Data provided: Tables of distribution functions

MAS6004

The University Of Sheffield.

SCHOOL OF MATHEMATICS AND STATISTICS **Spring Semester 2014–2015**

Inference **2 hours**

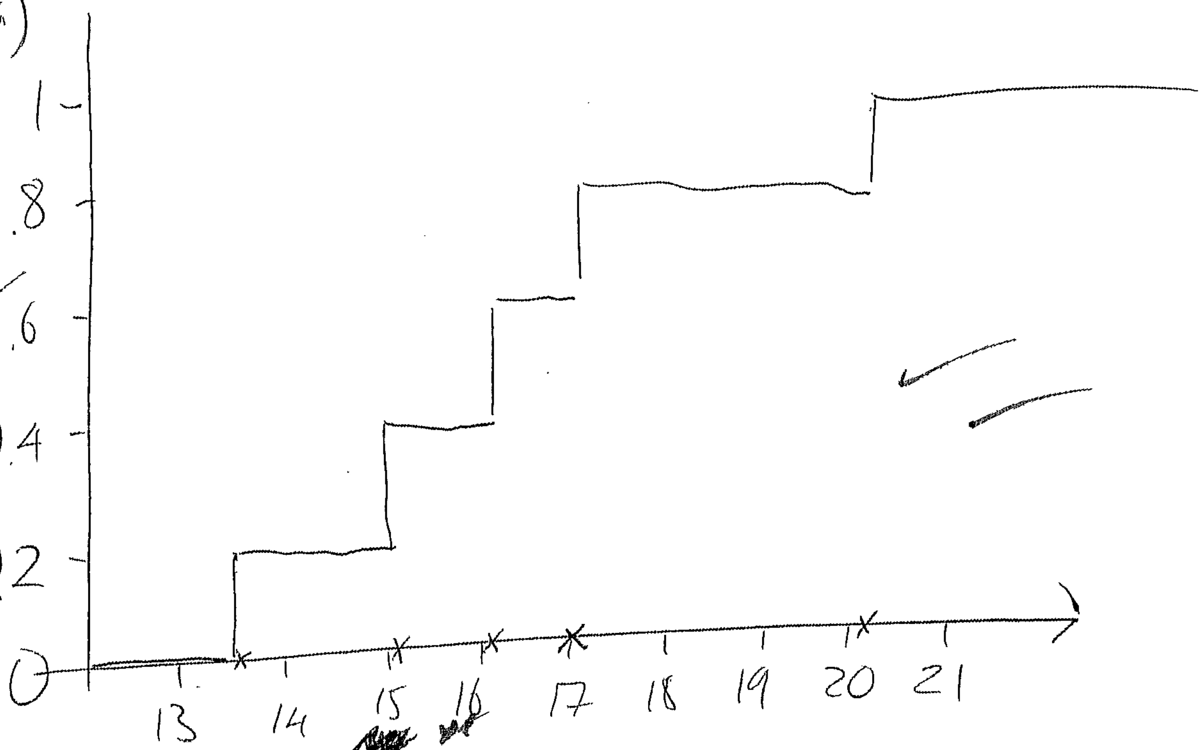
Empty rubric

Please leave this exam paper on your desk
Do not remove it from the hall

Registration number from U-Card (9 digits)
to be completed by student

Blank

1/ (i) a)



b) Pick X_1 if U_1 lies in $[0, 0.2)$

X_2 .. U_2 ..... $[0.2, 0.4)$ etc

[NB - other possible approaches here]

giving 13.6, 20.2, 20.2, 16.1, 16.1

& so the bootstrap median is 16.1

c) To estimate the std error of the median, we would repeat this to get B bootstrap estimates $m_1, \ldots, m_B$ say & then calculate the observed / sample std. deviation

$$s = \left(\frac{1}{B-1} \sum (m_i - \bar{m})^2 \right)^{1/2}$$

1) c) d) (iii) a) Use ~~$T = \text{Var}(x_i) = \frac{1}{S-1} \sum (t_i - \bar{t})^2$~~

~~as a test statistic.~~

To conduct the MC test, we first fit the exponential dist^n to the data & estimate

$$\hat{\lambda} = \frac{1}{\bar{t}}$$

We then do

For $i = 1, \ldots, B$

- Simulate $\tilde{t}_1, \ldots, \tilde{t}_{85} \sim \text{Exp}(\hat{\lambda})$
- Calculate $\tilde{T}_i = \frac{1}{85-1} \sum_{j=1}^{85} (\tilde{t}_j - \bar{\tilde{t}})^2$

& estimate the p-value by

$$\frac{1}{B} \left(\sum \mathbb{I}_{\tilde{T}_i \leq T_{obs}} + 1 \right)$$

~~as we expect~~ ~~Under H_0~~ If the null hypothesis of independence is incorrect, we expect the variance to be smaller & so we are looking to see if T_{obs} is smaller compared to what is expected under H_0

b) We could generate a p-value of 0.05 with as few as 19 ~~test~~ statistics.

This would not be advisable ∵ the test is random & so even if the true p-value is >0.05 there is a reasonable probability we would not observe any simulated values less than T_{obs} if we only used 19 samples.

Q2

(ii) $Var(\overline{X_{2n}}) = Var\left(\frac{1}{2n}\sum_{i=1}^{2n} X_i\right)$

$= \frac{1}{4n^2}\sum Var(X_i)$ ~~[illegible]~~

$= \frac{1}{2n} Var(X_1)$ ✓

$Var(X_1) = \mathbb{E}X_1^2 - (\mathbb{E}X_1)^2 = \frac{1}{1^2} = 1$ ✓ (can quote)

Q2 ctd (i)(a)

Thus $Var(\overline{X}_{2n}) = \frac{1}{2n}$

(b) $Var\left(\frac{\overline{X}_n + \overline{Y}_n}{2}\right) = \frac{1}{4} Var\left(\frac{\sum_{i=1}^{n}(X_i + Y_i)}{n}\right)$

$= \frac{n}{4n^2} Var(X_1 + Y_1)$ as

$X_i, Y_i \perp\!\!\!\perp X_j, Y_j$ for $j \neq i$

$Var(X_1 + Y_1) = Var(X_1) + Var(Y_1) + 2\,Cov(X_1, Y_1)$

$= 1 + 1 + 2\left(\mathbb{E}(X_1 Y_1) - \mathbb{E}X_1 \mathbb{E}Y_1\right)$

$= 2\mathbb{E}(X_1 Y_1)$

$= 2\int_0^1 (-\log u)(-\log(1-u))\,du$

$= 2\int_0^1 \log x \log(1-x)\,dx = 2I$

Thus $Var\left(\frac{\overline{X}_n + \overline{Y}_n}{2}\right) = \frac{I}{2n}$ as required.

(c) $\mathbb{E}(\overline{X_{2n}}) = \mathbb{E}\left(\frac{\overline{X_n} + \overline{Y_n}}{2}\right) = 1$

so both $\overline{X_{2n}}$ & $\frac{\overline{X_n} + \overline{Y_n}}{2}$ estimate the mean & they both use $2n$ random draws.

But $Var\left(\frac{\overline{X_n} + \overline{Y_n}}{2}\right) = \frac{1}{2n}\left(2 - \frac{\pi^2}{6}\right) \leq \frac{1}{2n} = Var(\overline{X_{2n}})$

as $1 < \frac{\pi^2}{6} < 2$

& so $\frac{\overline{X_n} + \overline{Y_n}}{2}$ is more accurate than $\overline{X_{2n}}$ (lower variance)

This is an example of the use of antithetic variables.

(iii) $L(t \mid \alpha, \beta) = \prod_{i=1}^{4} (\alpha\beta)(\beta t_i)^{\alpha-1} e^{-(\beta t_i)^\alpha}$

$= (\alpha\beta)^4 \beta^{4\alpha-4} \left(\prod t_i\right)^{\alpha-1} e^{-\sum(\beta t_i)^\alpha}$

So $l(t \mid \alpha, \beta) = 4\log\alpha + 4\alpha\log\beta - (\alpha-1)\sum\log t_i - \sum(\beta t_i)^\alpha$

$\frac{dl}{d\beta} = \frac{4\alpha}{\beta} - \alpha\beta^{\alpha-1}\sum t_i^\alpha$

& thus $\hat{\beta} = \left(\frac{4}{\sum t_i^\alpha}\right)^{1/\alpha}$

Hence $l(p) = 4\log\alpha + 4\log\left(\frac{4}{\sum t_i^\alpha}\right) + (\alpha-1)\sum\log t_i - 4$

(ii) $l_p(\hat{\alpha}) = -12.2$

$l_p(1) = 0 + 4\log\left(\frac{4}{42}\right) + 0 - 4$

$= -13.41$

So $D_p(1) = -2(l_p(1) - l_p(\hat{\alpha}))$

$= -2(-13.41 - -12.2)$

$= 2.42$

Compare to a χ^2_1 distribution

$2.42 < \chi^2_1(0.95) = 3.84$ so no evidence ~~to~~ to reject H_0 at 5% level.

~~3~~ a)

$$L(\lambda) = \prod_{i=1}^{n} \lambda e^{-\lambda t_i} \prod_{j=1}^{m} \lambda e^{-\lambda s_j} = \lambda^{m+n} \exp\left(-\lambda\left(\sum t_i + \sum s_j\right)\right)$$

So $$\log L(\lambda) = (m+n)\log\lambda - \lambda\left(\sum_{i=1}^{n} t_i + \sum_{j=1}^{m} s_j\right)$$

(b) $\mathbb{E}(S \mid S \leq h, \lambda)$ requires us to calculate the pdf of S given $S \leq h$.

$$\mathbb{P}(S \leq t \mid S \leq h) = \frac{\mathbb{P}(S \leq t)}{\mathbb{P}(S \leq h)} = \frac{1-e^{-\lambda t}}{1-e^{-\lambda h}}$$

Then $$\mathbb{E}(S \mid S \leq h, \lambda) = \int_0^h \mathbb{P}(S > t \mid S \leq h)\, dt$$

$$= \int_0^h 1 - \frac{1-e^{-\lambda t}}{1-e^{-\lambda h}}\, dt$$

$$= h - \frac{1}{1-e^{-\lambda h}}\left[t + \frac{e^{-\lambda t}}{\lambda}\right]_0^h$$

$$= h - \frac{1}{1-e^{-\lambda h}}\left(h + \frac{e^{-\lambda h}}{\lambda} - \frac{1}{\lambda}\right)$$

$$= \frac{1}{\lambda} - \frac{h e^{-\lambda h}}{1-e^{-\lambda h}}$$ as req^d.

[Or we can find pdf $\dfrac{dG(t)}{dt} = \dfrac{\lambda e^{-\lambda t}}{1-e^{-\lambda h}}$

& use $\displaystyle\int_0^h t\,g(t)\,dt = \mathbb{E}(S \mid S \leq h)$]

(c) $Q(\lambda \mid \lambda^{(r)}) = \mathbb{E}_{\lambda^{(r)}}\left(\log L(\lambda \mid t_1,..,t_n, S_1,..,S_m) \;\middle|\; \begin{matrix} t_1,..,t_n \\ r \text{ of the } S_i < h \\ m-r \text{ of the } S_i > h, \lambda^{(r)} \end{matrix}\right)$

$$= (m+n)\log\lambda - \lambda\left(\sum_{i=1}^{n} t_i + (m-r)\,\mathbb{E}(S \mid S>h) + r\,\mathbb{E}(S \mid S \leq h)\right)$$

$$= (m+n)\log\lambda - \lambda\left(n\bar{t} + (m-r)\left(h + \frac{1}{\lambda^{(r)}}\right) + r\left(\frac{1}{\lambda^{(r)}} - \frac{h e^{-h\lambda^{(r)}}}{1-e^{-h\lambda^{(r)}}}\right)\right)$$

as $\mathbb{E}(S \mid S>h) = h + \frac{1}{\lambda^{(r)}}$ because S is exponential & has the memoryless property

So that $\pi(s \mid S>h) = \begin{cases} e^{-\lambda(s-h)} & s \geq h \\ 0 & \text{otherwise} \end{cases}$

(d) Thus to find $\lambda^{(r+1)}$ we minimize $Q(\lambda | \lambda^{(r)})$

$$\frac{dQ}{d\lambda} = \frac{m+n}{\lambda} - \left(n\bar{t} + (m-r)\left(h + \frac{1}{\lambda^{(r)}}\right) + r\left(\frac{1}{\lambda^{(r)}} - \frac{he^{-h\lambda^{(r)}}}{1-e^{-h\lambda^{(r)}}}\right)\right)$$

& so $$\lambda^{(r+1)} = \frac{m+n}{n\bar{t} + (m-r)\left(h + \frac{1}{\lambda^{(r)}}\right) + r\left(\frac{1}{\lambda^{(r)}} - \frac{he^{-\lambda h^{(r)}}}{1-e^{-h\lambda^{(r)}}}\right)}$$

4 In microscopic imaging it is common to model the number of photos arriving at the lens in each frame, X, as $\mathrm{Po}(x \mid \lambda)$, where λ is the rate of photon emission per frame. Given a random sample, $\boldsymbol{x} = \{x_1, \ldots, x_n\}$,

(i) (a) Show that $\pi(\lambda) = \mathrm{Ga}(\lambda \mid a, b)$ is a conjugate prior and give explicit expressions for the posterior parameters. ***(5 marks)***

The likelihood is

$$\mathrm{L}(\lambda\,;\,\boldsymbol{x}) \propto \prod \lambda^{x_i} e^{-\lambda} \propto \lambda^{s} e^{-\lambda n}\ ,\ s = \sum x_i$$

Thus

$$\pi(\lambda \mid \boldsymbol{x}) \propto \lambda^{a^\star - 1} e^{-\lambda b^\star}$$

with $a^\star = a + s$, $b^\star = b + n$, is the kernel of a $\mathrm{Ga}(\lambda \mid a^\star, b^\star)$.

[3M 2A]

(b) Find the Bayes estimator for λ under 0-1 loss. ***(3 marks)***

The Bayes estimate is the posterior mode, $\widehat{\lambda} = (a^\star - 1)/b^\star$, which can be found by taking the derivative of the log posterior and solving for zero.

[1M 2A]

 Question 4 continued on next page

4 (continued)

(ii) (a) Calculate the predictive distribution of Y, the number of photons captured by the lens in the next random sample of m frames,

$$Y = \sum_{j=n+1}^{n+m} X_j$$

(7 marks)

Using standard probability results, $Y \sim \text{Po}(y \mid m\lambda)$. Thus,

$$\begin{aligned} f(y \mid \boldsymbol{x}) &= \int_0^\infty \frac{(m\lambda)^y}{y!} e^{-m\lambda} \frac{b^{\star a^\star}}{\Gamma[a^\star]} \lambda^{a^\star - 1} e^{-b^\star \lambda} \, \mathrm{d}\lambda \\ &= \frac{m^y}{y!} \frac{\Gamma(a+s+y)}{\Gamma(a+s)} \frac{(n+b)^{a+s}}{(n+b+m)^{a+s+y}} \end{aligned}$$

[5M 2A]

(b) The scientist a priori believes that $\mathbb{E}[\lambda] = 10/3$ and $\mathbb{V}[\lambda] = 50/9$. Calculate $\text{P}[Y \leq 1 \mid \boldsymbol{x}]$ if 3 photons were detected in a sample of $n = 10$ images. ***(5 marks)***

First, solve

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

to get $a = 2$, $b = 0.6$. Now, If $m = 1, n = 10, s = 3$

$$f(y \mid \boldsymbol{x}) = \frac{1}{y!} \frac{\Gamma(a+s+y)}{\Gamma(a+s)} \frac{(n+b)^{a+s}}{(n+b+1)^{a+s+y}}$$

and

$$\begin{aligned} \text{P}[Y \leq 1 \mid \boldsymbol{x}] &= \left(\frac{n+b}{n+b+1}\right)^{a+s} \left(1 + \frac{a+s}{n+b+1}\right) \\ &= \left(\frac{10.6}{11.6}\right)^5 \left(1 + \frac{5}{11.6}\right) = 0.912\,. \end{aligned}$$

[2M 3A]

5 Consider the hierarchical model,

$$
\begin{aligned}
X_i &\sim \mathrm{Ber}(x_i \mid \theta_i)\ ,\ \text{ind. } i = 1, \ldots, n \\
\pi(\theta_i) &= \mathrm{Be}(\theta_i \mid a, a)\ ,\ \text{ind. } i = 1, \ldots, n \\
\pi(a) &= \mathrm{Ga}(a \mid c, d)\ ,\ \text{with } \mathbb{E}[a] = \frac{c}{d}\ .
\end{aligned}
$$

(i) Write down the full conditional distributions for $\boldsymbol{\theta} = \{\theta_1, \ldots, \theta_n\}$ and a.

(13 marks)

The joint posterior is

$$
\pi(\boldsymbol{\theta}, a \mid \propto) \left[\prod_{i=1}^{n} \theta_i^{x_i} (1-\theta_i)^{1-x_i} \right] \left[\prod_{i=1}^{n} \frac{1}{\mathrm{B}(a,a)} \theta_i^{a-1} (1-\theta_i)^{a-1} \right] a^{c-1} e^{-da}
$$

And the full conditionals:

θ_i

$$
\pi(\theta_i \mid -) \propto \theta_i^{x_i + a - 1} (1-\theta_i)^{(1-x_i)+a-1}\ .
$$

A Beta distribution with parameters $(x_i + a, (1 - x_i) + a)$

a

$$
\pi(a \mid -) \propto a^{c-1} e^{-da} \frac{1}{\mathrm{B}(a,a)^n} \prod_{i=1}^{n} (\theta_i (1-\theta_i))^a
$$

[5M 3A]

(ii) Write pseudo-code for a Metropolis-within-Gibbs strategy to sample from $\pi(\boldsymbol{\theta}, a \mid \boldsymbol{x})$.

(15 marks)

Start by fixing the number of MCMC samples, M, and the starting point, $\left\{\boldsymbol{\theta}^0, a^{(0)}\right\}$. Then, for $k = 1, \ldots, M$

- For $i = 1, \ldots, n$ update $\theta_i^{(k)}$ from $\mathrm{Be}(\theta \mid x_i + a^{(k-1)}, 1 - x_i + a^{(k-1)})$
- Choose a proposal density q, with support in $\mathbb{R}^+$ and keep the proposed value a^p with probability

$$
\alpha = \min\left\{1, \frac{\pi(a^p \mid \boldsymbol{\theta}^{(k)}, \boldsymbol{w})\, q(a^{(k-1)} \mid a^p)}{\pi(a^{(k-1)} \mid \boldsymbol{\theta}^{(k)}, \boldsymbol{w})\, q(a^p \mid a^{(k-1)})}\right\}
$$

[9M 3A]

6 Assume

$$X_i \sim \mathrm{N}\left(x_i \,\middle|\, \mu, \frac{1}{a_i \lambda}\right),$$

independent for $i = 1, \ldots, n$. Where $\boldsymbol{a} = \{a_1, \ldots, a_n\}$ are known constants with $0 < a_i < 1$ and $\sum_{i=1}^{n} a_i = 1$.

(i) Show that

$$\pi(\mu, \lambda) = \mathrm{N}\left(\mu \,\middle|\, m, \frac{1}{p\lambda}\right)\mathrm{Ga}(\lambda \mid a, b)$$

is a conjugate prior. ***(15 marks)***

The likelihood can be written as

$$\begin{aligned} \mathrm{L}(\mu, \lambda\,;\,\boldsymbol{x}) &\propto \lambda^{n/2} \exp\left[-\frac{\lambda}{2}\sum a_i (x_i - \mu)^2\right] \\ &\propto \lambda^{n/2} \exp\left[-\frac{\lambda}{2}(s^2 + (\mu - \widehat{\mu})^2)\right] \end{aligned}$$

where $\widehat{\mu} = \sum a_i x_i$ is the MLE and $s^2 = \sum a_i x_i^2 - (\widehat{\mu})^2$. Using Bayes theorem,

$$\begin{aligned} \pi(\mu, \lambda \mid \boldsymbol{x}) &\propto\propto \lambda^{n/2} \exp\left[-\frac{\lambda}{2}(s^2 + (\mu - \widehat{\mu})^2)\right]\lambda^{1/2}\exp[-p\lambda/2(\mu - m)^2]\lambda^{a-1}e^{-b\lambda} \\ &\propto \lambda^{1/2}\exp[-p^\star\lambda/2(\mu - m^\star)^2]\lambda^{a^\star - 1}e^{-b^\star\lambda} \end{aligned}$$

with

$$p^\star = p + 1\,, \qquad m^\star = (mp + \widehat{\mu})/p^\star\,,$$
$$a^\star = a + n/2\,, \qquad b^\star = b + (s^2 + p/p^\star(m - \widehat{\mu})^2)/2$$

[10M 5A]

(ii) Show that

$$\mathbb{E}[\mu \mid \boldsymbol{x}] = w\widehat{\mu} + (1 - w)m$$

where $0 < w < 1$ and $\widehat{\mu}$ is the MLE. ***(5 marks)***

From handouts, the marginal posterior is Student with mean $m^\star$. Thus,

$$m^\star = \widehat{\mu}\frac{1}{p+1} + m\frac{p}{p+1} \qquad \text{and} \quad w = (p+1)^{-1}$$

[2M 3A]

End of Question Paper