Data provided: Tables of distribution functions.

MAS6004

The University Of Sheffield.

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

MAS6004 Inference **3 hours**

Candidates may bring to the examination a calculator that 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 Consider the following hierarchical model

$$x_i \sim \mathrm{N}(x_i \,|\, \mu_i, 1/\lambda)\ , \quad \text{independent for } i = 1, \ldots, n$$
$$\pi\,(\mu_i) = \mathrm{N}(\mu_i \,|\, \theta, 1/\gamma)\ , \quad \text{independent for } i = 1, \ldots, n$$
$$\pi\,(\lambda) = \mathrm{Ga}(\lambda \,|\, a, b)\ , \quad \pi\,(\theta) = \mathrm{N}(\theta \,|\, m, 1/p) \quad \text{and} \quad \pi\,(\gamma) = \mathrm{Ga}(\gamma \,|\, c, d)\ ,$$

with $\{m, p, a, b, c, d\}$ known constants.

[HINT: The quadratic form $ax^2 + bx + c$ can be written as $a(x + b/(2a))^2 + d$, with d not depending on x.]

(i) Prove that the full conditionals for

(a) μ_i are $\mathrm{N}\big(\mu_i \,|\, m_i^\star, 1/p^\star\big)$ with $p^\star = \lambda + \gamma$ and $m_i^\star = (\lambda x_i + \gamma\theta)/p^\star$.. ***(3 marks)***

(b) θ is $\mathrm{N}(\theta \,|\, q^\star, 1/v^\star)$ with $v^\star = n\gamma + p$ and $q^\star = (n\gamma\bar{\mu} + pm)/v^\star$. ***(3 marks)***

(c) λ is $\mathrm{Ga}(\lambda \,|\, a^\star, b^\star)$ with $a^\star = a + n/2$ and $b^\star = b + \dfrac{1}{2}\displaystyle\sum_{i=1}^{n}(x_i - \mu_i)^2$. ***(2 marks)***

(d) γ is $\mathrm{Ga}(\gamma \,|\, c^\star, d^\star)$ with $c^\star = c + n/2$ and $d^\star = d + \dfrac{1}{2}\displaystyle\sum_{i=1}^{n}(\mu_i - \theta)^2$. ***(2 marks)***

(ii) Write pseudo-code of a Gibbs sampler for exploring the posterior of the parameters for this model. ***(10 marks)***

2 A physicist studying the expansion of the universe has two sets of measurements covering the same section of the Milky Way. The difference between these measurements is related to redshift and, if current theory is correct, expected to be very close to zero.

(i) Let $\boldsymbol{d} = \{d_1, \ldots, d_n\}$ be the data available with d_i the i-th observed difference. Assume these are conditionally independent $d_i \sim \mathrm{N}(d_i \mid \mu, 1/\lambda)$, with known precision $\lambda = 0.1$. Derive the posterior distribution of μ using the conjugate prior,

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

and give explicit expressions for its parameters. ***(8 marks)***

(ii) Given the data, the scientist may report a discrepancy (call this decision a_1) or an agreement (decision a_2) with the current theory. If a real discrepancy is reported there is a good chance it would be published in a top journal; if the discrepancy is not real, his career would suffer a major drawback. After some consideration, he believes that the following loss function reflects well his preferences:

$$\mathrm{L}(a_1, \mu) = \begin{cases} 100 & |\mu| \le 0.5 \\ 0 & |\mu| > 0.5 \end{cases}, \qquad \mathrm{L}(a_2, \mu) = \begin{cases} 0 & |\mu| \le 0.5 \\ 30 & |\mu| > 0.5 \end{cases}.$$

A set of $n = 150$ measures is taken and the following statistics are recorded:

$$\bar{d} = \frac{1}{n}\sum_{i=1}^{n} d_i = -0.3\,, \quad s_d^2 = \frac{1}{n}\sum_{i=1}^{n}\left(d_i - \bar{d}\right)^2 = 10\,.$$

After elicitation, the scientist's prior parameters are $m = 0$, $p = 0.5$.

(a) Prove that the scientist should report a discrepancy if and only if

$$\frac{100}{130} < \mathrm{P}[|\mu| \le 0.5 \mid \boldsymbol{d}]$$

(5 marks)

(b) Which is the scientist's optimal decision?
[Additional information, if Z has a standard Gaussian distribution, $\mathrm{P}[Z \le 0.826] = 0.795$ and $\mathrm{P}[Z \le 3.112] = 0.999$.] ***(7 marks)***

3 Let x_i be the number of complaints filed to a consumer agency in a given day, and let $\mathbf{x} = \{x_1, \ldots, x_n\}$ be a random sample obtained from the agency's records. Assuming that $\mathrm{Po}(x_i \mid \lambda)$ is a suitable model:

(i) (a) Prove that the posterior from the non-informative (improper) prior $\pi(\lambda) \propto \lambda^{-1}$ is a Gamma distribution and write down the posterior parameters explicitly. ***(3 marks)***

(b) Prove that $\pi(\lambda) = \mathrm{Ga}(\lambda \mid a, b)$ is a conjugate prior and write down the posterior parameters explicitly. ***(2 marks)***

(ii) It is further assumed that the distribution of the x_i arise from a process where the distribution of the waiting time to the next complaint, t, is exponential with the same rate parameter λ.

(a) Using your results in (i) (a), prove that the predictive distribution of t is

$$f(t \mid \boldsymbol{x}) = n^s s (n+t)^{-(s+1)}, \quad \text{where} \quad s = \sum_{i=1}^{n} x_i$$

(5 marks)

(b) The agency's manager claims that it is more likely that the next complaint will be filed before midday than not; *i.e.* $t \leq 1/2$. Does the sample $\boldsymbol{x} = \{0, 2, 1, 4, 3, 4, 3, 0, 2, 1\}$ provide evidence to support this statement? ***(5 marks)***

(iii) (a) Using your results in (i)(b), prove that the predictive distribution of y, the number of complaints filed in next day is

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

with $\{a^\star, b^\star\}$ the parameters of the posterior distribution. ***(5 marks)***

4 (i) Let X be a random variable with the triangular density function

$$f_X(x) = \begin{cases} x & \text{for } 0 \le x \le 1, \\ 2 - x & \text{for } 1 < x \le 2, \\ 0 & \text{otherwise.} \end{cases}$$

(a) Find the cumulative distribution function of X. ***(3 marks)***

(b) Using your result from part a), explain how to generate a random value of X given a uniform random number using the inversion method. ***(4 marks)***

(ii) Suppose it is desired to generate a random variable X where X has the following density function:

$$f_X(x) = \begin{cases} \frac{3}{2}(1 - x^2) & \text{if } 0 \le x \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

(a) Explain carefully how to generate a random value of X using rejection sampling, with a uniform envelope density function. If the first candidate value is $Y = 0.5$, given the value $U = 0.8$ from the $U[0, 1]$ distribution, determine whether 0.5 is accepted as a random sample from the distribution of X. ***(8 marks)***

(b) Consider the alternative envelope density function

$$g(y) = \begin{cases} 2(1 - y) & \text{for } 0 \le y \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

Derive the expected number of candidate random variables Y required to obtain a single random X with this new envelope function. ***(5 marks)***

[**Hint:** $(1 - x^2) = (1 - x)(1 + x)$.]

5 A sample of independent random observations $X = \{X_1, \ldots, X_n\}$ are drawn from the following mixture distribution:

$$X_i \sim \begin{cases} Poisson(\lambda) & \text{with probability } 1-\omega, \\ Poisson(\mu) & \text{with probability } \omega. \end{cases}$$

Let $\hat{\lambda}$, $\hat{\mu}$ and $\hat{\omega}$ denote the maximum likelihood estimates of λ, μ and ω given the observed data X only. The corresponding 'missing' variables $Y = \{Y_1, \ldots, Y_n\}$ are defined as follows:

$$Y_i = \begin{cases} 0 & \text{if } X_i \text{ is drawn from } Poisson(\lambda), \\ 1 & \text{if } X_i \text{ is drawn from } Poisson(\mu). \end{cases}$$

Define $\theta = (\omega, \mu, \lambda)$, with $\hat{\theta} = (\hat{\omega}, \hat{\mu}, \hat{\lambda})$.

(i) Using the fact that $P(X, Y|\theta) = P(Y|\theta)P(X|Y, \theta)$ for random variables X and Y, show that the log-likelihood

$$l(X, Y|\mu, \lambda, \omega) = \sum_{i=1}^{n} \Big[(1 - Y_i)\{\log(1-\omega) + X_i \log\lambda - \log X_i! - \lambda\} + Y_i\{\log(\omega) + X_i \log\mu - \log X_i! - \mu\} \Big].$$

(5 marks)

(ii) Show that the maximum likelihood estimates of ω, μ, λ given both X and Y are

$$\hat{\omega} = \frac{\sum_{i=1}^{n} Y_i}{n}, \quad \hat{\lambda} = \frac{\sum_{i=1}^{n} X_i(1-Y_i)}{\sum_{i=1}^{n}(1-Y_i)}, \quad \hat{\mu} = \frac{\sum_{i=1}^{n} X_i Y_i}{\sum_{i=1}^{n} Y_i}.$$

(4 marks)

(iii) Using Bayes' theorem, derive an expression for $P(Y_i = 1|X_i, \theta)$. ***(4 marks)***

 Question 5 continued on next page

5 (continued)

(iv) Given X only, suppose the EM algorithm is to be used to obtain the maximum likelihood estimate $\hat{\theta}$ of θ. Let the current estimate of $\hat{\theta}$ be θ_{old}. By maximising

$$Q(\theta|\theta_{old}) = E[l(\theta; X, Y)|X, \theta = \theta_{old}],$$

show that the updated estimates of $\hat{\omega}, \hat{\mu}, \hat{\lambda}$ are

$$\begin{aligned} \omega_{new} &= \frac{\sum_{i=1}^{n} p_i}{n}, \\ \lambda_{new} &= \frac{\sum_{i=1}^{n} X_i(1-p_i)}{\sum_{i=1}^{n}(1-p_i)}, \\ \mu_{new} &= \frac{\sum_{i=1}^{n} X_i p_i}{\sum_{i=1}^{n} p_i}, \end{aligned}$$

where p_i is your expression for $P(Y_i = 1|X_i, \theta = \theta_{old})$ in part (iv).

(7 marks)

6 (i) A set of $n + m$ patients in a hospital with bowel disease were given a new drug and the *remission time* (the time until their symptoms disappeared) was observed in days. By the end of the trial not all of the patients had recovered, with some individuals still showing symptoms at the end of the study. The first n individuals were observed to fully recover with remission times $x_1, \ldots, x_n$. Patients $n+1, \ldots, n+m$ were still showing symptoms, with the i^{th} individual having been observed for c_i days, for $i = n+1, \ldots, n+m$.

The remission time X is modelled by a $Weibull(\lambda, k)$ distribution with density

$$f(x) = \begin{cases} \frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1} e^{-(x/\lambda)^k} & x \geq 0, \\ 0 & x < 0. \end{cases}$$

(a) Show that, for an individual who is still showing symptoms after c_i days,

$$P(X > c_i) = e^{-(c_i/\lambda)^k} \quad c_i \geq 0.$$

(2 marks)

(b) Let $\mathbf{y} = (x_1, \ldots, x_n, c_{n+1}, \ldots, c_{n+m})$. Show that the log-likelihood for λ, k is given by

$$l(\lambda, k; \mathbf{y}) = n \log k - nk \log \lambda + (k-1)\sum_{i=1}^{m} \log x_i - \frac{1}{\lambda^k}\left(\sum_{i=1}^{n} x_i^k + \sum_{i=n+1}^{n+m} c_i^k\right).$$

(5 marks)

(c) Find the profile log-likelihood for k i.e.

$$l_p(k; \mathbf{y}) = \max_{\lambda} l(\lambda, k; \mathbf{y}).$$

(6 marks)

(ii) Suppose importance sampling, using a normal approximation as the importance density, is to be used to sample from a $Weibull(4, 6)$ distribution

$$f(x) = \begin{cases} \frac{6}{4}\left(\frac{x}{4}\right)^{5} e^{-(x/4)^6} & x \geq 0, \\ 0 & x < 0. \end{cases}$$

By considering a Taylor series expansion of $\log f(x)$ about the mode of x, obtain good candidates for the mean and variance of the importance density.

(7 marks)

Note: The Taylor series expansion of a function $h(x)$ is

$$h(x) = h(a) + \frac{h'(a)}{1!}(x-a) + \frac{h''(a)}{2!}(x-a)^2 + \cdots .$$

to second order]

End of Question Paper

SOME DISCRETE DISTRIBUTIONS

Name	**Context**	**Notation**	**p.f.** $p(x \mid \theta)$	$\mathrm{E}[X \mid \theta]$	$\mathrm{Var}[X \mid \theta]$	**Applications**	**Comments**
Uniform (discrete)	Set of k equally likely outcomes (usually, not necessarily, the integers)	$\mathrm{U}(1, \ldots, k)$	$p(x) = 1/k$ $\mathcal{X} = \{1, \ldots, k\}$	$\frac{k+1}{2}$	$\frac{k^2-1}{12}$	Dice	
Bernoulli trial	Expt. with two outcomes: 'success' w.p. θ and 'failure' w.p. $1-\theta$ $X \equiv$ no. successes	$\mathrm{Ber}(x \mid \theta)$	$p(x) = \theta^x(1-\theta)^{1-x}$ $\mathcal{X} = \{0, 1\}$ $\Theta = (0, 1)$	θ	$\theta(1-\theta)$	Coins, constituent of more complex distributions	
Binomial	$X \equiv$ no. successes in n ind. $\mathrm{Ber}(x \mid \theta)$ trials	$\mathrm{Bi}(x \mid n, \theta)$	$p(x) = \binom{n}{x}\theta^x(1-\theta)^{n-x}$ $\mathcal{X} = \{0, 1, 2, \ldots, n\}$ $\Theta = (0, 1)$	$n\theta$	$n\theta(1-\theta)$	Sampling with replacement	$\mathrm{Bi}(x \mid 1, \theta) \equiv \mathrm{Ber}(x \mid \theta)$
Geometric	$X \equiv$ no. failures until 1st success in sequence of ind. $\mathrm{Ber}(x \mid \theta)$ trials	$\mathrm{Ge}(x \mid \theta)$	$p(x) = \theta(1-\theta)^x$ $\mathcal{X} = 0, 1, 2, \ldots$ $\Theta = (0, 1)$	$\frac{1-\theta}{\theta}$	$\frac{1-\theta}{\theta^2}$	Waiting times (for single events)	Alternative formulation in terms of $Y \equiv$ no. of trials to 1st success ($Y = X+1$)
Negative binomial (or Pascal)	$X \equiv$ no. failures to m-th success in sequence of ind. $\mathrm{Ber}(x \mid \theta)$ trials. Generalisation of Geometric	$\mathrm{NB}(x \mid m, \theta)$	$p(x) = \binom{m+x-1}{x}\theta^m(1-\theta)^x$ $\mathcal{X} = 0, 1, 2, \ldots$ $\Theta = (0, 1)$	$\frac{m(1-\theta)}{\theta}$	$\frac{m(1-\theta)}{\theta^2}$	Waiting times (for compound events)	$\mathrm{NB}(x \mid 1, \theta) \equiv \mathrm{Ge}(x \mid \theta)$
Poisson	Arises empirically or via Poisson Process (PP) for counting events. For PP rate ν the no. of events in time $t \sim \mathrm{Po}(x \mid \nu t)$. Also as an approx. to the Binomial	$\mathrm{Po}(x \mid \lambda)$	$p(x) = \frac{e^{-\lambda}\lambda^x}{x!}$ $\mathcal{X} = 0, 1, 2, \ldots$ $\Lambda = \mathbb{R}^+$	λ	λ	Counting events occurring 'at random' in space or time	$\mathrm{Bi}(x \mid n, \theta) \equiv \mathrm{Po}(x \mid n\theta)$ if n large, θ small

SOME CONTINUOUS DISTRIBUTIONS

Name	Notation	p.d.f. $f(x \mid \theta)$	$\mathrm{E}[X \mid \theta]$	$\mathrm{Var}[X \mid \theta]$	Applications	Comments
Uniform (continuous)	$\mathrm{Un}(x \mid \alpha, \beta)$	$f(x) = \frac{1}{\beta - \alpha}$ $\mathcal{X} = [\alpha, \beta]$ $\Theta = \{(\alpha, \beta) \in \mathbb{R}^2 : \alpha < \beta\}$	$\frac{\alpha + \beta}{2}$	$\frac{(\beta - \alpha)^2}{12}$	Rounding errors $\mathrm{Un}(x \mid -1/2, 1/2)$. Simulating other distributions from $\mathrm{Un}(x \mid 0, 1)$	
Exponential	$\mathrm{Ex}(x \mid \lambda)$	$f(x) = \lambda e^{-\lambda x}$ $\mathcal{X} = \mathbb{R}_+$ $\Lambda = \mathbb{R}_+$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	Inter-event times for Poisson Process. Models lifetimes of non-ageing items.	Also parameterised in terms of $1/\lambda$. $\mathrm{Ga}(x \mid 1, \lambda) \equiv \mathrm{Ex}(x \mid \lambda)$
Gamma	$\mathrm{Ga}(x \mid \alpha, \beta)$	$f(x) = \frac{\beta^\alpha x^{\alpha-1} e^{-\beta x}}{\Gamma[\alpha]}$ $\mathcal{X} = \mathbb{R}_+$ $\Theta = \{(\alpha, \beta) \in \mathbb{R}^2 : \alpha > 0, \beta > 0\}$	$\frac{\alpha}{\beta}$	$\frac{\alpha}{\beta^2}$	Times between k events for Poisson Process. Lifetimes of ageing items.	Also parameterised in terms of $1/\beta$ $\mathrm{Ga}(x \mid 1, \lambda) \equiv \mathrm{Ex}(x \mid \lambda)$, $\mathrm{Ga}(x \mid \nu/2, 1/2) \equiv \chi^2_{(\nu)}(x)$
Beta	$\mathrm{Be}(x \mid \alpha, \beta)$	$f(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{\mathrm{B}(\alpha, \beta)}$ $\mathcal{X} = (0, 1)$ $\Theta = \{(\alpha, \beta) \in \mathbb{R}^2 : \alpha > 0, \beta > 0\}$	$\frac{\alpha}{\alpha + \beta}$	$\frac{\alpha\beta(\alpha + \beta)^{-2}}{(\alpha + \beta + 1)}$	Useful model for variables with finite range. Also as a Bayesian conjugate prior.	$\mathrm{Be}(x \mid 1, 1) \equiv \mathrm{Un}(x \mid 0, 1)$ $\mathrm{Be}(x \mid \alpha, \beta)$ is reflection about $\frac{1}{2}$ of $\mathrm{Be}(x \mid \beta, \alpha)$. Can transform $\mathrm{Be}(x \mid \alpha, \beta)$ on $[0, 1]$ to any finite range $[a, b]$ by $Y = (b - a)X + a$
Normal (Gaussian)	$\mathrm{N}(x \mid \mu, \sigma^2)$	$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x - \mu}{\sigma}\right)^2\right]$ $\mathcal{X} = \mathbb{R}$ $\Theta = \{(\mu, \sigma^2) \in \mathbb{R}^2 : \sigma^2 > 0\}$	μ	σ^2	Empirically and theoretically (via CLT etc.) a good model in many situations. Often parameterised in terms of the precision $\lambda = 1/\sigma^2$	$Y = aX + b \sim \mathrm{N}(y \mid a\mu + b, a^2\sigma^2)$ $Z = \frac{X-\mu}{\sigma} \sim \mathrm{N}(z \mid 0, 1)$ $\mathrm{P}[X \in (u, v)] = \mathrm{P}\left[Z \in \left(\frac{u - \mu}{\sigma}, \frac{v - \mu}{\sigma}\right)\right]$
Chi-square	$\chi^2_{(\nu)}(x)$	$f(x) = 2^{-\nu/2}\Gamma(\nu/2)^{-1} x^{\nu/2-1} e^{-x/2}$ $\mathcal{X} = \mathbb{R}_+$; $\Theta = \mathbb{R}_+$	ν	2ν	Sum of squares of ν standard normals	$\chi^2_{(\nu)}(x) \equiv \mathrm{Ga}(x \mid \nu/2, 1/2)$
Student t	$\mathrm{St}(x \mid \mu, \lambda, \nu)$	$f(x) = \Gamma[(\nu+1)/2]/\Gamma[\nu/2]\left(\frac{\lambda}{\nu\pi}\right)^{1/2}$ $\left(1 + \lambda(x - \mu)^2/\nu\right)^{-(\nu+1)/2}$ $\mathcal{X} = \mathbb{R}$ $\Theta = \mathbb{R}_+$	μ (if $\nu > 1$)	$\lambda^{-1}\frac{\nu}{\nu - 2}$ (if $\nu > 2$)	Useful alternative to Normal for variables with heavy tails.	If $X \sim \mathrm{N}(x \mid 0, 1)$ and $Y \sim \chi^2_{(\nu)}(y)$ independent then $\frac{X}{\sqrt{Y/\nu}} \sim \mathrm{t}_\nu$. $\mathrm{t}_1 \equiv$ Cauchy. $\mathrm{t}^2_\nu \equiv \mathrm{F}_{1,\nu}$. If $Y = \sqrt{\lambda}(x - \mu)$ then $Y \sim t_1(y)$