Imperial College London

Module:	MATH97123
Setter:	Young
Checker:	Ray
Editor:	Papatsouma/Whitney
External:	Woods
Date:	April 7, 2021

MSc EXAMINATIONS (STATISTICS)
January 2021

MATH97123 Fundamentals of Statistical Inference

Time: 2 hours

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

1. (i) Define what is meant by a *minimax* decision rule. Define what is meant by a *Bayes* decision rule. What is a *least favourable prior*?

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: $MR(d) \leq MR(d')$ 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) Let d_π be the Bayes rule with respect to a prior $\pi(\theta)$ and let $r_\pi = r(\pi, d_\pi)$ be the associated Bayes risk. Suppose that for an arbitrary prior $\pi^*(\theta)$ the associated Bayes rule is d_{π^*} and let $r_{\pi^*} = r(\pi^*, d_{\pi^*})$. Then the prior π is **least favourable** if $r_\pi \geq r_{\pi^*}$ for all priors π^*. [2 marks]

Suppose d_π is a decision rule that is Bayes with respect to a prior $\pi(\theta)$. Suppose that the risk function of d_π satisfies

$$R(\theta, d_\pi) \leq r(\pi, d_\pi),$$

for all θ, where $r(\pi, d_\pi)$ is the Bayes risk of d_π with respect to the prior $\pi(\theta)$.

Show that $\pi(\theta)$ is a least favourable prior.

ANSWER: (Seen a version) Let π^* be an arbitrary prior, with associated Bayes rule d_{π^*}. Then,

$$r_{\pi^*} = \int_{\Omega_\theta} R(\theta, d_{\pi^*})\pi^*(\theta)d\theta \leq \int_{\Omega_\theta} R(\theta, d_\pi)\pi^*(\theta)d\theta \leq r(\pi, d_\pi)\int_{\Omega_\theta} \pi^*(\theta)d\theta \equiv r(\pi, d_\pi),$$

the first inequality holding since d_{π^*} is Bayes with respect to π^*, and the second by assumption So, π is least favourable. [3 marks]

When is a Bayes decision rule minimax?

ANSWER: (Unseen) A Bayes rule is minimax if it is Bayes with respect to a least favourable prior. [Various specific results could be quoted here: if a Bayes rule has constant risk, it is minimax etc.] [2 marks]

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

(ii) Let $Y_1, \ldots, Y_n$ be independent, identically distributed random variables, with common density

$$f(y;\theta) = \frac{y}{\theta^2} e^{-y/\theta}, \; y > 0.$$

What is the maximum likelihood estimator of θ?

Find the form of the Jeffreys prior for θ, and the posterior distribution for θ under this prior.

What is the Bayes estimator of θ under squared error loss function, $L(\theta, \hat{\theta}) = (\hat{\theta} - \theta)^2$?

Verify that the Bayes estimator of θ under the loss function

$$L(\theta, \hat{\theta}) = \exp\left\{a\left(\frac{\hat{\theta}}{\theta} - 1\right)\right\} - a\left(\frac{\hat{\theta}}{\theta} - 1\right) - 1,$$

with $a \neq 0$, is given by

$$\hat{\theta}_B = \frac{\sum_{i=1}^n Y_i}{a}\left(1 - e^{-\frac{a}{2n+1}}\right).$$

ANSWER: (Seen similar) The likelihood function is

$$L(\theta) = \{\prod_{i=1}^n y_i\} \exp\{-\sum_{i=1}^n y_i/\theta\}/\theta^{2n},$$

so taking logarithm to obtain the log-likelihood $l(\theta)$ and setting the derivative to zero, the MLE $\hat{\theta}$ satisfies

$$-2n/\theta + \{\sum_{i=1}^n y_i\}/\theta^2 = 0,$$

giving $\hat{\theta} = \sum_{i=1}^n Y_i/2n$. [2 marks]

ANSWER: (Seen similar) The Jeffreys prior $\pi(\theta) \propto I(\theta)$, where $I(\theta)$ is the Fisher information

$$I(\theta) = E\{-\frac{\partial^2 l(\theta)}{\partial\theta^2}\}.$$

Here,

$$-\frac{\partial^2 l(\theta)}{\partial\theta^2} = 2n/\theta^2 - 2\{\sum_{i=1}^n y_i\}/\theta^3,$$

so, using $E(Y_i) = 2\theta$, following the hint, we have $I(\theta) = 2n/\theta^2$, and the Jeffreys prior has $\pi(\theta) \propto 1/\theta$. [2 marks]

ANSWER: (Seen similar) Then,

$$\pi(\theta|y) \propto L(\theta)\pi(\theta) \propto \theta^{-(2n+1)} e^{-\sum_{i=1}^n y_i/\theta},$$

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

identifying, in the notation of the hint, the posterior distribution as inverse-Gamma($2n, \sum_{i=1}^n y_i$). [2 marks]

ANSWER: (Seen similar) The Bayes estimator under squared error loss function is the mean of the posterior distribution, which is $\sum_{i=1}^n Y_i/(2n-1)$, from the given information. [1 mark]

ANSWER: (Seen similar) Denote by E_p expectation with respect to θ, under the posterior distribution. The Bayes estimator $\hat{\theta}$ minimises the expected posterior loss, $E_p\{L(\theta,\hat{\theta})\}$.

For the given loss,

$$E_p\{L(\theta,d)\} = e^{-a}\int_0^\infty e^{\frac{ad}{\theta}}\pi(\theta|y)d\theta - adE_p(\frac{1}{\theta}) - 1.$$

Differentiate with respect to d, then solve

$$e^{-a}E_p\left(\frac{a}{\theta}e^{\frac{ad}{\theta}}\right) - aE_p\left(\frac{1}{\theta}\right) = 0. \tag{1}$$

We have

$$E_p\left(\frac{1}{\theta}\right) = \frac{2n}{\sum_{i=1}^n Y_i},$$

and

$$\begin{aligned} E_p\left(\frac{1}{\theta}e^{\frac{ad}{\theta}}\right) &= \frac{(\sum_{i=1}^n Y_i)^{2n}}{\Gamma(2n)}\int_0^\infty \frac{1}{\theta}\cdot\frac{1}{\theta^{2n+1}}\exp\left\{-\frac{(\sum_{i=1}^n Y_i - ad)}{\theta}\right\}d\theta \\ &= \frac{(\sum_{i=1}^n Y_i)^{2n}}{\Gamma(2n)}\int_0^\infty \theta^{-(2n+2)}\exp\left\{-\frac{(\sum_{i=1}^n Y_i - ad)}{\theta}\right\}d\theta \\ &= \frac{(\sum_{i=1}^n Y_i)^{2n}}{\Gamma(2n)}\frac{\Gamma(2n+1)}{(\sum_{i=1}^n Y_i - ad)^{(2n+1)}}, \end{aligned}$$

so solving(1) gives the Bayes estimator as

$$\hat{\theta}_B = \frac{\sum_{i=1}^n Y_i}{a}\left(1 - e^{-\frac{a}{2n+1}}\right).$$

[7 marks]

[Total 25 marks]

[If Y has the inverse-Gamma(α,β) distribution with density

$$f(y;\theta) = \frac{\beta^\alpha}{\Gamma(\alpha)}y^{-(\alpha+1)}\exp(-\beta/y),\ y>0,$$

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

then $E(Y) = \beta/(\alpha - 1)$, $E(1/Y) = \alpha/\beta$.

If Y has the Gamma(k, θ) distribution with density

$$f(y; k, \theta) = \frac{y^{k-1} e^{-y/\theta}}{\theta^k \Gamma(k)}, \ y > 0,$$

then $E(Y) = k\theta$.]

2. Let $Y_1, \ldots, Y_p$ be independent random variables such that $Y_i \sim N(\theta_i, 1)$. Write $Y = (Y_1, \ldots, Y_p)^T$ and $\theta = (\theta_1, \ldots, \theta_p)^T$. Let $\hat{\theta} \equiv \hat{\theta}(Y) = (\hat{\theta}_1(Y), \ldots, \hat{\theta}_p(Y))^T$ be an estimator of θ, and let $g(Y) \equiv (g_1(Y), \ldots, g_p(Y))^T = \hat{\theta} - Y$. Denote by $\|\cdot\|$ the Euclidean norm, $\|Y\|^2 = Y_1^2 + \ldots + Y_p^2$.

Suppose that $D_i(Y) = \partial g_i(Y)/\partial Y_i$ exists. Then it is known that

$$\hat{R}\{\hat{\theta}(Y)\} = p + 2\sum_{i=1}^{p} D_i(Y) + \sum_{i=1}^{p}\{g_i(Y)\}^2$$

is an unbiased estimator of the risk of $\hat{\theta}$, under squared error loss $L(\theta, \hat{\theta}) = \|\hat{\theta} - \theta\|^2$. [You are not required to show this].

(i) The James-Stein estimator is

$$\delta_{JS}(Y) = \Big(1 - \frac{p-2}{\|Y\|^2}\Big)Y.$$

Show that an unbiased risk estimator for $\delta_{JS}(Y)$ is

$$\hat{R}\{\delta_{JS}(Y)\} = p - (p-2)^2/\|Y\|^2.$$

Deduce that Y is inadmissible as an estimator of θ.
Is $\delta_{JS}(Y)$ admissible?

ANSWER: (Seen similar) For the James-Stein estimator we have

$$g_i(Y) = -\frac{(p-2)}{\|Y\|^2}Y_i,$$

so,

$$D_i(Y) = -(p-2)\cdot\frac{\|Y\|^2 - 2Y_i^2}{\|Y\|^4},$$

by the quotient rule. Then

$$\sum_{i=1}^{p} D_i(Y) = -(p-2)\left\{\frac{p}{\|Y\|^2} - 2\frac{1}{\|Y\|^2}\right\} = -(p-2)^2/\|Y\|^2,$$

and

$$\sum_{i=1}^{p}\{g_i(Y)\}^2 = (p-2)^2\frac{\sum_{i=1}^{p} Y_i^2}{\|Y\|^4} = (p-2)^2/\|Y|^2,$$

so that

$$\hat{R}\{\delta_{JS}(Y)\} = p - 2(p-2)^2/\|Y|^2 + (p-2)^2/\|Y\|^2 = p - (p-2)^2/\|Y\|^2.$$

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

[6 marks]

ANSWER: (Seen similar) Take expectation: since $E_\theta\left[\hat{R}\{\delta_{JS}(Y)\}\right] = R(\theta, \delta_{JS}(Y))$ it follows that the risk of $\delta_{JS}(Y)$ is

$$R(\theta, \delta_{JS}(Y)) = p - (p-2)^2 E_\theta(1/||Y||^2) < p \equiv R(\theta, Y),$$

since $R(\theta, Y) = \sum_{i=1}^p E_\theta(Y_i - \theta_i)^2 = \sum_{i=1}^p \text{var}(Y_i) = p$. The expectation $E_\theta(1/||Y||^2)$ is certainly > 0. So, $\delta_{JS}(Y)$ strictly dominates Y as an estimator of θ, and Y is inadmissible. [4 marks]

ANSWER: (Seen) The James-Stein estimator $\delta_{JS}(Y)$ is **not** admissible. It is strictly dominated by the 'positive part James-Stein estimator'

$$\delta^+_{JS}(Y) = \left(1 - \frac{p-2}{||Y||^2}\right)^+ Y,$$

where $(1 - \frac{p-2}{||Y||^2})^+ = \max\{0, 1 - \frac{p-2}{||Y||^2}\}$. [3 marks]

(ii) Describe a Bayesian model for which the shrinkage estimator

$$\hat{\theta}_S(Y) = \frac{\lambda}{1+\lambda} Y$$

is a Bayes estimator of θ.

ANSWER: (Seen, but requires insight) Suppose we specify a $N(0, \lambda)$ prior on θ_i, independently , $i = 1, \ldots, p$. Then the posterior distribution of θ_i is $N(\frac{\lambda}{1+\lambda} Y_i, \frac{\lambda}{1+\lambda})$. Under the assumed loss function, the Bayes estimator of θ is the posterior mean, which is precisely $\hat{\theta}_S(Y)$. [4 marks]

Suppose λ is *unspecified* in the Bayesian model. Find the form of the estimator obtained by choosing the value of λ which minimises the unbiased risk estimator $\hat{R}\{\hat{\theta}_S(Y)\}$.

ANSWER: (Unseen) Then, in the notation of the question, $g_i(Y) = -\frac{Y_i}{1+\lambda}$, $D_i(Y) = -\frac{1}{1+\lambda}$. Then

$$\begin{aligned}\hat{R}\{\hat{\theta}_S(Y)\} &= p - \frac{2p}{1+\lambda} + \frac{\sum_{i=1}^p Y_i^2}{(1+\lambda)^2} \\ &= p\frac{\lambda - 1}{1+\lambda} + \frac{\sum_{i=1}^p Y_i^2}{(1+\lambda)^2}.\end{aligned}$$

Set the derivative of $\hat{R}\{\hat{\theta}_S(Y)\}$ to zero to find the value of λ, $\hat{\lambda}_S$ say, that minimises the unbiased risk estimator, noting that we need $\lambda \geq 0$. Then $\hat{\lambda}_S$ solves

$$\frac{\sum_{i=1}^p Y_i^2}{(1+\lambda)^3} - \frac{p}{(1+\lambda)^2} = 0.$$

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

If there is no solution ,$\hat{\lambda}_S = 0$ Then, the resulting estimator is

$$\hat{\theta}_S(Y) = \frac{\hat{\lambda}_S}{1 + \hat{\lambda}_S} Y = (1 - \frac{p}{||Y||^2})^+ Y.$$

[This is almost the positive part James-Stein estimator, but with p instead of $p - 2$]. [8 marks]

[Total 25 marks]

[You may assume Stein's Lemma: if $Y \sim N(\mu, \sigma^2)$ then, for differentiable function g with $E\{g'(Y)\} < \infty$,

$$E\{g(Y)(Y - \mu)\} = \sigma^2 E\{g'(Y)\}.]$$

3. (i) Define what is meant by a *uniformly most powerful (UMP) test* of a statistical hypothesis. What is a *uniformly most powerful unbiased (UMPU) test* of a statistical hypothesis? Explain the construction of UMP and UMPU tests in the context of a one parameter exponential family distribution.

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]

ANSWER: (Seen) Consider a random vector Y with density belonging to a one parameter exponential family of the form

$$f(y; \theta) = h(y) \exp\{\theta t(y) - k(\theta)\}.$$

Then UMP one-sided tests on the natural parameter θ can be constructed using $T = t(Y)$ as test statistic. For instance, to test $H_0 : \theta = \theta^*$ against $H_1 : \theta > \theta^*$, for specified θ^*, a UMP test is of the form: reject H_0 if $T > c$, where c is fixed so that $P_{\theta=\theta^*}(T > c) = \alpha$, for a test of size α. UMPU two-sided tests are similarly based on the statistic T. For instance, to test $H_0 : \theta = \theta^*$ against $H_1 : \theta \neq \theta^*$, a UMPU test is of the form: reject H_0 if $T > c_1$ or $T < c_2$, where c_1, c_2 are fixed so that the probability of rejection is α under H_0 and the test is unbiased (so that the derivative of the power function is 0 at θ^*). [4 marks]

(ii) Let $Y_1, \ldots, Y_n$ be independent, identically distributed, with common inverse Gaussian $IG(\mu, \lambda)$ probability density function

$$f(y; \mu, \lambda) = \frac{\sqrt{\lambda}}{\sqrt{2\pi}} y^{-3/2} \exp\left\{-\frac{\lambda(y-\mu)^2}{2\mu^2 y}\right\}, \ y > 0, \ \lambda > 0, \ \mu > 0.$$

Show that this assumed distribution constitutes a full exponential family model and identify the natural statistics.

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

ANSWER: (Seen similar) We can re-write the density of $IG(\mu, \lambda)$ in the form

$$f(y;\mu,\lambda) = h(y)\exp\left\{-\frac{\lambda y}{2\mu^2} - \frac{\lambda}{2y} + \frac{\lambda}{\mu} + \frac{1}{2}\log(\lambda)\right\},$$

to see that the assumed distribution constitutes a full $(2,2)$ exponential family, with natural parameters $\phi^1 = -\frac{\lambda}{2\mu^2}, \phi^2 = -\frac{\lambda}{2}$.

Based on the sample $Y_1, \ldots, Y_n$ the natural statistics are $S_1 = \sum_{i=1}^n Y_i$ and $S_2 = \sum_{i=1}^n 1/Y_i$. [5 marks]

Verify that the distribution of $\frac{\lambda}{\mu^2}(Y-\mu)^2/Y$ is χ_1^2, chi-squared on 1 degree of freedom.

ANSWER: (Unseen) Let $Z = \frac{\lambda}{\mu^2}(Y-\mu)^2/Y$. Then the moment generating function of Z is

$$\begin{aligned} M_Z(t) &= E(e^{tZ}) \\ &= \int_0^\infty \sqrt{\frac{\lambda}{2\pi y^3}}\exp\left\{\frac{t\lambda(y-\mu)^2}{\mu^2 y} - \frac{1}{2}\frac{\lambda(y-\mu)^2}{\mu^2 y}\right\}dy \\ &= \int_0^\infty \sqrt{\frac{\lambda}{2\pi y^3}}\exp\left\{-\frac{1}{2}\lambda(1-2t)\frac{(y-\mu)^2}{\mu^2 y}\right\}dy \\ &= \frac{\sqrt{\lambda}}{\sqrt{\lambda(1-2t)}} = (1-2t)^{-1/2}, \end{aligned}$$

relating the integrand to $IG(\mu, \lambda(1-2t))$. We recognise this as the moment generating function of χ_1^2. [3 marks]

Explain in detail how to test the null hypothesis $H_0 : \lambda \le \lambda_0$ against the alternative hypothesis $H_1 : \lambda > \lambda_0$, assuming that μ is *known*, justifying the construction carefully and explaining any UMP or UMPU properties of the test.

ANSWER: (Unseen) The joint pdf of an IID sample $Y = \{Y_1, \ldots, Y_n\}$ is

$$f(y;\mu,\lambda) = \prod_{i=1}^n \left(\frac{\lambda}{2\pi y_i^3}\right)^{1/2}\exp\left\{-\frac{\lambda}{2}\frac{(y_i-\mu)^2}{\mu^2 y_i}\right\}.$$

If μ is known, this defines a one parameter exponential family, with natural parameter $-\lambda/2$. The general theory above implies the existence of a uniformly most powerful size α test of H_0 against H_1 of the form: reject H_0 if $T(Y) < c$, where the natural statistic is

$$T(Y) = \sum_{i=1}^n \frac{(Y_i-\mu)^2}{\mu^2 Y_i}.$$

We have that $\lambda T(Y)$ is distributed as χ_n^2, being a sum of n IID χ_1^2 variables. So, set $\lambda_0 c$ to be the α quantile of χ_n^2 for a UMP test of size α. [5 marks]

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

[Total 25 marks]

[Note that if Z is distributed as χ^2_k, its moment generating function $M_Z(t) = E(e^{tZ}) = (1-2t)^{-k/2}, t < 1/2$.]

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

(i) The importance of sufficiency and completeness in parameter estimation.

(ii) The Conditionality and Likelihood Principles of statistical inference

(iii) Bayesian approaches to hypothesis testing

(iv) Large sample tests based on maximum likelihood estimators.
ANSWER: Entirely descriptive question. Notes can been seen at

http://www2.imperial.ac.uk/~ayoung/math97123coursenotes2020-21.pdf.

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]