)\overline{X}_2$ be the minimum variance unbiased estimator of μ, **what** must w equal?

A) $\frac{\sigma^2}{n_1+n_2}$ **B)** $\sigma^2(n_1+n_2)$ **C)** $\frac{n_1 n_2}{n_1+n_2}$ **D)** $\frac{n_1+n_2}{n_1 n_2}$ **E)** $\frac{n_2}{n_1+n_2}$ **F)** $\frac{n_1}{n_1+n_2}$.

A6. Let a be the minimum number of Galwegians that should be sampled at random so that with probability (at least) 0.95 the sample proportion of smokers will not differ from the unknown population proportion θ of smokers by more than ± 0.04, and let $b \pm c$ be an approx. 95.44% confidence interval for θ when 36 smokers are actually observed in a random sample of $n = 100$ Galwegians. **What** are a and $b \pm c$ respectively?
Note: If $Z \sim N(0,1)$, then $P(Z > 1.96) = 0.025, P(Z > 2) = 0.0228$.

A) 1068, 0.36 ± 0.096 **B)** 1068, 0.36 ± 0.096 **C)** 601, 0.36 ± 0.096 **D)** 601, 0.36 ± 0.048.

A7. Let $X = \#$ heads in n flips of a fair coin. **What** number does $\frac{4X}{n} + 1$ converge to in probability?

A) 1 **B)** 2 **C)** 3 **D)** 4 **E)** 5 **F)** 6 .

A8. Let $X_1, X_2, \ldots, X_n$ be a random sample from an infinite population that has density $f(x;\theta), -\infty < x < \infty$. Let $\hat{\theta}$ be an estimator of θ. Then **it is true that**

A) if $\hat{\theta}$ is an unbiased estimator of θ, then $\hat{\theta}$ must be a consistent estimator of θ
B) if $\hat{\theta}$ is a consistent estimator of θ, then $\hat{\theta}$ must be an unbiased estimator of θ
C) if $\hat{\theta}$ is an estimator of θ that utilizes all the information about θ that is contained in the sample, then $\hat{\theta}$ is called a sufficient statistic
D) exactly two of statements A), B), and C) must be true
E) all three of statements A), B), and C) must be true
F) none of statements A), B), and C) must be true.

A9. Let $X_1, X_2, \ldots, X_n$ be a random sample of size $n > 1$ from a population that has mean μ and variance σ^2. Then **we have**

A) $E\left[\overline{X}^2\right] = \frac{\sigma^2}{n} + \mu^2$
B) $E(\overline{X} - \mu)^2 = \sigma^2$
C) $E\left[\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \mu)^2\right] = \sigma^2$
D) exactly two of A), B), and C) are true
E) all three of A), B), and C) are true.

A10. Suppose that based on a random sample of size 10 and the formula $\bar{x} \pm t_{n-1\,\alpha}\frac{s}{\sqrt{n}}$, a 95% confidence interval for the mean μ of a population was found to be $50 < \mu < 60$. We can then **say that:**

A) the probability is 0.95 that μ lies in the interval $(50, 60)$
B) of all possible samples of size 10 that could be taken from the population, 95% of the intervals that would be obtained (using the same formula as above) would contain μ
C) 95% of the means of all possible samples of size 10 that could be taken from the population would lie in the interval $(50, 60)$
D) if the appropriate t-test of the alternatives $H_0 : \mu = 53$ versus $H_1 : \mu \neq 53$ was conducted using a level of significance $\alpha = 0.05$, H_0 would be rejected
E) exactly two of statements A), B), C) and D) are true
F) exactly three of statements A), B), C) and D) are true
G) all four of statements A), B), C) and D) are true.

PART B

B1. Throughout this question, assume that the daily profits X (ignore units) of a certain company have a normal distribution with unknown mean θ and standard deviation $\sigma = 8$, and that we wish to test the null hypothesis $H_0 : \theta \geq 100$ versus the alternative $H_1 : \theta < 100$. Some probabilities relating to a standard normal variable Z that you will need below are $P(Z < -1) = 0.1587$, $P(Z < -2) = 0.0228$ and $P(Z < -3) = 0.0013$.

a. *[15 marks]* Suppose that the profits over a random sample of $n = 64$ days will be observed by two analysts. Analyst A will reject H_0 if the sample mean daily profits satisfies $\overline{x} < 98$, while B will reject H_0 if $\overline{x} < 99$. **Calculate,** and **exhibit** in a table, the values of the power functions of the tests used by A and B for $\theta = 99, 100$ and 101. **Draw** rough graphs of the two power functions. (You do not need to use graph paper, but label your axes, and clearly show which function is uniformly at least as large as the other.)

b. *[10 marks]* The uniformly most powerful test has the form "reject H_0 if $\overline{x} < c$". **Find** n *and* c so that the rule "reject H_0 if $\overline{x} < c$" will have P(reject H_0 when $\theta = 100) = 0.0228$ *and* P(reject H_0 when $\theta = 98) = 0.9772$.

c. *[10 marks]* Examine each of statements (i) and (ii) below separately and then **prove that it is true or else that it is false.** *Hint:* The UMP level α test based on n observations rejects H_0 if $\overline{x} < c$ with $c = 100 - z_\alpha \frac{\sigma}{\sqrt{n}}$.

(i) For any fixed sample size n, and any level of significance α, the power of the uniformly most powerful test is larger at $\theta = 97$ than at $\theta = 98$.

(ii) For any fixed sample size n, and any two fixed numbers α_1 and α_2 satisfying $0 < \alpha_1 < \alpha_2 < 1$, then the power at $\theta = 97$ of the uniformly most powerful test is smaller if a level of significance α_1 is used than if a level of significance α_2 is used.

B2.

a. *[3+7 = 10 marks]* Let $X_1, X_2, \ldots, X_n$ be a random sample from an infinite population that has mean μ and (finite) variance σ^2. **Show** that $\overline{X}$ and $S^2 := \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ are unbiased estimators of μ and σ^2, respectively. (You may accept that $Var(\overline{X}) = \frac{\sigma^2}{n}$.)

b. Let $X_1, X_2, \ldots, X_n$ be a random sample from the Bernoulli(θ) distribution, $0 < \theta < 1$.

(i) *[10 marks]* **Show** that the sample proportion $\hat{\theta}$ is the minimum variance unbiased estimator of θ.

(ii) *[6 marks]* **Show** that $\hat{\theta}$ is a consistent estimator of θ. [You may use the following version of Chebychev's Inequality: If U is a random variable with mean $E(U)$ and finite variance $Var(U)$ then for any $\epsilon > 0$, $P(|U - E(U)| \geq \epsilon) \leq \frac{Var(U)}{\epsilon^2}$.]

(iii) *[6 marks]* **Show** that $\hat{\theta}$ is the maximum likelihood estimator of θ.

(iv) *[3 marks]* Briefly **describe** an application of the above theory of estimation of a Bernoulli parameter θ.

B3.

a. *[15 marks]* Let X be one random observation from an exponential density
$f(x;\theta) = \begin{cases} \frac{1}{\theta}\exp(-\frac{x}{\theta}), & x > 0 \\ 0, \text{ otherwise.} \end{cases}$ with $\theta > 0$. **Find** explicitly the uniformly most powerful test of size α of $H_0 : \theta = 1,\ H_1 : \theta > 1$ by first using the Neyman Pearson lemma to test $H_0 : \theta = 1,\ H_1' : \theta = \theta_1$ where θ_1 is any fixed number greater than 1.

b. Recall that the density of a $N(\theta, 1)$ random variable is
$f_X(x;\theta) = \frac{1}{\sqrt{2\pi}}\exp\{-\frac{1}{2}(x-\theta)^2\}, -\infty < x < \infty$. Let $X_1, X_2, \ldots, X_n$ be i.i.d. $N(\theta, 1)$ random variables. We desire to use the likelihood ratio (LR) procedure to test the alternatives $H_0 : \theta = \theta_0,\ H_1 : \theta \neq \theta_0$ at size α.

(i) *[7 marks]* **Show** that the maximum likelihood estimate of θ over the entire parameter space $\Omega = \{\theta \mid -\infty < x < \infty\}$ is $\bar{x}$.

(ii) *[13 marks]* **Find** the LR test of the alternatives above. Ensure that you show how to make the test have size α. *Hint:* First show that the value of the likelihood ratio statistic is $\exp\{-n(\bar{x} - \theta_0)^2/2\}$.