Data provided: Tables of distribution functions

The University Of Sheffield.

MAS6004

SCHOOL OF MATHEMATICS AND STATISTICS

Spring Semester 2017–2018

Inference

3 hours

Solutions

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.*

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) (Bookwork)

$$F_X(x) = \mathbb{P}(X \leq x) \checkmark$$

(ii) (Bookwork)

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

$$\begin{aligned}
\mathbb{E}(\hat{F}_X(x)) &= \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n} \mathbb{I}_{X_i \leq x}\right) \\
&= \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\left(\mathbb{I}_{X_i \leq x}\right) \qquad \text{(as } \mathbb{E} \text{ is a linear operator)} \checkmark \\
&= \frac{n}{n}\mathbb{P}\left(X_i \leq x\right) \checkmark \\
&= \mathbb{P}\left(X_1 \leq x\right) = F_X(x) \checkmark.
\end{aligned}$$

(iii) (Bookwork) As $n \to \infty$, $\hat{F}_X(x) \to F_X(x) \checkmark$ with probability one (mark just for saying it converges), and by the CLT, it has an approximate normal distribution $\checkmark$.

$$\widehat{F}_X(x) \sim N\left(F_X(x), \frac{F_X(x)(1 - F_X(x))}{n}\right) \checkmark$$

(iv) (Unseen) By the plug-in principle we need to find m s.t.

$$\begin{aligned}
\frac{1}{2} &\leq \int_{-\infty}^{m} \mathrm{d}\hat{F}(x) \checkmark \\
&= \int_{-\infty}^{m} \frac{1}{n}\sum \delta(X_i - x)\mathrm{d}x \checkmark \\
&= \frac{1}{n}\sum \mathbb{I}_{X_i \leq m} \checkmark
\end{aligned}$$

and

$$\begin{aligned}
\frac{1}{2} &\leq \int_{m}^{\infty} \mathrm{d}\hat{F}(x) \\
&= \frac{1}{n}\sum \mathbb{I}_{X_i \geq m}
\end{aligned}$$

So $\hat{m} = X_{(\frac{n+1}{2})}$, i.e., the midpoint/median of the dataset$\checkmark$.

 Question 1 continued on next page

1 (continued)

(v) (Unseen) For $i = 1, \ldots, B$ ✓

- Sample $X_1^{(i)}, \ldots, X_n^{(i)}$ with replacement from $\{X_1, \ldots, X_n\}$. ✓
- Set $\hat{m}^{(i)} = X^{(i)}_{(\frac{n+1}{2})}$. ✓

Calculate the standard error as

$$se(\hat{m}) = \frac{1}{B-1}\sum_{i=1}^{B}(\hat{m}^{(i)} - \bar{\hat{m}})^2 \checkmark$$

where $\bar{\hat{m}} = \frac{1}{B}\sum_{i=1}^{B}\hat{m}^{(i)}$. ✓

(vi) (Routine) A 95% CI can be found by calculating the 2.5th and 97.5th percentiles of $\hat{m}^{(1)}, \ldots, \hat{m}^{(B)}$. ✓✓

Or, as $\bar{\hat{m}} \pm 1.96 se(\hat{m})$, but only if the $\hat{m}^{(i)}$ are approximately normally distributed.

2 A precision weighing device yields unbiased measurements within half a gramme, modelled as $x \sim \mathrm{Un}(x \mid \theta - 1/2,\, \theta + 1/2)$, where θ is the unknown weight. A priori, it is believed $\theta \sim \mathrm{Un}(\theta \mid 10,\, 20)$.

(i) Find the posterior distribution of θ if a single measurement, $x = 12$, is made. *(5 marks)*

The likelihood is uniform in $x > \theta - 1/2$ and $x < \theta + 1/2$, thus constant in $\theta \in \{11.5,\, 12.5\}$. When multiplied by the prior, the posterior is constant in this same region and thus

$$\pi(\theta \mid x = 12) = 1 \;; \qquad 11.5 < \theta < 12.5 \,.$$

4M 1A

(ii) Using a new set of six measurements, $\boldsymbol{x} = \{11,\, 11.5,\, 11.7,\, 11.1,\, 11.4,\, 10.9\}$.

(a) Find the posterior distribution of θ. *(8 marks)*

The likelihood is constant in the region defined by $x_{(1)} > \theta - 1/2$ and $x_{(n)} < \theta + 1/2$ and thus constant in $\left\{x_{(n)} - 1/2,\;\; x_{(1)} + 1/2\right\} = \{11.2,\, 11.4\}$. Given that the prior is also constant in this region, the posterior is

$$\pi(\theta \mid \boldsymbol{x}) = 5 \;; \qquad 11.2 < \theta < 11.4 \,.$$

6M 2A

(b) Show that the posterior mean and variance are 11.3 and 0.27, respectively. *(2 marks)*

From the distributions handout, the posterior mean and variance are

$$\mathbb{E}[\theta \mid \boldsymbol{x}] = \frac{11.4 + 11.2}{2} = 11.3 \,,$$
$$\mathbb{V}[\theta \mid \boldsymbol{x}] = \frac{(11.4 - 11.2)^2}{12} = \frac{0.01}{3} \approx 0.003 \,.$$

1M 1A

(c) Provide an equally tailed posterior interval of probability 0.95 and explain why this is a HPD interval. *(5 marks)*

Given that the posterior is uniform, any interval of probability $\alpha \in (0,\, 1)$ is HPD. In particular, the equally tailed interval, $[a,\, b]$ is determined by

$$(a - 11.2) \times 5 = 0.025 \quad \text{and} \quad (11.4 - b) \times 5 = 0.025 \;;$$

hence, $a = 11.205$ and $b = 11.395$.

2M 3A

3 (i) (Routine) Estimates are $\hat{M} = 10419.7$ and $\hat{P}_1 = 65/1000 = 0.065$.✓✓

Confidence intervals are

$$\hat{M} \pm 1.96\sqrt{\frac{141763122}{1000}}: \ (9682,\ 11158)✓✓$$

$$\hat{P} \pm 1.96\sqrt{\frac{0.065 \times 0.935}{1000}}: \ (0.050,\ 0.080)✓✓$$

The width of the CI for M is $2 \times 1.96 \times \sqrt{\frac{141763122}{1000}}$. So to make the width less than 10 we would need

$$n = \left(2 \times 1.96 \times \frac{\sqrt{141763122}}{10}\right)^2 = 21783888.✓$$

(ii) (Unseen) Inversion sampling has been used to generate x✓, with antithetic sampling ✓used to generate negatively correlated pairs. This reduces the variance of the sample mean✓.

$$\begin{aligned} Var(\bar{c}) &= Var\left\{\frac{1}{1000}\sum_{i=1}^{1000} c_i\right\} \\ &= \frac{1}{1000^2}\{1000 \times Var(c_i) + 2 \times 500 \times Cov(c_i,\ c_{i+500})\}✓✓ \\ &= \frac{1}{1000}\{153930901 \times (1 - 0.505)\}✓ \\ &= 76083.✓ \end{aligned}$$

So 95% confidence interval is $10794 \pm 1.96\sqrt{76083}$, i.e. (10253, 11334).✓

3 (continued)

(iii) (Unseen) We want to calculate

$$\begin{aligned}
M &= \mathbb{E}c(X, Y) \\
&= \int c(x, y)\pi_X(x)\pi_Y(y)\mathrm{d}x\mathrm{d}y \\
&= \int c(x, y)\pi_X(x)\frac{\pi_Y(y)}{g(y)}g(y)\mathrm{d}x\mathrm{d}y \checkmark\checkmark \\
&\approx \frac{1}{n}\sum c(x_i, y_i)h(y_i) \checkmark
\end{aligned}$$

where $g(y)$ is the $N(10, 4)$ pdf, and

$$\begin{aligned}
h(y) &= \frac{\pi_Y(y)}{g(y)} \\
&= \frac{\frac{1}{4}ye^{-y/2}}{\frac{1}{\sqrt{8\pi}}e^{-(y-10)^2/8}}\mathbb{I}_{y>0} \\
&= \frac{\sqrt{8\pi}\,ye^{-y/2+(y-10)^2/8}}{4}\mathbb{I}_{y>0} \checkmark
\end{aligned}$$

Thus, an estimate of M is

$$\hat{M} = \frac{1}{1000}\sum_{i=1}^{1000} c_i \frac{y_i \exp(-0.5y_i + (y_i - 10)^2/8)\sqrt{8\pi}}{4}\mathbb{I}_{y_i>0} \checkmark$$

4 Assume that the waiting time, t, of a client in a bank can be modelled with an exponential distribution with unknown parameter λ,

$$f(t \mid \lambda) = \lambda \, \exp[-\lambda \, t] \, , \quad \lambda > 0 \, .$$

and that the prior distribution is Gamma with parameters (a, b):

$$\pi(\lambda) = \frac{b^a}{\Gamma[a]} \lambda^{a-1} \exp[-b \, \lambda] \, ; \quad a, \, b > 0 \, .$$

(i) Find the prior parameters if we believe $\mathbb{E}[\lambda] = 0.2$ and $\mathbb{V}[\lambda] = 1$. *(1 mark)*

Using the distributions handout, $a/b = 0.2$ and $a/b^2 = 1$ yields $a = 0.04$ and $b = 0.2$.

0.5M each

(ii) An average waiting time, $\bar{t} = 3.8$, is recorded from observing 20 clients at random. Show that the prior is conjugate and provide the posterior parameters. *(5 marks)*

The likelihood is

$$\mathrm{L}(\lambda \, ; \, \boldsymbol{x}) \propto \lambda^n \exp\left[-\lambda \sum_{i=1}^{n} t_i\right]$$

and thus the posterior is

$$\pi(\lambda \mid \boldsymbol{x}) = \mathrm{Ga}(\lambda \mid a^\star, \, b^\star)$$

with $a^\star = a + n = 20.04$ and $b^\star = b + n\bar{t} = 76.2$.

4M 1A

(iii) The coefficient of variation of a random quantity with nonzero mean, μ and standard deviation $\sigma > 0$ is defined as σ/μ. What is the smallest sample size required to reduce the posterior coefficient of variation to 0.1? *(6 marks)*

$$\frac{\sigma}{\mu} = \frac{\sqrt{a^\star / b^{\star 2}}}{a^\star / b^\star} = \frac{1}{\sqrt{a^\star}} = \frac{1}{\sqrt{a+n}} = 0.1$$

Thus, $a + n = 100$ and $n \geq 100$.

3M 3A

 Question 4 continued on next page

4 (continued)

(iv) Explain why the highest predictive probability interval of the waiting time for a randomly chosen new client is of the form $(0, c)$ and show that $c = 12.286$.

(8 marks)

The predictive distribution of the waiting time, y, of a new client is

$$f(y \mid \boldsymbol{x}) = \int_0^\infty \lambda e^{-\lambda y} \frac{b^{\star a^\star}}{\Gamma[a^\star]} \lambda^{a^\star - 1} \exp[-\lambda b^\star] \, \mathrm{d}\lambda$$
$$= \frac{a^\star}{b^\star}\left(1 + \frac{y}{b^\star}\right)^{-(a^\star+1)},$$

a decreasing function of y. Hence the predictive mode is at 0 and the upper limit, of the HPD interval is determined by

$$\int_0^c \frac{a^\star}{b^\star}\left(1 + \frac{y}{b^\star}\right)^{-(a^\star+1)} \mathrm{d}y = 0.95$$
$$1 - \left(1 + \frac{c}{b^\star}\right)^{-a^\star} = 0.95$$
$$c = b^\star\left(\left(\frac{1}{0.05}\right)^{1/a^\star} - 1\right) = 12.286 .$$

6M 2A

5 (i) (a) (routine) Method I is a two sample randomisation test.✓
Method II is a Monte Carlo hypothesis test. ✓
Null hypothesis is that group means are equal $\mu_x = \mu_y$.
Alternative is that $\mu_x \neq \mu_y$.✓

(b) (bookwork) Assumption is that subjects have been allocated to the two groups randomly.✓

(c) (routine) There are ${}^{12}C_6 = 924$ possible allocations of patients into group. ✓Smallest p-value obtained when, for the observed data, every measurement in one group is greater than every measurement in the other. So p-value in this case would be $2/924 = 0.0022$ for two-sided alternative.✓

(d) (routine) p-value for the randomisation test is 0.041, and for the Monte Carlo test it is 0.031. So in both cases we would reject H_0.✓

(e) (bookwork) Could use `mean(x)-mean(y)`.✓

 Question 5 continued on next page

5 (continued)

(ii) (a) (routine) We require

$$\int_{-k}^{k} f(x)\mathrm{d}x = 1.\checkmark$$

But we know

$$\int_{-k}^{k} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\mathrm{d}x = \Phi(k) - \Phi(-k)\checkmark$$

where $\Phi(\cdot)$ is the CDF of a standard normal random variable. Thus

$$r = \frac{1}{\Phi(k) - \Phi(-k)}.\checkmark$$

.

(b) (unseen) If we use a uniform proposal, then $g(x) = \dfrac{1}{2k}$ ✓and thus the max of $f(x)/g(x)$ occurs at $x = 0$ ✓and we find

$$M = \sup \frac{f(x)}{g(x)} = \frac{2kr}{\sqrt{2\pi}}\checkmark$$

So the rejection algorithm in this case is:

- Simulate $Y \sim U[-k, k]$ and $U \sim U[0, 1]$✓
- If $U \leq e^{-X^2/2}$ set $X = Y$. Otherwise return to step 1. ✓

  The acceptance rate of this algorithm will be

$$\frac{1}{M} = \frac{\sqrt{2\pi}}{2kr}.\checkmark$$

(c) (unseen) If we use a truncated normal as a proposal, then the acceptance rate of the rejection algorithm is simply

$$\frac{1}{r} = \Phi(k) - \Phi(-k).\checkmark$$

Thus the uniform proposal has a higher acceptance rate if

$$\frac{\sqrt{2\pi}}{2kr} > \frac{1}{r}$$

which happens if and only if

$$k < \sqrt{\frac{\pi}{2}}.\checkmark\checkmark$$

6 Consider the regression model,

$$y_i = \alpha_i + \beta x_i + \varepsilon_i \; ; \quad i = 1, \ldots, n$$

with $\varepsilon_i \sim \mathrm{N}(\varepsilon_i \,|\, 0,\, 1/\lambda)$, i.i.d. And prior structure

$$\begin{aligned} \alpha_i &\sim \mathrm{N}(\alpha_i \,|\, \mu,\, 1/p) \; ; \quad \text{independent for } i = 1, \ldots, n \\ \mu &\sim \mathrm{N}(\mu \,|\, a,\, 1/r) \\ \beta &\sim \mathrm{N}(\beta \,|\, b,\, 1/q) \end{aligned}$$

and

$$\lambda \sim \mathrm{Ga}(\lambda \,|\, c,\, d)$$

(i) Show that the full conditional of:

(a) Each of the individual intercepts, α_i, is Gaussian and provide explicit expressions for the parameters. *(3 marks)*

$$\begin{aligned} \pi(\alpha_i \,|\, -) &\propto \exp\left[-\frac{\lambda}{2}(y_i - \alpha_i - \beta x_i)^2\right] \exp\left[-\frac{p}{2}(\alpha_i - \mu)^2\right] \\ &\propto \exp\left[-\frac{p^\star}{2}(\alpha_i - a^\star)^2\right] \end{aligned}$$

with $p^\star = p + \lambda$ and $a^\star = (\lambda(y_i - \beta x_i) + p\mu)/p^\star$.

1M 1A

(b) The mean intercept, μ, is Gaussian and provide explicit expressions for the parameters. *(3 marks)*

$$\begin{aligned} \pi(\mu \,|\, -) &\propto \exp\left[-\frac{p}{2}\sum_{i=1}^{n}(\alpha_i - \mu)^2\right] \exp\left[-\frac{r}{2}(\mu - a)^2\right] \\ &\propto \exp\left[-\frac{r^\star}{2}(\mu - m^\star)^2\right] \end{aligned}$$

with $r^\star = r + np$ and $m^\star = (np\bar{\alpha} + ra)/r^\star$.

2M 1A

 Question 6 continued on next page

6 (continued)

(c) The regression slope, β, is Gaussian and provide explicit expressions for the parameters. *(3 marks)*

$$\begin{aligned}\pi(\beta \mid -) &\propto \exp\left[-\frac{\lambda}{2}\sum_{i=1}^{n}(y_i - \alpha_i - \beta x_i)^2\right] \exp\left[-\frac{q}{2}(\beta - b)^2\right] \\ &\propto \exp\left[-\frac{q^\star}{2}(\beta - b^\star)^2\right]\end{aligned}$$

with $q^\star = \lambda \sum x_i^2 + q$ and $b^\star = \left(\lambda \sum x_i(y_i - \alpha_i) + qb\right)/q^\star$.

2M 1A

(d) The regression precision, λ, is Gamma and provide explicit expressions for the parameters. *(3 marks)*

$$\begin{aligned}\pi(\lambda \mid -) &\propto \lambda^{n/2} \exp\left[-\frac{\lambda}{2}\sum_{i=1}^{n}(y_i - \alpha_i - \beta x_i)^2\right] \lambda c - 1 e^{-d\lambda} \\ &\propto \lambda^{c^\star - 1} \exp[-d^\star \lambda]\end{aligned}$$

with $c^\star = c + n/2$ and $d^\star = d + \frac{1}{2}\sum(y_i - \alpha_i - \beta x_i)^2$.

2M 1A

(ii) Write pseudo-code for an MCMC sampling scheme for exploring the posterior distribution. *(8 marks)*

6(continued)

We can setup a Gibbs sampler, cycling trough the full conditionals. First, select the length of the chain, M. Then

procedure GIBBS SAMPLER
  Set $\left\{\boldsymbol{\alpha}^{(0)}, \mu^{(0)}, \beta^{(0)}, \lambda^{(0)}\right\}$
  for $j = 1, \ldots, M$ **do**
    Sample $\lambda^{(j)}$ from $\mathrm{Ga}(\cdot \,|\, a, b)$, with

$$a = c + n/2 \quad \text{and} \quad b = d + \frac{1}{2}\sum\left(y_i - \alpha_i^{(j-1)} - \beta^{(j-1)}x_i\right)^2$$

    Sample $\beta^{(j)}$ from $\mathrm{N}(\cdot \,|\, \nu, 1/\tau)$, with

$$\tau = \lambda^{(j-1)}\sum x_i^2 + q \quad \text{and} \quad \nu = \left(\lambda^{(j-1)}\sum x_i\left(y_i - \alpha_i^{(j-1)}\right) + qb\right)/\tau$$

    Sample $\mu^{(j)}$ from $\mathrm{N}(\cdot \,|\, \nu, 1/\tau)$, with

$$\tau = r + np \quad \text{and} \quad \nu = (np\bar{\alpha}^{(j-1)} + ra)/\tau$$

    for $i = 1, \ldots, n$ **do**
      Draw $\alpha_i^{(j)}$ from $\mathrm{N}(\cdot \,|\, \nu, 1/\tau)$, with

$$\tau = p + \lambda^{(j)} \quad \text{and} \quad \nu = \left(\lambda^{(j)}(y_i - \beta^{(j)}x_i) + p\mu^{(j)}\right)/\tau$$

    end for
    Record $\left\{\boldsymbol{\alpha}^{(j)}, \nu^{(j)}, \beta^{(j)}, \lambda^{(j)}\right\}$
  end for
end procedure

Award 2M for each correct parameter/step. Deduct 1M if the updated state of the chain has not been taken considered explicitly.
Award full marks if MH is used instead of Gibbs, iff an appropriate proposal has been put forward and the acceptance rate included explicitly.

End of Question Paper