Imperial College London

Module:	MATH70078
Setter:	Young
Checker:	Tang
Editor:	Varty
External:	Everitt
Date:	November 27, 2024

MSc EXAMINATIONS (STATISTICS)
January 2025

MATH70078 Fundamentals of Statistical Inference

Time: 2 hours

Setter's signature	Checker's signature	Editor's signature
. . . .	. . . .	. . . .

Questions on next page

1. Explain carefully what is meant by *minimax* and *Bayes* decision rules. How is a Bayes decision rule found in general?

ANSWER: (Seen) Given a loss function $L(\theta, a)$, the risk function of a decision rule $d = d(Y)$ is defined for $\theta \in \Omega_\theta$,the parameter space, by

$$R(\theta, d) = E_\theta L(\theta, d(Y)),$$

where E_θ denotes expectation with respect to distribution of Y, assuming Y has distribution defined by parameter value θ.

The maximum risk of a decision rule d is $MR(d) = \sup_{\theta \in \Omega_\theta} R(\theta, d)$. A decision rule is **minimax** if it minimises maximum risk: $MR(d) \le MR(d')$ for all decision rules d'.
[2 marks]

ANSWER: (Seen) Given a prior $\pi(\theta)$, the Bayes risk of a decision rule d is $r(\pi, d) = \int_{\Omega_\theta} R(\theta, d)\pi(\theta)d\theta$. A decision rule d is a **Bayes rule** with respect to the given prior $\pi(\theta)$ if it minimises the Bayes risk:

$$r(\pi, d) = \inf_{d'} r(\pi, d').$$

[2 marks]

ANSWER: (Seen) In general, a Bayes decision rule is found by minimising the expected posterior loss: for given data y, we identify $d(y)$ as the action which minimises

$$\int_{\Omega_\theta} L(\theta, d(y))\pi(\theta|y)d\theta,$$

where $\pi(\theta|y) \propto \pi(\theta)f(y;\theta)$ is the posterior density of θ, with $\pi(\theta)$ the assumed prior and $f(y;\theta)$ the likelihood function. [3 marks]

Consider testing two simple hypotheses $H_0 : \theta = \theta_0$ against $H_1 : \theta = \theta_1$, with action space $\mathcal{A} = \{0, 1\}$, where $a = 0$ and $a = 1$ correspond respectively to acceptance and rejection of H_0. Consider the loss function

$$L(\theta, a) = \begin{cases} K_0 a, & \theta = \theta_0 \\ K_1(1 - a), & \theta = \theta_1, \end{cases}$$

for constants $K_0, K_1 > 0$. Assume prior probabilities $\pi_0 > 0$ and $\pi_1 = 1 - \pi_0 > 0$ for H_0 and H_1 respectively.

(i) Derive the Bayes rule for this inference problem.

[This question continues on the next page . . .]

ANSWER: (Seen similar) Let $d(y) = 0$ if H_0 is accepted and $d(y) = 1$ otherwise. The expected posterior loss of a rule d is then

$$E\{L(\theta, d)|y\} = \begin{cases} K_1 p(\theta = \theta_1|y), & d(y) = 0, \\ K_0 p(\theta = \theta_0|y), & d(y) = 1. \end{cases}$$

The resulting Bayes rule minimises this expected posterior loss, so is

$$d_\pi(y) = \begin{cases} 0, & K_1 p(\theta = \theta_1|y) < K_0 p(\theta = \theta_0|y), \\ 1, & \text{otherwise.} \end{cases}$$

This is equivalent to accepting H_0 if

$$\frac{f(y|\theta_1)}{f(y|\theta_0)} < \frac{K_0\pi_0}{K_1\pi_1},$$

and rejecting H_0 otherwise. [5 marks]

(ii) Interpret the Bayes rule in terms of the frequentist approach to testing. Is it a likelihood ratio test? What is the critical value of the corresponding test statistic?

ANSWER: (Seen similar) From the frequentist perspective, this Bayes rule corresponds to a likelihood ratio test (so a most powerful test) that rejects the null hypothesis if the likelihood ratio $\Lambda(y) = f(y;\theta_1)/f(y;\theta_0) \geq C$, with critical value $C = \frac{K_0\pi_0}{K_1\pi_1}$. [3 marks]

(iii) Show that any likelihood ratio test which rejects H_0 if $f(y;\theta_1)/f(y;\theta_0) \geq C, C \geq 0$, is admissible for the assumed form of loss function.

ANSWER: (Seen similar) Consider a likelihood ratio test that rejects H_0 if $\Lambda(y) \geq C, C \geq 0$. From the above, this is a Bayes rule corresponding to $\pi_0 = (CK_1/K_0)/(1 + CK_1/K_0)$. Then by the Theorem from lectures that says that a Bayes rule for a finite decision problem, with parameter space $\Omega_\theta = \{\theta_0, \theta_1\}$, is admissible, such a likelihood ratio test is admissible. [5 marks]

(iv) Let α and β be the Type 1 and Type 2 error probabilities of such a likelihood ratio test. Show that if the critical value C of the test is fixed so that $K_0\alpha = K_1\beta$, then it is also a minimax test.

ANSWER: (Seen similar) Let $\alpha = p(\Lambda(Y) \geq C; \theta_0)$ and $\beta = p(\Lambda(Y) < C; \theta_1)$ be the Type 1 and Type 2 error probabilities of the likelihood ratio test. Then the risk function is seen to be given by $K_0\alpha$ if $\theta = \theta_0$ and $K_1\beta$ if $\theta = \theta_1$. This risk does not depend on θ, so that the decision rule is an equaliser rule, if $K_0\alpha = K_1\beta$. Then by the Theorem of lectures that says that a Bayes rule which has constant risk function is minimax, we reach the desired conclusion. [5 marks]

[Total 25 marks]

2. Let $Y_1, \ldots, Y_n$ be independent, identically distributed normal random variables, with mean μ and variance $\mu^2, \mu > 0$.

(i) Show that this model constitutes an example of a curved exponential family. Why would the Conditionality Principle be relevant to inference on μ?

ANSWER: (Seen)

The log likelihood is, apart from an additive constant,

$$l(\mu) = -n\log(\mu) - \frac{1}{2\mu^2}T_2 + \frac{1}{\mu}T_1,$$

where $(T_1, T_2) = (\sum_{i=1}^n Y_i, \sum_{i=1}^n Y_i^2)$. The form of the log-likelihood identifies a curved (2,1) exponential family with natural parameters $(\pi_1, \pi_2) = (1/\mu, -1/(2\mu^2))$. We have

$$E(T_1) = n\mu, E(T_2) = 2n\mu^2.$$

We can write $Y_i = \mu Z_i$, where $Z_1, \ldots, Z_n$ are IID $N(1, 1)$. Then $A \equiv T_1/\sqrt{T_2}$ may be written as a function of $Z_1, \ldots, Z_n$, and is therefore distribution constant. We have (taking a 1-1 function of (T_1, T_2), which is minimal sufficient) that (T_2, A) is minimal sufficient, with A distribution constant. Therefore A is ancillary and the Conditionality Principle would indicate that inference about μ should be based on the conditional distribution of $T_2|A = a$. [6 marks]

(ii) Calculate the Cramér-Rao lower bound for the variance of an unbiased estimator of μ.

ANSWER: (Seen similar)

We have

$$\frac{\partial l(\mu)}{\partial\mu} = -\frac{n}{\mu} + \frac{1}{\mu^3}T_2 - \frac{1}{\mu^2}T_1,$$

so, differentiating again and calculating $I_n(\mu) = E(-\partial^2 l(\mu)/\partial\mu^2)$, gives $3n/\mu^2$, using $E(T_1), E(T_2)$. The Cramér-Rao lower bound tells us that the variance of an unbiased estimator of μ is no less than $1/I_n(\mu) = \mu^2/(3n)$. [4 marks]

(iii) Show that $\hat{\mu}_1 = \bar{Y} = n^{-1}\sum_{i=1}^n Y_i$ and $\hat{\mu}_2 = c_n\sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}$, with

$$c_n = \Gamma\left(\frac{n-1}{2}\right) / \left\{\sqrt{2}\Gamma\left(\frac{n}{2}\right)\right\},$$

are both unbiased estimators of μ.

ANSWER: (Seen similar) We have that $\bar{Y}$ is distributed as $N(\mu, \mu^2/n)$ and

[This question continues on the next page . . .]

$\sum_{i=1}^{n}(Y_i - \bar{Y})^2)/\mu^2$ is distributed as χ^2_{n-1}, independently of $\bar{Y}$. So $E(\hat{\mu}_1) = \mu$ and $E\left\{\sqrt{\sum_{i=1}^{n}(Y_i - \bar{Y})^2)}\right\} = \mu/c_n$, by the hint, and so $E(\hat{\mu}_2) = \mu$, as required to show that $\hat{\mu}_1, \hat{\mu}_2$ are both unbiased. [5 marks]

(iv) Consider the class of unbiased estimators

$$d = \alpha\hat{\mu}_2 + (1 - \alpha)\hat{\mu}_1, 0 \le \alpha \le 1.$$

Show that the value of α which minimises the variance of d is

$$\alpha^* = \frac{1}{1 + n\{(n-1)c_n^2 - 1\}}.$$

ANSWER: (Seen similar) Since $\hat{\mu}_1$ and $\hat{\mu}_2$ are independent

$$\text{var}(d) = \alpha^2\text{var}(\hat{\mu}_2) + (1 - \alpha)^2\text{var}(\hat{\mu}_1).$$

We have $\text{var}(\hat{\mu}_1) = \mu^2/n$ and $\text{var}(\hat{\mu}_2) = E(\hat{\mu}_2^2) - \{E(\hat{\mu}_2)\}^2 = c_n^2 E\{\sum_{i=1}^{n}(Y_i - \bar{Y})^2)\} - \mu^2 = \mu^2 c_n^2(n-1) - \mu^2$, since $E(\chi^2_{n-1}) = n-1$. Setting the derivative of var(d) with respect to α to zero yields the given expression α^*. [5 marks]

(v) Let $d^* = \alpha^*\hat{\mu}_2 + (1 - \alpha^*)\hat{\mu}_1$. Show that $n\text{var}(d^*) \to \mu^2/3$ as $n \to \infty$. In terms of asymptotic variance, is d^* preferable to the maximum likelihood estimator $\hat{\mu}$?

ANSWER: (Seen similar) Then $n\text{var}(d^*) = \mu^2\{\alpha^{*2}e_n + (1 - \alpha^*)^2\}$, where $e_n = n\{(n-1)c_n^2 - 1\}$. We have $\alpha^* \to 2/3, 1 - \alpha^* \to 1/3$ and $e_n \to 1/2$, by hint. Putting together, $n\text{var}(d^*) \to \mu^2(\frac{4}{9} \cdot \frac{1}{2} + \frac{1}{9}) = \mu^2/3$. Recall that the MLE has $\text{var}(\hat{\mu}) \to$ Cramér-Rao lower bound. Therefore, asymptotically d^* has same variance as the MLE. Note that d^* is unbiased for all n, by construction, while the MLE is only asymptotically unbiased. Since they have the same asymptotic variances, d^* can be considered preferable.

[5 marks]

[Total 25 marks]

[You may note that if $V \sim \chi^2_n$, then $E(\sqrt{V}) = \sqrt{2}\Gamma\left(\frac{n+1}{2}\right)/\Gamma\left(\frac{n}{2}\right)$. Also, you may assume that $n\{(n-1)c_n^2 - 1\} \to 1/2$ as $n \to \infty$.]

3. (i) Explain in detail the optimality notions of *uniformly most powerful* and *uniformly most powerful unbiased* tests of a null hypothesis $H_0 : \theta \in \Omega_0$ against an alternative hypothesis $H_1 : \theta \in \Omega_1$.

ANSWER: (Seen) Given the parameter space Ω_θ for a parameter θ, consider testing the null hypothesis $H_0 : \theta \in \Theta_0$ against alternative $H_1 : \theta \in \Theta_1$, where Θ_0 and Θ_1 are *disjoint* subsets of Ω_θ.

Define a test in terms of its test function $\phi(Y)$, so that $\phi(y)$ is the probability that H_0 is rejected when $Y = y$, and define the power function of the test ϕ to be $w(\theta) = E_\theta\{\phi(Y)\}$.

A **uniformly most powerful** (UMP) test of size α is a test $\phi_0(\cdot)$ for which: (a) $E_\theta\{\phi_0(Y)\} \leq \alpha$ for all $\theta \in \Theta_0$; (b) given any other test $\phi(\cdot)$ for which $E_\theta\{\phi(Y)\} \leq \alpha$ for all $\theta \in \Theta_0$, we have $E_\theta\{\phi_0(Y)\} \geq E_\theta\{\phi(Y)\}$ for all $\theta \in \Theta_1$. [4 marks]

ANSWER: (Seen) A test ϕ of $H_0 : \theta \in \Theta_0$ against $H_1 : \theta \in \Theta_1$ is unbiased of size α if $\sup_{\theta\in\Theta_0} E_\theta\{\phi(Y)\} = \alpha$ and $E_\theta\{\phi(Y)\} \geq \alpha$ for all $\theta \in \Theta_1$. A test which is uniformly most powerful amongst the class of all unbiased tests is **uniformly most powerful unbiased** (UMPU): a test $\phi_0(\cdot)$ is UMPU if $E_\theta\{\phi_0(Y)\} \geq E_\theta\{\phi(Y)\}$ for all $\theta \in \Theta_1$, and any test $\phi(\cdot)$ which is unbiased of size α. [4 marks]

(ii) Let $Y_1, \dots, Y_n$ be independent, identically distributed $N(\mu, \sigma^2)$.

Explain how to test the null hypothesis $H_0 : \sigma = \sigma_0$ against the alternative $H_1 : \sigma \neq \sigma_0$, for specified σ_0, for the cases: (a) μ is *known*; (b) μ is *unknown*. What optimality properties does the test have in each case?

ANSWER: (Seen)

The joint pdf of $Y = (Y_1, \dots, Y_n)$ is

$$f(y; \mu, \sigma) \propto \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \mu)^2\right\} \tag{1}$$

$$\propto \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n} y_i^2 + \frac{\mu}{\sigma^2}\sum_{i=1}^{n} y_i\right\}. \tag{2}$$

In the case when μ is known, (1) identifies a one-parameter exponential family, with natural parameter $\theta^1 = -\frac{1}{2\sigma^2}$. Testing $H_0 : \sigma = \sigma_0$ against $H_1 : \sigma \neq \sigma_0$ is equivalent to testing $H_0 : \theta^1 = \theta^{1*} = -\frac{1}{2\sigma_0^2}$ against $H_1 : \theta^1 \neq \theta^{1*}$. Then, from lecture theory, a uniformly most powerful unbiased (UMPU) test exists, of the form: reject H_0 if

$$\frac{1}{\sigma_0^2}\sum_{i=1}^{n}(Y_i - \mu)^2 \notin (c_1, c_2),$$

[This question continues on the next page . . .]

where c_1, c_2 are such that

$$P_{\sigma=\sigma_0}\left(\frac{1}{\sigma_0^2}\sum_{i=1}^n (Y_i-\mu)^2 \notin (c_1,c_2)\right) \equiv P(\chi_n^2 \notin (c_1,c_2)) = \alpha,$$

and

$$\frac{d}{d\sigma}P_\sigma\left(\frac{1}{\sigma_0^2}\sum_{i=1}^n (Y_i-\mu)^2 \notin (c_1,c_2)\right)\Bigg|_{\sigma=\sigma_0} = 0,$$

for a UMPU test of size α. [This part was seen on a problem sheet, and it might be noted that the latter condition reduces to $c_1^{n/2}e^{-c_1/2} = c_2^{n/2}e^{-c_2/2}$, but this is not required for full credit].

In the case when μ is unknown, (2) identifies a two-parameter exponential family, with natural parameters $\theta^1 = -\frac{1}{\sigma^2}, \theta^2 = \frac{\mu}{\sigma^2}$. Now we wish to test $H_0 : \theta^1 = \theta^{1*}$ against $H_1 : \theta^1 \neq \theta^{1*}$, with θ^2 nuisance. From theory, a UMPU test exists, of conditional form, conditional on the observed data value of $\sum_{i=1}^n Y_i$, or equivalently $\bar{Y}$. The size α test is: reject H_0 if $\sum_{i=1}^n Y_i^2 \notin (k_1,k_2)$, say, where

$$P_{\sigma=\sigma_0}\left(\sum_{i=1}^n Y_i^2 \notin (k_1,k_2)|\bar{Y}=\bar{y}\right) = \alpha,$$

where $\bar{y}$ is the observed sample mean, and where the derivative of the (conditional) power function is zero at $\sigma=\sigma_0$.

Let $S^2 = \sum_{i=1}^n (Y_i-\bar{Y})^2 = \sum_{i=1}^n Y_i^2 - n\bar{Y}^2$. Then, S^2 is a increasing function of $\sum_{i=1}^n Y_i^2$ for fixed $\bar{Y}$. Then, 'a useful result' gives that the UMPU conditional test is equivalent to the test based on the marginal distribution of S^2. So, reject H_0 if $S^2/\sigma_0^2 \notin (c_1,c_2)$, where now

$$P_{\sigma=\sigma_0}\left(\frac{S^2}{\sigma_0^2} \notin (c_1,c_2)\right) \equiv P(\chi_{n-1}^2 \notin (c_1,c_2)) = \alpha,$$

for a level α test, with the derivative of the power function zero at $\sigma=\sigma_0$. This construction is the *same* as that in the case above, with $n \to n-1$ [8 marks]

(iii) Suppose in (ii) that $\mu = 0$ and consider a Bayesian analysis for σ, with (improper) prior $\pi(\sigma) \propto 1/\sigma$.

Let $\pi(\sigma|Y)$ be the posterior density of σ, given $Y_1,\ldots,Y_n$, and let $L \equiv L(Y)$ satisfy

$$\int_0^L \pi(\sigma|Y)d\sigma = 1-\alpha,$$

the $1-\alpha$ quantile of the posterior distribution of σ.

Show that L is a $1-\alpha$ frequentist confidence limit for σ, with

$$P(\sigma < L(Y)) = 1-\alpha,$$

[This question continues on the next page . . .]

where $P(\cdot)$ here means the probability with respect to the $N(0, \sigma^2)$ sampling distribution of $Y_1, \ldots, Y_n$.

ANSWER: (Unseen) With $S = \sum_{i=1}^n Y_i^2$, the posterior density is of the form

$$\pi(\sigma|Y) \propto \frac{1}{\sigma^{n+1}} \exp\left\{-\frac{1}{2\sigma^2}S\right\}.$$

Then, using the Hint, explicitly we have

$$\pi(\sigma|Y) = \frac{2(S/2)^{n/2}}{\Gamma(n/2)} \frac{1}{\sigma^{n+1}} \exp\left\{-\frac{1}{2\sigma^2}S\right\}.$$

Then, L satisfies

$$\int_0^L \pi(\sigma|Y)d\sigma = 1 - \alpha,$$

and substituting $t = S/\sigma^2$, so that $d\sigma = \frac{S^{1/2}}{2t^{3/2}}dt$ this becomes

$$\int_{S/L^2}^{\infty} \frac{t^{n/2-1}e^{-t/2}}{2^{n/2}\Gamma(n/2)}dt = 1 - \alpha.$$

We recognise the integrand here as the density function of the chi-squared distribution on n degrees of freedom, χ_n^2. So, $P(\chi_n^2 > S/L^2) = 1 - \alpha$, so that S/L^2 is equal to q_α, the α quantile of χ_n^2. Then $L(Y) \equiv \sqrt{S/q_\alpha}$. [6 marks]

ANSWER: (Unseen) From a frequentist perspective, We have $S/\sigma^2 \sim \chi_n^2$. So,

$$P(\sigma < L(Y)) = P(\sigma^2/S < L^2/S) = P(\sigma^2/S < 1/q_\alpha) = P(S/\sigma^2 > q_\alpha) = 1 - \alpha,$$

as required to verify that $L(Y)$ is a $1 - \alpha$ frequentist confidence limit for σ.

[3 marks]

[Total 25 marks]

[Recall that $\int_0^\infty \frac{1}{x^{\kappa+1}} \exp(-\lambda x^{-\alpha})dx = \frac{1}{\alpha\lambda^{\kappa/\alpha}}\Gamma(\kappa/\alpha)$.]

4. Write a brief account, with careful definitions and examples as appropriate, of **TWO** of the following:

(i) difficulties with the likelihood and conditionality principles of statistical inference;

(ii) the importance of James-Stein estimators;

(iii) inference based on Bayes factors;

(iv) large sample testing procedures based on the likelihood .

ANSWER: Entirely descriptive question. All topics **Seen**. Requires extracting and synthesizing material. The course summary notes are available.

Marked as 13+13, capped at 25. For each of the two chosen parts, marking according to: basic definitions etc. 5; appropriate examples, illustration 5; bonus/style 3

[25 marks]

[Total 25 marks]