GX 2031

OLLSCOIL NA hÉIREANN, GAILLIMH
NATIONAL UNIVERSITY OF IRELAND, GALWAY

SEMESTER II EXAMINATIONS 2003–2004

MODULE CODE: MA 236
MODULE: STATISTICAL INFERENCE

External Examiner: Dr. T.C. Bailey
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.

A1. If X_1, X_2, X_3 is a random sample of size 3 from a population that has mean θ, what should the number a be if it is desired that $\hat{\theta} = X_1 + X_2 + aX_3$ be an unbiased estimator of θ?

(a) 1 (b) -1 (c) 2 (d) -2 (e) 3 (f) -3.

A2. A random sample of size $n = 2$ will be taken with replacement from a population that has mean 50 and variance 4. Let $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ be the variance of the random sample. **What** is $E(S^2)$?

(a) 54 (b) 50 (c) 42 (d) $\sqrt{2}$ (e) 4 (f) 2 (g) 1 .

A3. Let $\overline{X}_1$ and $\overline{X}_2$ be the means of two independent random samples of sizes $n_1 > 1$ and $n_2 > 1$ from an infinite population that has mean μ and (finite) variances σ_1^2 and σ_2^2, respectively, where $\sigma_1^2 > 0$, $\sigma_2^2 > 0$, $\sigma_1^2 \neq \sigma_2^2$. Thus $\overline{X}_1$ and $\overline{X}_2$ are independent, and of course $E(\overline{X}_1) = \mu$, $Var(\overline{X}_1) = \frac{\sigma_1^2}{n_1}$, $E(\overline{X}_2) = \mu$ and $Var(\overline{X}_2) = \frac{\sigma_2^2}{n_2}$. If we desire that the statistic $\hat{\theta} = w\overline{X}_1 + (1-w)\overline{X}_2$ be the minimum variance unbiased estimator of μ, **what** must w equal?

(a) $\dfrac{\sigma_1^2 + \sigma_2^2}{n_1 + n_2}$ (b) $(\sigma_1^2 + \sigma_2^2)(n_1 + n_2)$ (c) $\dfrac{\frac{\sigma_2^2}{n_2}}{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$ (d) $\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$ (e) $\frac{\sigma_1^2}{n_1}$ (f) $\dfrac{n_1}{n_1 + n_2}$.

A4. If $X_1, X_2, \ldots, X_n$ are independent random variables having the χ_1^2 distribution (i.e. chi-squared distribution with 1 degree of freedom), **what** is the limiting distribution of $\dfrac{\overline{X} - 1}{\sqrt{2/n}}$?

Hint: The mean and variance of a χ_ν^2 random variable are ν and 2ν, respectively.

(a) χ_2^2 (b) χ_1^2 (c) $N(0,1)$ (d) $N(0,2)$ (e) $N(1,2)$ (f) $N(1,1)$.

A5. If a random sample of size n is taken without replacement from the finite set $\{1,2,\ldots,K\}$, accept that $E(X_{(n)}) = \frac{n(K+1)}{n+1}$, where $X_{(n)}$ is the largest of the n observations that will be picked. Suppose now that a country's military intelligence knows that an enemy has built new tanks numbered $1,2,\ldots,K$, and wants to estimate the unknown constant K. If three of the tanks are captured and their serial numbers are 101, 50 and 210, **what** is an unbiased estimate of K?

(a) 210 (b) 361 (c) 279 (d) 100 (e) 500 (f) 400.

A6. 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) $E\left[\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\mu)^2\right] = \sigma^2$ (b) $E\left[\frac{1}{n}\sum_{i=1}^{n}X_i^2\right] = \sigma^2+\mu^2$ (c) $E\left[\overline{X}^2\right] = \frac{\sigma^2}{n}+\mu^2$

(d) exactly two of (a), (b) and (c) are true (e) all three of (a), (b) and (c) are true.

Questions A7.–A10. below concern the following problem: 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 random variable Z that you may need are $P(Z > 1) = 0.1587$, $P(Z > 2) = 0.0228$ and $P(Z > 3) = 0.0013$.

A7. Suppose that the profits over a random sample of $n = 64$ days will be observed by an analyst and that he will reject H_0 if the sample mean daily profits satisfies $\overline{x} < 98$. **What** is the value of the power function of the test if in fact $\theta = 99$.

(a) 0.0228 (b) 0.4772 (c) 0.1587 (d) 0.3413 (e) 0.9772 (f) 0.5.

A8. The uniformly most powerful test has the form "reject H_0 if $\overline{x} < c$". If the size of the test is to be $\alpha = 0.0228$, **what** should c be? (Assume that the analyst has $n = 64$.)

(a) 96 (b) 97 (c) 98 (d) 99 (e) 100 (f) 101.

A9. **What** values of n and c should the analyst use if it is desired that the rule "reject H_0 if $\overline{x} < c$" will have $P(\text{reject } H_0 \text{ when } \theta = 100) = 0.0228$ *and* $P(\text{reject } H_0 \text{ when } \theta = 98) = 0.9772$.

(a) $n = 100$ and $c = 98$ (b) $n = 100$ and $c = 100$ (c) $n = 256$ and $c = 99$

(d) $n = 200$ and $c = 98$ (e) $n = 64$ and $c = 99$ (f) $n = 36$ and $c = 98$.

A10. Consider the following two statements (i) and (ii):

(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 $0 < \alpha_1 < \alpha_2 < 1$, 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.

Choose one answer below regarding the veracity of these statements.

Hint: The UMP level α test based on n observations rejects H_0 if $\overline{x} < c$ with $c = 100 - z_\alpha \frac{\sigma}{\sqrt{n}}$.

(a) Both of statements (i) and (ii) are true
(b) Statement (i) is true but statement (ii) is false
(c) Statement (i) is false but statement (ii) is true
(d) Both of statements (i) and (ii) are false.

PART B

B1.

(a) *[10 marks]* Let $X_1, X_2, \ldots, X_n$ be a random sample from a $N(\mu, \sigma^2)$ distribution. Accept that the random variable $\sum_{i=1}^{n}(X_i - \bar{X})^2$ can be written as $\sum_{i=1}^{n-1} Y_i^2$ where the Y_i are iid $N(0, \sigma^2)$ random variables. **Show** that the random variable $\dfrac{(n-1)S^2}{\sigma^2}$ has a chi-squared distribution with $n-1$ degrees of freedom, where $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$. Ensure that you state clearly any relevant distribution results/properties used in your proof.

(b) Suppose that the return X on a certain stock investment is an $N(\mu, \sigma^2)$ random variable and that a financial analyst has stated that A financial analyst made the following statement concerning the return X (ignore units) on a certain stock investment: "95.44% of the time, the return will be within ± 4 units of the population mean return, μ".

(i) *[5 marks]* **Show** that the value of σ^2 that corresponds to the analyst's statement is 4.0 .
Note: If $Z \sim N(0,1)$, then $P(Z > 1) = 0.1587$, $P(Z > 2) = 0.0228$ and $P(Z > 3) = 0.0013$.

In answering (ii) and (iii) below, you'll need the theory in B1(a) above. Also note that $\chi^2_{10\,,\,0.95} = 3.940$, $\chi^2_{10\,,\,0.975} = 3.247$, $\chi^2_{11\,,\,0.95} = 4.575$, $\chi^2_{11,\,0.975} = 3.816$, $\chi^2_{10\,,\,0.025} = 20.483$.

(ii) *[15 marks]* Suppose that you will invest in the stock if there is sufficient statistical evidence to conclude that $\sigma^2 < 4$. A random sample of 11 stock returns showed a sample variance $s^2 = 2.0$. **Carry out** a test of the alternatives $H_0 : \sigma^2 \geq 4$, $H_1 : \sigma^2 < 4$ using a level of significance $\alpha = 0.05$. Ensure as usual that you give a clear conclusion.

(iii) *[5 marks]* **Find** a two-sided 95% confidence interval for σ^2.

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) In the study of the effectiveness of a new migraine drug, it is desired to estimate θ, the population proportion of migraine sufferers who experience less pain after being put on the drug. Each individual in a random sample of migraine sufferers who volunteered to take the new drug for a month will respond 'yes' or 'no' to the question "are you experiencing less pain since being put on the drug?".

(i) *[10 marks]* **Show** that the sample proportion $\hat{\theta}$ of 'yes' responses 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]* **Is** $\frac{n+1}{n}\hat{\theta}$ a consistent estimator of θ? Briefly justify your answer.

B3.

(a) Let X be one random observation from an exponential density

$$f(x;\theta) = \begin{cases} \frac{1}{\theta}\exp(-\frac{x}{\theta}), & x > 0 \\ 0, & \text{otherwise} \end{cases} \quad \text{with } \theta > 0.$$

(i) *[8 marks]* Use the Neyman Pearson Lemma to **find** explicitly the uniformly most powerful test of size α of the alternatives $H_0 : \theta = 1,\ H_1 : \theta = \theta_1$ where θ_1 is any fixed number greater than 1.

(ii) *[8 marks]* Since the test you obtained in (i) above is independent of the number θ_1, as long as $\theta_1 > \theta_0$, the same test is uniformly most powerful of size α for the alternatives $H_0 : \theta = 1,\ H_1 : \theta > 1$. **Derive** an expression for the power function of this test of these alternatives $H_0 : \theta = 1,\ H_1 : \theta > 1$.

(b) Recall that the density of a $N(\theta, 1)$ random variable is

$$f_X(x;\theta) = \frac{1}{\sqrt{2\pi}}\exp\{-\tfrac{1}{2}(x-\theta)^2\}, -\infty < x < \infty.$$

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

(ii) *[11 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 $\exp\{-n(\bar{x} - \theta_0)^2/2\}$.
