Imperial College London

MSc EXAMINATIONS (STATISTICS)

January 2020

M5MS02

# Fundamentals of Statistical Inference (Solutions)

| Setter's signature | Checker's signature | Editor's signature |
| --- | --- | --- |
| ................ | ................ | ................ |

1. (i) 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 seen ⇓

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

where $E_\theta$ denotes expectation with respect to distribution of $Y$, assuming $Y$ has distribution defined by parameter value $\theta$.

A decision rule $d$ is *strictly dominated* by rule $d'$ if $R(\theta, d') \leq R(\theta, d)$ for *all* $\theta \in \Omega_\theta$ *and* $R(\theta, d') < R(\theta, d)$ for *at least one point* $\theta \in \Omega_\theta$. A decision rule which is **not** strictly dominated by another rule is **admissible**. 2

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

The Bayes decision rule is found by minimising the expected posterior loss. For any given data $y$, the Bayes rule $\delta(y)$ is characterised by: $\delta(y)$ is the action which minimises

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

where $\pi(\theta|y) \propto \pi(\theta) \times f(y; \theta)$ is the posterior density of $\theta$, given $y$. 3

Suppose decision rule $d$ is *unique* Bayes for prior $\pi(\theta)$, but is inadmissible. Then $d$ is strictly dominated, as defined above, by a rule, $\delta$ say. But then we would have $r(\pi, \delta) = \int_{\Omega_\theta} R(\theta, \delta)\pi(\theta)d\theta \leq \int_{\Omega_\theta} R(\theta, d)\pi(\theta)d\theta = r(\pi, d)$. Since $d$ is Bayes, we would then require $r(\pi, \delta) = r(\pi, d)$, so that $\delta$ is also Bayes, contradicting uniqueness, so $d$ must be admissible. 4

(ii) We have, as functions of $\theta$, $\boxed{\text{part seen} \Downarrow}$

$$f(y;\theta) \propto \exp\left\{y^T\theta/\sigma^2 - \theta^T\theta/(2\sigma^2)\right\},$$

and

$$\pi(\theta) \propto \exp\left\{\theta^T\theta_0/\tau_0^2 - \theta^T\theta/(2\tau_0^2)\right\}.$$

Therefore,

$$\pi(\theta|y) \propto \exp\left\{\theta^T[\theta_0/\tau_0^2 + y/\sigma^2] - \theta^T\theta[1/\sigma^2 + 1/\tau_0^2]/2\right\} = \exp\left\{\theta^T\theta_1/\tau_1^2 - \theta^T\theta/(2\tau_1^2)\right\},$$

where

$$1/\tau_1^2 = 1/\tau_0^2 + 1/\sigma^2,\ \theta_1 = \frac{1/\tau_0^2}{1/\tau_0^2 + 1/\sigma^2}\theta_0 + \frac{1/\sigma^2}{1/\tau_0^2 + 1/\sigma^2}y,$$

of the form $w\theta_0 + (1-w)y$. So, the posterior distribution is $N_p(\theta_1, \tau_1^2 I_p)$. $\boxed{5}$

Under the given loss function, the *unique* Bayes estimator is the posterior mean, $\theta_1$. $\boxed{2}$

Then, given $w \in (0,1)$, fix $\tau_0^2$ so that $\frac{1/\tau_0^2}{1/\tau_0^2+1/\sigma^2} = w$. We have that $\delta(Y)$ is the unique Bayes estimator under the normal prior $N_p(\theta_0, \tau_0^2 I_p)$, and hence admissible by the result in (i). $\boxed{3}$

Yes, $\delta(Y)$ **is** an admissible estimator in the case $\sigma^2$ is unknown. If not, it would be strictly dominated by some estimator $\delta'(Y)$. Then we would have $\boxed{\text{unseen} \Downarrow}$

$$\begin{aligned}
\forall\ (\theta,\sigma^2),\ R((\theta,\sigma^2),\delta') &\leq R((\theta,\sigma^2),\delta),\\
\exists\ (\theta_0,\sigma_0^2),\ R((\theta_0,\sigma_0^2),\delta') &< R((\theta_0,\sigma_0^2),\delta).
\end{aligned}$$

But, this would contradict admissibility of $\delta(Y)$ for estimating $\theta$ when $\sigma^2$ is known to equal $\sigma_0^2$. $\boxed{4}$

$\boxed{\text{Total } 25}$

2. (i) Given $Y$ from a model function $f(y;\theta)$, a statistic $S = s(Y)$ is **sufficient** for parameter $\theta$ if the conditional distribution of $Y|S = s$ does not depend on $\theta$, for all $s$. $S$ is **minimal sufficient** if it can be expressed as a function of every other sufficient statistic. seen ⇓ 3

S is said to be **complete** if $E_\theta\{g(S)\} = 0 \; \forall\theta \implies P_\theta\{g(S) = 0\} = 1 \; \forall\theta$. The importance of completeness in point estimation is contained in the result that says that if there exists an unbiased estimator of $\theta$ which is a function of a complete sufficient statistic, then it is the *unique* such estimator. Let $S$ be a complete sufficient statistic for a parameter $\theta$ and let $\phi(S)$ be any estimator based only on $S$. Then $\phi(S)$ is the unique minimum variance estimator of its expectation. 3

unseen ⇓

(ii) Suppose $\widehat{\theta}$ is UMVU and let $U$ satisfy $E_\theta(U) = 0, E_\theta(U^2) < \infty$. For arbitrary $\lambda \in \mathbb{R}$, define $\widehat{\theta}_\lambda = \widehat{\theta} + \lambda U$. Then

$$\begin{aligned} 0 &\leq \text{var}_\theta(\widehat{\theta}_\lambda) - \text{var}_\theta(\widehat{\theta}) = 2\lambda\text{cov}_\theta(\widehat{\theta}, U) + \lambda^2 E_\theta(U^2) \\ &= E_\theta(U^2)\left(\lambda + \frac{\text{cov}_\theta(\widehat{\theta}, U)}{E_\theta(U^2)}\right)^2 - \frac{\text{cov}^2_\theta(\widehat{\theta}, U)}{E_\theta(U^2)}. \end{aligned}$$

This inequality can only be satisfied at $\lambda = -\frac{\text{cov}_\theta(\widehat{\theta},U)}{E_\theta(U^2)}$ if $\text{cov}_\theta(\widehat{\theta}, U) = 0$, or equivalently $E_\theta(\widehat{\theta}U) = 0$, for all $\theta \in \Omega_\theta$. 8

unseen ⇓

(iii) From the given density function,

$$E_\theta(Y_{(n)}) = \int_\theta^{2\theta} y\frac{n}{\theta}(\frac{y-\theta}{\theta})^{n-1}dy = \int_0^\theta (x+\theta)\frac{n}{\theta}x^{n-1}dx = \frac{2n+1}{n+1}\theta.$$

Use symmetry, $E_\theta(Y_{(n)} - \frac{3}{2}\theta) = E_\theta(\frac{3}{2}\theta - Y_{(1)})$, to obtain $E(Y_{(1)}) = \frac{n+2}{n+1}\theta$, so that $E_\theta(\widehat{\theta}) = \theta$, as required to show unbiased. 3

Then $U = (2n+1)Y_{(1)} - (n+2)Y_{(n)}$ is an unbiased estimator of 0. Directly, using the given information,

$$\begin{aligned} 3E_\theta(\widehat{\theta}U) &= E_\theta[(Y_{(1)} + Y_{(n)})\{(2n+1)Y_{(1)} - (n+2)Y_{(n)}\}] \\ &= (2n+1)E_\theta(Y^2_{(1)}) + (n-1)E_\theta(Y_{(1)}Y_{(n)}) - (n+2)E_\theta(Y^2_{(n)}) \\ &= \frac{(n-1)}{(n+1)(n+2)}\theta^2 \neq 0, \end{aligned}$$

on simplification. So, $\widehat{\theta}$ is **not** UMVU. 8

Total 25

3. (i) Given the parameter space $\Omega_\theta$ for a parameter $\theta$, consider testing the null hypothesis $H_0 : \theta \in \Theta_0$ against alternative $H_1 : \theta \in \Theta_1$, where $\Theta_0$ and $\Theta_1$ are *disjoint* subsets of $\Omega_\theta$. $\boxed{\text{seen} \Downarrow}$

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 $\phi$ to be $w(\theta) = E_\theta\{\phi(Y)\}$.

A **uniformly most powerful** (UMP) test of size $\alpha$ 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 $\theta \in \Theta_1$. $\boxed{4}$

A test $\phi$ of $H_0 : \theta \in \Theta_0$ against $H_1 : \theta \in \Theta_1$ is unbiased of size $\alpha$ 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 $\alpha$. $\boxed{4}$

Consider a random vector $Y$ with density belonging to a full exponential family of the form $\boxed{\text{seen} \Downarrow}$

$$f(y; \theta) = h(y) \exp\left\{\sum_{i=1}^{m} t_i(y)\theta^i - k(\theta)\right\}.$$

One-parameter case: $m = 1$. Then UMP one-sided tests on the natural parameter $\theta^1$ can be constructed using $T_1 = t_1(Y)$ as test statistic. For instance, to test $H_0 : \theta^1 = \theta^{1*}$ against $H_1 : \theta^1 > \theta^{1*}$, for specified $\theta^{1*}$, a UMP test is of the form: reject $H_0$ if $T_1 > c$, where $c$ is fixed so that $P_{\theta^1=\theta^{1*}}(T_1 > c) = \alpha$, for a test of size $\alpha$. UMPU two-sided tests are similarly based on the statistic $T_1$. For instance, to test $H_0 : \theta^1 = \theta^{1*}$ against $H_1 : \theta^1 \neq \theta^{1*}$, a UMPU test is of the form: reject $H_0$ if $T_1 > c_1$ or $T_1 < c_2$, where $c_1, c_2$ are fixed so that the probability of rejection is $\alpha$ under $H_0$ and the test is unbiased (so that the derivative of the power function is 0 at $\theta^{1*}$). $\boxed{3}$

Multi-parameter case: $m > 1$. Write $T_i = t_i(Y)$ and suppose we wish to test a hypothesis about $\theta^1$, with $\theta^2, \ldots, \theta^m$ nuisance. Then optimal (UMPU) tests are based on the conditional distribution of $T_1$, given $T_i = t_i, i = 2, \ldots, m$, where $t_i$ is the observed data value of $T_i$. The precise form of test will depend on the hypothesis being tested [one-side, two-sided, point null hypothesis], but, as illustration, to test $H_0 : \theta^1 \leq \theta^{1*}$ against $H_1 : \theta^1 > \theta^{1*}$, for a specified $\theta^{1*}$, the test is of the form: reject $H_0$ if $T_1 > t_1^c$, where we require

$$P_{\theta^{1*}}(T_1 > t_1^c | T_2 = t_2, \ldots, T_m = t_m) = \alpha,$$

for a test of required size $\alpha$. $\boxed{4}$

(ii) The joint pdf of $Y = (Y_1, \ldots, Y_n)$ is  seen similar $\Downarrow$

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

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

In the case when $\mu$ 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 $H_1 : \sigma > \sigma_0$ is equivalent to testing $H_0 : \theta^1 = \theta^{1*} = -\frac{1}{2\sigma_0^2}$ against $H_1 : \theta^1 > \theta^{1*}$. Then, from lecture theory, a **UMP** test exists, of the form: reject $H_0$ if

$$\frac{1}{\sigma_0^2}\sum_{i=1}^n (Y_i - \mu)^2 > c,$$

where $c$ is such that

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

for a test of size $\alpha$. 4

In the case when $\mu$ 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 $H_1 : \theta^1 > \theta^{1*}$, with $\theta^2$ 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 $\alpha$ test is: reject $H_0$ if $\sum_{i=1}^n Y_i^2 > k$, say, where

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

where $\bar{y}$ is the observed sample mean.

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 $S^2/\sigma_0^2 > c'$, where

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

for a level $\alpha$ test. 6

Total 25

4. Entirely descriptive question. Notes can been seen at

http://www2.imperial.ac.uk/$\sim$ayoung/m5ms02coursenotes2019-20.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. $\boxed{\text{Total } 25}$