Imperial College London

Module:	MATH70078
Setter:	Young
Checker:	Ray
Editor:	Varty
External:	Woods
Date:	January 9, 2023

MSc EXAMINATIONS (STATISTICS)
January 2023

MATH70078 Fundamentals of Statistical Inference

Time: 2 hours

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

1. Let $Y_1, \dots, Y_n$ be an independent and identically distributed sample from the normal distribution $N(\mu, \sigma^2)$, with known σ^2, and consider testing the null hypothesis $H_0 : \mu \leq 0$ against the alternative hypothesis $H_1 : \mu > 0$.

   Explain carefully how an optimal frequentist test of H_0 against H_1 of size α would be carried out.

   ANSWER: (Seen)

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

$$\begin{aligned} f(y; \mu, \sigma) &\propto \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \mu)^2\right\} \\ &\propto \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n} y_i^2 + \frac{\mu}{\sigma^2}\sum_{i=1}^{n} y_i\right\}. \end{aligned}$$

   With σ^2 known, this joint density has monotone likelihood ratio with respect to $\sum_{i=1}^n Y_i$, so an optimal (uniformly most powerful) frequentist test of H_0 against H_1 is of the form: reject H_0 if $\sum_{i=1}^n Y_i$ is large, or equivalently $\bar{Y} = n^{-1}\sum_{i=1}^n Y_i > c$, where c is such that $P(\bar{Y} > c; \mu = 0) = \alpha$, the required size of test. When $\mu = 0$, $\bar{Y}$ is distributed as $N(0, \sigma^2/n)$, so c is calculated so that $P(Z > \frac{1}{\sigma/\sqrt{n}}c) = \alpha$, when Z is $N(0,1)$, so that $c = \sigma/\sqrt{n}\Phi^{-1}(1-\alpha)$. [6 marks]

   Given data $y_1, \dots, y_n$ with mean $\bar{y} = n^{-1}\sum_{i=1}^n y_i$, explain why the appropriate p−value for testing H_0 against H_1 can be expressed as $1 - \Phi(\sqrt{n}\bar{y}/\sigma)$, in terms of the $N(0,1)$ distribution function $\Phi(\cdot)$.

   ANSWER: (Seen)

   The appropriate p−value is $\sup_{H_0} P(\bar{Y} \geq \bar{y})$, which is achieved at $\mu = 0$ by the monotone likelihood ratio property, so this is $P(\bar{Y} \geq \bar{y}; \mu = 0)$, which is the given expression by the above analysis. [3 marks]

   Consider now a Bayesian analysis, where μ is assigned a $N(0, \tau^2)$ prior, with τ^2 known.

   Calculate the posterior probability that H_0 is true, $P(\mu \leq 0 | y_1, \dots, y_n)$.

   ANSWER: (Seen)

   The posterior distribution of $\mu | y_1, \dots, y_n$ is normal with mean $\{\tau^2/(\tau^2 + \sigma^2/n)\}\bar{y}$ and variance $\tau^2/(1 + n\tau^2/\sigma^2)$, using the result derived in lecture material.

   Then

$$P(\mu \leq 0 | y_1, \dots, y_n) = P\left(Z \leq \frac{0 - \{\tau^2/(\tau^2 + \sigma^2/n)\}\bar{y}}{\sqrt{\tau^2/(1 + n\tau^2/\sigma^2)}}\right) = P\left(Z \geq \frac{\tau}{\sqrt{(\sigma^2/n)(\tau^2 + \sigma^2/n)}}\bar{y}\right),$$

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

where, as above, Z is $N(0, 1)$. [6 marks]

For the case $\sigma^2 = \tau^2 = 1$, compare this posterior probability with the frequentist p−value, for values of $\bar{y} > 0$, and show that the Bayesian probability is always greater than the p−value.

ANSWER: (Unseen)

For $\sigma^2 = \tau^2 = 1$ we have

$$P(\mu \leq 0|y_1, \dots, y_n) = P\left(Z \geq \frac{1}{\sqrt{(1/n)(1+1/n)}}\bar{y}\right),$$

while the p−value is

$$P(Z \geq \frac{1}{\sqrt{1/n}}\bar{y}).$$

Because

$$\frac{1}{\sqrt{(1/n)(1+1/n)}} < \frac{1}{\sqrt{1/n}},$$

the Bayesian probability is larger than the p−value if $\bar{y} > 0$. [5 marks]

What is $\lim_{\tau^2\to\infty} P(\mu \leq 0|y_1, \dots, y_n)$?

ANSWER: (Unseen)

As $\tau^2 \to \infty$, the coefficient of $\bar{y}$ in the Bayesian probability

$$\frac{\tau}{\sqrt{(\sigma^2/n)(\tau^2 + \sigma^2/n)}} = \frac{1}{\sqrt{(\sigma^2/n)\{1 + \sigma^2/(\tau^2 n)\}}} \to \frac{1}{\sigma/\sqrt{n}},$$

the coefficient in the p−value. So, in the limit as $\tau^2 \to \infty$, the Bayesian probability and the frequentist p−value coincide. [5 marks]

[Total 25 marks]

2. Let $(X_1, Y_1), \dots, (X_n, Y_n)$ be an independent, identically distributed sample of size n from the distribution with joint density

$$f(x, y; \psi) = \exp(-\psi x - \psi^{-1} y), \;\; x > 0, y > 0,$$

depending on the parameter $\psi > 0$.

Show that this distribution constitutes an example of a curved exponential family.

ANSWER: (Seen similar)

Writing $z = (x, y)$, the given density can be written in the form

$$f(z; \psi) = \exp\{\pi_1(\psi) T_1(z) + \pi_2(\psi) T_2(z)\},$$

with $\pi_1(\psi) = -\psi, \pi_2(\psi) = -1/\psi$ and $T_1(z) = x$, $T_2(z) = y$. This general form (noting that $\pi_2(\psi) = 1/\pi_1(\psi)$) describes a curved exponential family. Note that if $Z = (X, Y)$, we interpret the distribution as having X, Y independent, exponential with means $1/\psi$ and ψ respectively. [3 marks]

Derive the maximum likelihood estimator $\hat{\psi}$ of ψ. What is the Fisher information $I_n(\psi)$ about ψ contained in the sample of size n? What is the asymptotic distribution of $\hat{\psi}$?

ANSWER: (Seen similar)

With $X = \sum_{i=1}^n X_i$, $Y = \sum_{i=1}^n Y_i$, the likelihood function based on outcomes $X = x, Y = y$ is

$$L(\psi; x, y) = f(x, y; \psi) = \exp(-\psi x - \psi^{-1} y),$$

and log-likelihood $l(\psi) = -\psi x - \psi^{-1} y$. Setting the derivative $l'(\psi) = 0$ gives $\hat{\psi} = \sqrt{Y/X}$.

The Fisher information $I_n(\psi) = E\{-l''(\psi)\} = \frac{2}{\psi^3} E(Y)$. By above observation $E(Y) = n\psi$, so $I_n(\psi) = 2n/\psi^2$.

Asymptotically, $\sqrt{n}(\hat{\psi} - \psi)$ is distributed as $N(0, 1/I_1(\psi))$, that is $N(0, \psi^2/2)$.

[6 marks]

Verify that $\hat{\psi}$ is not minimal sufficient, but that $(\hat{\psi}, A)$ is, where $A = n\sqrt{\bar{X}\bar{Y}}$, with $\bar{X} = n^{-1} \sum_{i=1}^n X_i$, $\bar{Y} = n^{-1} \sum_{i=1}^n Y_i$, and A is ancillary.

ANSWER: (Seen similar)

From the form of the likelihood function, the minimal sufficient statistic is 2-dimensional, and can be taken as $(X, Y) \equiv (\sqrt{Y/X}, \sqrt{XY}) \equiv (\hat{\psi}, A)$. So, $\hat{\psi}$ is not minimal sufficient, but $(\hat{\psi}, A)$ is.

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

Notice that $X = \sum_{i=1}^n X_i$ is distributed as $\mathsf{Gamma}(n, \psi)$ and (independently) $Y = \sum_{i=1}^n Y_i$ is $\mathsf{Gamma}(n, \psi^{-1})$. We have that $X \equiv \psi^{-1} W_1$ and $Y \equiv \psi W_2$, where W_1, W_2 are independent Gamma random variables free of ψ. Then, immediately, XY is free of ψ, so A is. [6 marks]

Suppose now $n = 1$ and let $\xi = \log \psi$.

Explain what is meant by the Conditionality Principle and explain why, in the current example, this would indicate that the relevant distribution for carrying out inference on ξ should be the conditional distribution of the maximum likelihood estimator $\hat{\xi}$ given $A = a$, its observed value. Show that this conditional distribution has probability density

$$f(\hat{\xi}; \xi | A = a) = \{2K_0(2a)\}^{-1} \exp\{-2a \cosh(\hat{\xi} - \xi)\}, \ \hat{\xi} \in \mathbb{R},$$

in terms of the function $K_0(\cdot)$, where

$$K_0(x) = \int_0^\infty e^{-x \cosh t} dt.$$

ANSWER: (Seen similar)

Note that $\hat{\xi} = \log(\hat{\psi})$ and by the above $(\hat{\xi}, A)$ is minimal sufficient, with A ancillary, i.e. distribution constant. The Conditionality Principle says that if (T, A) is minimal sufficient for a parameter θ, with A ancillary, inference should be made based on the conditional distribution of $T|A = a$: we should condition on the observed value of the ancillary. Applying this here, inference on ξ should be based on the conditional distribution of $\hat{\xi}$ given $A = a$.

Write $\xi = \log \psi$, so that $\psi = e^\xi$. Then $\hat{\xi} = \log \hat{\psi}$, and consider the transformation $(X, Y) \to (\hat{\xi}, A) = (\frac{1}{2} \log(\frac{Y}{X}), \sqrt{XY})$.

The inverse transformation has $X = Ae^{-\hat{\xi}}$, $Y = Ae^{\hat{\xi}}$, and the Jacobian is

$$J = \begin{bmatrix} \frac{dX}{dA} & \frac{dX}{d\hat{\xi}} \\ \frac{dY}{dA} & \frac{dY}{d\hat{\xi}} \end{bmatrix} = \begin{bmatrix} e^{-\hat{\xi}} & -Ae^{-\hat{\xi}} \\ e^{\hat{\xi}} & Ae^{\hat{\xi}} \end{bmatrix},$$

with $|J| = 2A$.

Since $f(x, y; \xi) = \exp\{-e^\xi x - e^{-\xi} y\}$, the joint density of $(\hat{\xi}, A)$ is

$$f(\hat{\xi}, a; \xi) = 2a \exp\{-e^\xi a e^{-\hat{\xi}} - e^{-\xi} a e^{\hat{\xi}}\} = 2a \exp\{-2a \cosh(\hat{\xi} - \xi)\}.$$

Then the marginal density of A is

$$f(a) = 2a \int_{-\infty}^\infty e^{-2a \cosh(\hat{\xi} - \xi)} d\hat{\xi} = 4aK_0(2a),$$

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

and the exact conditional density of $\hat{\xi}|A = a$ is

$$f(\hat{\xi}|a;\xi) = \frac{f(\hat{\xi}, a;\xi)}{f(a)} = \frac{1}{2K_0(2a)} \exp\{-2a \cosh(\hat{\xi} - \xi)\}.$$

[10 marks]

[Total 25 marks]

[Recall that $\cosh(z) = (e^z + e^{-z})/2$.]

3.

(i) Outline briefly the use and properties of conditional hypothesis tests in multi-parameter exponential families.

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) $E_\theta\{\phi_0(Y)\} \le \alpha$ for all $\theta \in \Theta_0$; (b) given any other test $\phi(\cdot)$ for which $E_\theta\{\phi(Y)\} \le \alpha$ for all $\theta \in \Theta_0$, we have $E_\theta\{\phi_0(Y)\} \ge E_\theta\{\phi(Y)\}$ for all $\theta \in \Theta_1$.

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 $E_\theta\{\phi(Y)\} \ge \alpha$ 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 $E_\theta\{\phi_0(Y)\} \ge E_\theta\{\phi(Y)\}$ for all $\theta \in \Theta_1$, and any test $\phi(\cdot)$ which is unbiased of size α.

Consider a random vector Y with density belonging to a full exponential family of the form

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

We are considering the multi-parameter case: $m > 1$. Write $T_i = t_i(Y)$ and suppose we wish to test a hypothesis about θ^1, with $\theta^2, \dots, \theta^m$ nuisance. Then optimal (UMPU) tests are based on the conditional distribution of T_1, given $T_i = t_i, i = 2, \dots, m$, 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 $H_0 : \theta^1 \le \theta^{1*}$ against $H_1 : \theta^1 > \theta^{1*}$, for a specified θ^{1*}, the test is of the form: reject H_0 if $T_1 > t_1^c$, where we require

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

for a test of required size α. [10 marks]

(ii) Let $X_1, \dots, X_n$ be independent, identically distributed exponential random variables with common mean λ, and $Y_1, \dots, Y_n$ be independent, identically distributed exponential random variables with common mean μ. Suppose the two samples are independent.

Consider testing $H_0 : \lambda \le \mu$ against $H_1 : \lambda > \mu$. What is the form of the optimal, uniformly most powerful unbiased, test?

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

ANSWER: (Seen similar)

The joint density can be written, in terms of $X = \sum_{i=1}^n X_i$ and $Y = \sum_{i=1}^n Y_i$, as

$$\begin{aligned} f(x, y; \lambda, \mu) &= \frac{1}{(\lambda\mu)^n} \exp\{-\frac{1}{\lambda}X - \frac{1}{\mu}Y\} \\ &\propto \exp\{-(\frac{1}{\lambda} - \frac{1}{\mu})X - \frac{1}{\mu}(X+Y)\} \\ &\equiv \exp\{\pi_1 X + \pi_2(X+Y)\}, \end{aligned}$$

say, where, $\pi_1 = -(\frac{1}{\lambda} - \frac{1}{\mu})$ and

$$\lambda \leq \mu \equiv \frac{1}{\lambda} \geq \frac{1}{\mu} \equiv \frac{1}{\lambda} - \frac{1}{\mu} \geq 0 \equiv \pi_1 \leq 0.$$

So, we want to test $H_0 : \pi_1 \leq 0$ against $H_1 : \pi_1 > 0$, with π_2 a nuisance parameter. The theory described in (i) tells us that a UMPU test exists and is of the conditional form: Reject H_0 if $X > c(v)$, where v is the observed value of $V = X + Y$, and $c(v)$ is fixed so that

$$P_{\pi_1=0}(X > c(v)|V = v) = \alpha,$$

the required size of test. [7 marks]

Professor Z remarks: 'If we define

$$T = \frac{X}{X+Y},$$

where $X = \sum_{i=1}^n X_i$ and $Y = \sum_{i=1}^n Y_i$, then it is obvious that if H_0 is false T will tend to be large, so instead of messing around doing a conditional test, just reject H_0 if $T > c$, where c is fixed so that $P(T > c)$ is the size α of test we are after, when $\lambda = \mu$. If $\lambda = \mu$, T has a Beta distribution, or something.'

What do you think of Professor Z's ideas?

ANSWER: (Seen similar)

Note that the conditional UMPU test can equivalently be expressed as: Reject H_0 if $T > d(v)$ conditional on $V = v$, with

$$P_{\pi_1=0}(T > d(v)|V = v) = \alpha.$$

We have X distributed as $\mathsf{Gamma}(n, 1/\lambda)$ and Y distributed as $\mathsf{Gamma}(n, 1/\mu)$, independently.

Consider the transformation $T = X/(X+Y), V = X + Y$. This has inverse $X = TV, Y = V - TV$, and the Jacobian is

$$J = \begin{bmatrix} \frac{dX}{dT} & \frac{dX}{dV} \\ \frac{dY}{dT} & \frac{dY}{dV} \end{bmatrix} = \begin{bmatrix} V & T \\ -V & 1-T \end{bmatrix},$$

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

with $|J| = V$.

When $\lambda = \mu$, the joint probability density of (T, V) is then

$$
\begin{aligned}
f(t, v; \lambda = \mu) &= \left(\frac{1}{\lambda}\right)^{2n} \frac{(tv)^{n-1} v^{n-1} (1-t)^{n-1} e^{-v/\lambda}}{\Gamma(n)^2} \cdot v \\
&\equiv \frac{\Gamma(2n) t^{n-1} (1-t)^{n-1}}{\Gamma(n)\Gamma(n)} \cdot \frac{v^{2n-1} e^{-v/\lambda}}{\lambda^{2n}\Gamma(2n)},
\end{aligned}
$$

from which we see that T, V are independent, with T distributed as $\text{Beta}(n, n)$ and V distributed as $\text{Gamma}(2n, 1/\lambda)$. [It is feasible that this could be quoted: if done accurately, no deduction]. This independence when $\lambda = \mu(\pi_1 = 0)$ implies that the UMPU conditional test is equivalent to a test which rejects if $T > c$, where c is such that $P(\text{Beta}(n, n) > c) = \alpha$. So, Professor Z's ideas are exactly equivalent to the UMPU conditional test. [8 marks]

[Total 25 marks]

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

   (i) the James-Stein estimator and admissibility;

   (ii) optimal point estimation in the Inverse Gaussian distribution;

   (iii) principles of decision theory;

   (iv) large sample testing procedures based on maximum likelihood estimators.

   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]