Data provided: Tables of distribution functions

MAS6004

The University Of Sheffield.

SCHOOL OF MATHEMATICS AND STATISTICS **Spring Semester 2016–2017**

Inference **Solutions** **2 hours**

Candidates may bring to the examination a calculator which conforms to University regulations.

Marks will be awarded for your best ***five*** *answers. Total marks 100.*

Standard results from the lecture notes may be used without derivation, but must be clearly stated.

1 (i) (a) (Unseen) $H_0 : \beta = 0$✓

(b) (Unseen) We've used a randomisation test. Under H_0, there is no relationship between x and y, and so it should not matter if we permute the elements of x.✓✓

(c) (Routine) The test gives an estimated p-value of $31/1000 = 0.031$, and so we reject H_0 at the 5% level, but not at the 1% level.✓✓

(d) (Routine) The plot suggests that the distributional assumptions made about ϵ for the t-test may not hold, in particular, the constant variance assumption looks questionable. The randomization test, because it does not rely on a constant variance assumption, may be preferred to the classical t-test. ✓✓

(e) (Unseen) If there were only 5 observations, then there are $5! = 120$ different permutations. If the observed ordering gives the largest values of $\hat{\beta}$, then the reverse ordering would give $-\hat{\beta}$, and so the smallest possible p-value would be $2/120$.✓✓

(ii) (Unseen) For $n = 1, \ldots, N$

- Simulate $\tilde{y}_i \sim N(0, x_i^2)$ for $i = 1, \ldots, 100$✓
- Calculate $\hat{\beta}_n = \arg\min_{\beta} \sum_{i=1}^{100} \left(\frac{\tilde{y}_i - \beta x_i}{x_i} \right)^2$ ✓

Find

$$p = \frac{1}{N} \sum_{n=1}^{N} \mathbb{I}_{|\hat{\beta}_m| \geq |\hat{\beta}_{obs}|}$$

✓and reject H_0 if $p < 0.01$, otherwise conclude there is insufficient evidence against H_0.✓

(iii) (a) (Bookwork) If $L(\alpha, \beta|\mathbf{x}) = \prod_{i=1}^{n} f(x_i; \alpha, \beta)$ is the likelihood function, then the profile likelihood for α is

$$L_p(\alpha) = \max_{\beta} L(\alpha, \beta|\mathbf{x})✓✓$$

(b) (Bookwork) If $l_p(\alpha) = \log L_p(\alpha)$ is the profile log-likelihood, then we define the profile deviance to be

$$D_p(\alpha) = 2\{l_p(\hat{\alpha}) - l_p(\alpha)\}✓✓$$

For n large, at the true value of α, $D_p(\alpha) \sim \chi_1^2$,✓and so a 95% confidence interval for α takes the form

$$C_\alpha = \{\alpha : D_p(\alpha) \leq c_{0.05}\}✓✓$$

where $c_{0.05}$ is the 95th percentage point of the χ_1^2 distribution.

2 (i) (Bookwork)

$$\hat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}_{X_i \leq x} \checkmark\checkmark\checkmark$$

(ii) (Bookwork)

$$n\hat{F}_n(x) \sim Bin(n, F(x))\checkmark\checkmark$$

and so $\hat{F}_n(x) \sim \frac{1}{n}Bin(n, F(x))\checkmark$

(iii) (Routine)

$$\begin{aligned}\gamma(\hat{F}_n) &= \frac{\mathbb{E}_{\hat{F}_n}(X - \mathbb{E}_{\hat{F}_n}(X))^3}{\left(\mathbb{E}_{\hat{F}_n}(X - \mathbb{E}_{\hat{F}_n}(X))^2\right)^{3/2}}\checkmark\checkmark \\ &= \frac{\frac{1}{n}\sum(X_n - \bar{X}_n)^3}{\left(\frac{1}{n}\sum(X_n - \bar{X}_n)^2\right)^{3/2}}\checkmark\checkmark\checkmark\end{aligned}$$

(iv) (Bookwork) By sampling with replacement from $\{X_1, \ldots, X_n\}\checkmark\checkmark$

(v) (Unseen) For $b = 1, \ldots, B\checkmark$

- Sample $X_1*, \ldots, X_n^* \sim \hat{F}_n(\cdot)\checkmark$
- Calculate $T_b = \gamma(\mathbf{X}^*)\checkmark$

Then estimate the standard error using

$$\text{std. err}(\hat{\gamma}) = \left(\frac{1}{B-1}\sum_{b=1}^{B}(T_b - \bar{T})^2\right)^{1/2}$$

where $\bar{T} = \frac{1}{B}\sum_{b=1}^{B} T_b\checkmark\checkmark$

A 95% CI could be formed either by looking at the 2.5th and 97.5th percentiles of the $\{T_b\}_{b=1}^{B}$ or by

$$\hat{\gamma} \pm 1.96\,\text{std. err}(\hat{\gamma})$$

$\checkmark\checkmark$

3 A common model used in experimental design to investigate whether the mean of several populations is the same or not can be written as

$$y_{ij} = \mu_j + \varepsilon_i \; ; \qquad i = 1, \dots, n_j \; ,$$
$$\mu_j \sim \mathrm{N}\big(\mu_j \,\big|\, \eta, 1/t\big) \; , \quad j = 1, \dots, k \; ,$$

and $\varepsilon_i \sim \mathrm{N}(\cdot \,|\, 0, 1/\lambda)$, all independent. Where $\boldsymbol{\mu} = \{\mu_1, \dots, \mu_k\}$, η and λ are unknown parameters. Let the prior be

$$\pi(\eta, \lambda) = \mathrm{N}(\eta \,|\, m, 1/p)\, \mathrm{Ga}(\lambda \,|\, a, b) \; .$$

with $\{t, m, p, a, b\}$ fixed.

(i) (a) Show that the full conditional distribution of each μ_j is $\mathrm{N}\big(\mu_j \,\big|\, m_j^\star, 1/t_j^\star\big)$ and give explicit expressions for the parameters. ***(4 marks)***

To derive the full conditional of the random coefficients, note that

$$\pi(\mu_j \,|\, \text{—}) \propto \left[\prod_{i=1}^{n_j} \exp\left[-\frac{\lambda}{2}\big(y_{ij} - \mu_j\big)\right]\right] \exp\left[-\frac{t}{2}(\mu_j - \eta)^2\right]$$
$$\propto \exp\left[\frac{t_j^\star}{2}(\mu_j - m_j{}^\star)^2\right]$$

with $t_j^\star = n_j\lambda + t$ and $m_j{}^\star = \big(n_j\lambda\bar{y}_j + t\eta\big)/t_j^\star$.

3M 1A

(b) Show that the full conditional distribution of η is $\mathrm{N}(\eta \,|\, m^\star, 1/p^\star)$ and give explicit expressions for the parameters. ***(6 marks)***

For the overall mean,

$$\pi(\eta \,|\, \text{—}) \propto \exp\left[-\frac{t}{2}\sum_{j=1}^{k}(\mu_j - \eta)^2\right] \exp\left[-\frac{p}{2}(\eta - m)^2\right]$$
$$\propto \exp\left[\frac{p^\star}{2}(\eta - m^\star)^2\right]$$

with $p^\star = p + kt$ and $m^\star = (kt\bar{\mu} + pm)/p^\star$.

4M 2A

(c) Show that the full conditional distribution of λ is $\mathrm{Ga}(\lambda \,|\, a^\star, b^\star)$ and give explicit expressions for the parameters. ***(3 marks)***

This is straightforward, $\pi(\lambda \,|\, \text{—}) = \mathrm{Ga}(\lambda \,|\, a^\star, b^\star)$, with

$$a^\star = a + \frac{1}{2}\sum_{j=1}^{k} n_j \quad \text{and} \quad b^\star = b + \frac{1}{2}\sum_{i=1}^{n_j}(y_{ij} - \mu_j)^2 \; .$$

2M 1A

 Question 3 continued on next page

3 (continued)

(ii) Write down pseudo-code for an MCMC scheme to explore the posterior distribution $\pi(\boldsymbol{\mu}, \eta, \lambda \mid \boldsymbol{y})$. ***(7 marks)***

Start by fixing $\left\{\boldsymbol{\mu}^{(0)}, \eta^{(0)}, \lambda^{(0)}\right\}$. For $d = 1, \dots, M$

(a) sample ${\mu_j}^{(d)}$ from $\mathrm{N}\left(\cdot \,\middle|\, \left(n_j \lambda^{(d-1)} \bar{y}_j + t\eta^{(d-1)}\right)/t_j^{\star(d-1)}, t_j^{\star(d-1)}\right)$, for $j = 1, \dots, k$.

(b) sample $\eta^{(d)}$ from $\mathrm{N}\left(\cdot \,\middle|\, (kt\bar{\mu}^{(d)} + pm)/p^{\star(d-1)}, p^{\star(d-1)}\right)$

(c) sample $\lambda^{(d)}$ from $\mathrm{Ga}\left(\cdot \,\middle|\, a + \frac{1}{2}\sum_{j=1}^{k} n_j, b + \frac{1}{2}\sum_{i=1}^{n_j}\left(y_{ij} - {\mu_j}^{(d)}\right)^2\right)$

4M 3A

4 A branch manager is interested in the rate of clients served in a day, θ. Through a typical period he records a random sample of clients served by day $\boldsymbol{x} = \{x_1, \ldots, x_n\}$ and assumes $x_i \sim \mathrm{Po}(x_i \mid \theta)$. He decides to use $\pi(\theta) = \mathrm{Ga}(\theta \mid a, b)$ as a prior.

(i) Show that his posterior distribution is $\mathrm{Ga}(\theta \mid a^\star, b^\star)$ and provide explicit expressions for the posterior parameters. ***(2 marks)***

$$\pi(\theta \mid \boldsymbol{x}) \propto \theta^{s+a-1} \exp[-\theta(b+n)] = \mathrm{Ga}(\theta \mid a^\star, b^\star)\,.$$

with $s = \sum x_i$. Here $a^\star = a + s$ and $b^\star = b + n$

1M 1A

Question 4 continued on next page

4 (continued)

(ii) Using past records of similar branches the manager elicits $\mathbb{E}[\theta] = 10/3$ and $\mathbb{V}[\theta] = 50/9$ and obtains $n = 40$ and $\sum_{i=1}^{40} x_i = 425.3$ from the sample.

(3 marks)

To get the prior parameters, solve

$$\frac{a}{b} = \frac{10}{3}, \quad \frac{a}{b^2} = \frac{50}{9} \qquad \Longrightarrow a = 2, b = 0.6.$$

2M 1A

(a) Calculate his prior and posterior point estimates under a quadratic loss function,

$$\mathcal{L}(\theta, \widehat{\theta}) = \left(\theta - \widehat{\theta}\right)^2,$$

and the associated expected losses. ***(5 marks)***

This are the prior/posterior mean and variance, respectively. Thus

$$\mathbb{E}[\theta] = 10/3, \ \mathbb{V}[\theta] = 50/9, \ \mathbb{E}[\theta \mid \boldsymbol{x}] = 10.52\,, \ \mathbb{V}[\theta \mid \boldsymbol{x}] = 0.26\,.$$

3M 1A

(b) Calculate his posterior point estimate under the absolute loss function,

$$\mathcal{L}(\theta, \widehat{\theta}) = \left|\theta - \widehat{\theta}\right|,$$

assuming the posterior distribution can be approximated by a Gaussian. ***(5 marks)***

Using the Gaussian approximation, the posterior mean, media and mode are the same, thus this is the posterior mean.

3M 2A

(c) Using a zero-one loss function,

$$\mathcal{L}(\theta, \widehat{\theta}) = \begin{cases} 0 & |\theta - \widehat{\theta}| < c \\ 1 & |\theta - \widehat{\theta}| \geq c \end{cases},$$

and assuming $c \to 0$, calculate his prior and posterior point estimates.

(5 marks)

This is the prior/posterior mode:

$$\mathrm{Mode}[\theta] = \frac{5}{3} \qquad \mathrm{Mode}[\theta \mid \boldsymbol{x}] = 10.5\,.$$

2M 1A

5 (i) (a) (Bookwork) Let $w(x) = \dfrac{f(x)}{g(x)}$. To estimate S, for $i = 1, \ldots, N$

- Simulate $X_i \sim g(\cdot)$
- Set $w_i = w(X_i)$ ✓

Then estimate S by

$$\hat{S} = \frac{1}{n} \sum w_i X_i^2 \checkmark$$

(b) (Routine) First find the mode.

$$\frac{d}{dx} h(x) = \frac{3}{x} - 4x^3 \checkmark$$

which has a minimum at

$$m = \left(\frac{3}{4}\right)^{1/4} \checkmark$$

A second order Taylor expansion of $h(x)$ about m is

$$h(x) = h(m) + \frac{1}{2} h''(m)(x-m)^2 \checkmark$$

where

$$h''(x) = -\frac{3}{x^2} - 12x^2$$

and so

$$h''(m) = -12 \left(\frac{4}{3}\right)^{1/2} = -13.9 \checkmark$$

and so

$$h(x) = h(m) - \frac{1}{2}(x-m)^2 / 0.0721$$

So the optimal importance sampling distribution is

$$N\left(\left(\frac{3}{4}\right)^{1/4}, \frac{1}{12}\left(\frac{3}{4}\right)^{1/2}\right) = N(0.93, 0.072) \checkmark \checkmark$$

 Question 5 continued on next page

5 (continued)

(ii) (a) (Unseen) The probability a $N(0,\sigma^2)$ rv is greater than a is $1-\Phi(a/\sigma)=\Phi(-a/\sigma)$ $\checkmark$. Thus, the amount of time we have to wait until we see an acceptance has a $\mathrm{Geometric}(\Phi(-a/\sigma))$ $\checkmark$ distribution, which has mean

$$1/\Phi(-a/\sigma)\checkmark$$

(b) (Routine)

$$\begin{aligned} M &= \sup_x \frac{f(x)}{g(x)} = \sup_x \frac{c\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{x^2}{2\sigma^2}}\mathbb{I}_{x>a}}{\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}}\checkmark \\ &= \sup_x c\exp\left(-\frac{1}{2\sigma^2}(2x\mu-\mu^2)\right)\mathbb{I}_{x>a}\checkmark \end{aligned}$$

This is maximised where $2x\mu-\mu^2$ is minimized, which is at $x=a$ (given that $\mu>0$ and $x>a$).$\checkmark$

So

$$M = c\exp\left(-\frac{1}{2\sigma^2}(2a\mu-\mu^2)\right)$$

$\checkmark$

So the rejection algorithm is

- Simulate $X\sim N(\mu,\sigma^2)$
- Accept X with probability $\dfrac{f(X)}{Mg(X)}\checkmark\checkmark$

(c) (Unseen) The efficiency of the rejection algorithm is determined by the acceptance rate, which is $1/M$ $\checkmark$. So we should choose μ to minimize M (and thus maximize $1/M$).$\checkmark$

M is minimized when we maximize $2a\mu-\mu^2$, which occurs at $\mu=a$ $\checkmark$

6 An engineer is testing a new precision weighing device. In her experimental design n pieces of titanium of identical known weight are measured and the relative discrepancy in measure, $\boldsymbol{y} = \{y_1, \ldots, y_n\}$ is recorded and it is assumed $y_i \sim \mathrm{Un}(y_i \mid 0, \theta)$, where θ represents the maximum technical discrepancy of the device.

(i) Sketch the likelihood function and show that $\widehat{\theta} = y_{(n)} = \max\{y_1, \ldots, y_n\}$ is the MLE. ***(3 marks)***

Prior and likelihood are plotted below. Notice that the likelihood is decreasing and non-zero only for values of $\theta \geq y_{(n)}$, thus this is the MLE.

2M 1A

(ii) The engineer decides to use a Pareto $\mathrm{Pa}(\theta \mid a, b)$

$$\pi(\theta) = ab^a \theta^{-(a+1)}, \qquad \theta > b, \quad a, b > 0,$$

as a prior distribution.

(a) Sketch the engineer's prior distribution. ***(3 marks)***

See figure above.

2M 1A

(b) Show that her posterior distribution is $\mathrm{Pa}(\theta \mid a^\star, b^\star)$, with $a^\star = n + a$ and $b^\star = \max\left\{b, \widehat{\theta}\right\}$. ***(7 marks)***

The posterior is non-zero only in the overlapping region where both prior and likelihood are non-zero, thus

$$\pi(\theta \mid \boldsymbol{y}) \propto \theta^{-(n+a+1)} \qquad \theta \geq t = \max\left\{b, \widehat{\theta}\right\}$$

the normalising constant is

$$\int_t^\infty \theta^{-(n+a+1)}\, \mathrm{d}\theta = \frac{1}{n+a} t^{-(n+a)}.$$

Thus the posterior is $\mathrm{Pa}(\theta \mid a^\star, b^\star)$, with $a^\star = a + n$ and $b^\star = t$.

5M 2A. Give full marks if posterior is given up to a proportionality constant.

(c) Discuss the implications on the Bayesian learning process if $b > \widehat{\theta}$. ***(3 marks)***

If $b^\star = b$ then the posterior distribution does not contain any sufficient statistics from the data.

Award 1M for commenting on the sample space restriction. Give 3M for commenting the posterior does not depend on a sufficient statistic.

 Question 6 continued on next page

6 (continued)

(iii) Provide the HPD interval of size 0.95 if $n = 10$, $\widehat{\theta} = 0.5$, $a = 3$ and $b = 0.4$.

(4 marks)

The HPD must contain the mode $b^\star$ and clearly is a one sided interval. Thus the upper limit, q, is given by

$$0.95 = \int_{b^\star}^{q} (n + a)\, {b^\star}^{n+a} \theta^{-(n+a+1)}\, \mathrm{d}\theta$$
$$\Longrightarrow q = b^\star (1 - 0.95)^{-1/a^\star}$$

For the data at hand this is $(0.5, 0.63)$.

1M 3A

End of Question Paper