DEPARTMENT OF MATHEMATICAL SCIENCES
Outline solution to Examination Questions

UNIVERSITY OF BATH

Unit Code MA40189	Unit Title TOPICS IN BAYESIAN STATISTICS
Academic Year 2019/20	Examiner SIMON SHAW
Semester II	Question No. NOTES ON QUESTIONS — Page 1 of 2

Part | Mark

1 (a) Bookwork.

(b) Bookwork.

(c) Routine problem.

(d) (i). Routine problem but students will have to take care with the constants for the kernel in the integrand.

(ii) Students have seen similar ideas before but this should begin to stretch them.

(e) Unseen, genuine first class level material. They haven't really seen this four parameter prior and will have to be on top of their understanding to spot the conjugacy. Worth 4 marks (20%).

A student who has engaged with the course should be able to obtain 8 marks (40%) from at least parts (a), (b), and (c).

2. (a) (i) Routine problem.

(ii) Routine discussion: they've seen many examples of this type.

(iii) Routine, familiar, problem. Good students will benefit from knowing the Gamma, Inv-gamma relationship.

(iv) Familiar problem which they've seen examples of but should start to stretch them as they need to take care with the proportionality constants.

(b) (i) Bookwork.

(ii) Unseen, genuine first class level material. Worth 4 marks (20%).

A student who has engaged with the course should be able to obtain 8 marks (40%) from at least parts (a)(i), (a)(ii) and (b)(i).

Total

DEPARTMENT OF MATHEMATICAL SCIENCES
Outline solution to Examination Questions

UNIVERSITY OF BATH

Unit Code MA40189	Unit Title TOPICS IN BAYESIAN STATISTICS		
Academic Year 2019/20	Examiner SIMON SHAW		
Semester II	Question No. NOTES ON QUESTIONS	Page 2	of 2

Part | Mark

3 (a) Routine problem.
(b) Routine problem though will need to keep on top of notation.
(c). Bookwork.
(d). Bookwork though importance sampling has infrequently appeared on past papers so should be less familiar.
(e) Unseen, genuine first class level material. Worth 4 marks (20%).

A student who has engaged with the course should be able to obtain (at least) 8 marks from parts (a), (b), and (c) and so score at least 40%.

4. (a) Routine derivation.
(b) Routine problem: they will save time (and should) by using (a).
(c) Routine problem. In these parts, students will have to take a little care with the gamma functions.
(d) Stretching of a routine problem. The risk of the sampling procedure (assuming (b), (c) correct) is straightforward for this example.
(e) Unseen, genuine first class level material. Should test understanding of decision theory away from simple point estimation. For the careful student, the differentiation should be straightforward provided they are thinking. Worth 4 marks (20%).

A student who has engaged with the course should be able to obtain (at least) 8 marks (40%) from parts (a), (b), and (c).

Total

Unit Code MA40189	Unit Title TOPICS IN BAYESIAN STATISTICS	
Academic Year 2019/20	Examiner SIMON SHAW	
Semester II	Question No. 1	Page 1 of 4

Part | Mark

(a) A class Π of prior distributions is said to form a conjugate family with respect to a likelihood $f(x|\theta)$ if the posterior density is in the class of Π for all x whenever the prior density is in Π. 2

(b) If $f(x|\theta)$ belongs to k-parameter regular exponential family then

$$f(x|\theta) = \exp\left\{\sum_{j=1}^{k} \phi_j(\theta)u_j(x) + g(\theta) + h(x)\right\}$$

Take as our prior the following $(k+1)$-parameter exponential family over θ,

$$f(\theta) = \exp\left\{\sum_{j=1}^{k} a_j\, \phi_j(\theta) + d\, g(\theta) + c(a,d)\right\}$$

where $a = (a_1, \ldots, a_k)$ and $c(a,d)$ is the normalising constant,

$$c(a,d) = -\log \int_{\theta} \exp\left\{\sum_{j=1}^{k} a_j\, \phi_j(\theta) + d\, g(\theta)\right\} d\theta$$

Then, $f(\theta|x) \propto f(x|\theta)f(\theta)$

$$\propto \exp\left\{\sum_{j=1}^{k} [a_j + u_j(x)]\, \phi_j(\theta) + (d+1)g(\theta)\right\}$$

which, up to a constant of proportionality, has the same form as the prior giving conjugacy. 3

Total

Unit Code MA40189	Unit Title TOPICS IN BAYESIAN STATISTICS	
Academic Year 2019/20	Examiner SIMON SHAW	
Semester II	Question No. 1	Page 2 of 4

Part | Mark

(c) $$f(x \mid \theta) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi v}} \exp\left\{-\frac{1}{2v}(x_i-\mu)^2\right\}$$
$$= (2\pi)^{-\frac{n}{2}} v^{-\frac{n}{2}} \exp\left\{-\frac{1}{2v}\sum_{i=1}^{n}(x_i-\mu)^2\right\}$$

Now, $$\sum_{i=1}^{n}(x_i-\mu)^2 = \sum_{i=1}^{n}\left((x_i-\bar{x})+(\bar{x}-\mu)\right)^2$$
$$= \sum_{i=1}^{n}(x_i-\bar{x})^2 + n(\bar{x}-\mu)^2 = (n-1)s^2 + n(\bar{x}-\mu)^2$$

Thus, $$f(x \mid \theta) = (2\pi)^{-\frac{n}{2}} v^{-\frac{n}{2}} \exp\left\{-\frac{1}{2v}\left[(n-1)s^2+n(\bar{x}-\mu)^2\right]\right\}$$
$$\propto v^{-\frac{n}{2}} \exp\left\{-\frac{1}{2v}\left[(n-1)s^2+n(\bar{x}-\mu)^2\right]\right\}$$
$$= g(\theta, \bar{x}, s^2).$$

Hence, as $f(\theta \mid x) \propto f(x \mid \theta) f(\theta) \propto g(\theta, \bar{x}, s^2) f(\theta)$, $\bar{X}$ and S^2 are sufficient for X for learning about θ. 3

(d) (i) $$f(\mu, v \mid x) \propto v^{-\frac{n}{2}} \exp\left\{-\frac{1}{2v}\left[(n-1)s^2+n(\bar{x}-\mu)^2\right]\right\} v^{-1}$$
$$= v^{-\left(\frac{n}{2}+1\right)} \exp\left\{-\frac{1}{2v}\left[(n-1)s^2+n(\bar{x}-\mu)^2\right]\right\}$$

Total

Unit Code MA40189	Unit Title TOPICS IN BAYESIAN STATISTICS	
Academic Year 2019/20	Examiner SIMON SHAW	
Semester II	Question No. 1	Page 3 of 4

Part | Mark

Thus, $f(v|x) \propto \int_{-\infty}^{\infty} v^{-(\frac{n}{2}+1)} \exp\left\{-\frac{1}{2v}\left[(n-1)s^2 + n(\bar{x}-\mu)^2\right]\right\} d\mu$

$$= v^{-(\frac{n}{2}+1)} \exp\left\{-\frac{1}{2v}(n-1)s^2\right\} \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2v} n(\bar{x}-\mu)^2\right\} d\mu$$

We recognise the integrand as a kernel of $N(\bar{x}, v/n)$ so that

$$f(v|x) \propto v^{-(\frac{n}{2}+1)} \exp\left\{-\frac{1}{2v}(n-1)s^2\right\} v^{\frac{1}{2}}$$

$$= v^{-(\frac{n}{2}-\frac{1}{2}+1)} \exp\left\{-\frac{1}{2v}(n-1)s^2\right\}$$ 4

We recognise this as a kernel of Inv-gamma$\left(\frac{n-1}{2}, \frac{(n-1)s^2}{2}\right)$ and thus this is the distribution of $v|x$.

(ii) Note that if $Z \sim$ Inv-gamma(α, β) then $Z^{-1} \sim$ Gamma(α, β).

Hence, $v^{-1}|x \sim$ Gamma$\left(\frac{n-1}{2}, \frac{(n-1)s^2}{2}\right)$ and, from the hint,

$(n-1)s^2 v^{-1}|x \sim$ Gamma$\left(\frac{n-1}{2}, \frac{1}{2}\right)$ which is the chi-squared distribution with $(n-1)$ degrees of freedom. Hence, $Y|x \sim \chi^2_{n-1}$.

This corresponds to the pivot $\frac{(n-1)S^2}{v}$ used to construct confidence intervals in the classical model. Thus, classical and Bayesian inference will coincide showing that the prior is noninformative. 4

Total

Unit Code MA40189	Unit Title TOPICS IN BAYESIAN STATISTICS	
Academic Year 2019/20	Examiner SIMON SHAW	
Semester II	Question No. 1	Page 4 of 4

Part | Mark

(e)
$$f(\mu,\nu) = f(\mu|\nu)f(\nu)$$
$$\propto \nu^{-\frac{1}{2}}\exp\left\{-\frac{\lambda}{2\nu}(\mu-\mu_0)^2\right\}\nu^{-(\alpha+1)}\exp\left\{-\frac{\beta}{\nu}\right\}$$
$$= \nu^{-(\alpha+\frac{3}{2})}\exp\left\{-\frac{1}{2\nu}\left[\lambda(\mu-\mu_0)^2+2\beta\right]\right\}$$

$$f(\mu,\nu|x) \propto \nu^{-(\alpha+\frac{n}{2}+\frac{3}{2})}\exp\left\{-\frac{1}{2\nu}\left[\lambda(\mu-\mu_0)^2+n(\mu-\bar{x})^2+2\beta+(n-1)s^2\right]\right\}$$

Now, from the hint, $\lambda(\mu-\mu_0)^2+n(\mu-\bar{x})^2 = (\lambda+n)\left(\mu-\frac{\lambda\mu_0+n\bar{x}}{\lambda+n}\right)^2 + \frac{n\lambda}{\lambda+n}(\mu_0-\bar{x})^2$

so that
$$f(\mu,\nu|x) \propto \nu^{-(\alpha+\frac{n}{2}+\frac{3}{2})}\exp\left\{-\frac{1}{2\nu}\left[(\lambda+n)\left(\mu-\frac{\lambda\mu_0+n\bar{x}}{\lambda+n}\right)^2 + 2\left(\beta + \frac{(n-1)s^2}{2} + \frac{n\lambda}{2(\lambda+n)}(\mu_0-\bar{x})^2\right)\right]\right\}$$

which is of the same form as the prior giving conjugacy.

[In terms of the hyperparameters, $\alpha \mapsto \alpha+\frac{n}{2}$, $\lambda \mapsto \lambda+n$,
$\mu_0 \mapsto \frac{\lambda\mu_0+n\bar{x}}{\lambda+n}$, $\beta \mapsto \beta + \frac{(n-1)s^2}{2} + \frac{n\lambda}{2(\lambda+n)}(\mu_0-\bar{x})^2$.] 4

Total 20

Unit Code MA40189	Unit Title TOPICS IN BAYESIAN STATISTICS
Academic Year 2019/20	Examiner SIMON SHAW
Semester II	Question No. 2 — Page 1 of 4

Part | Mark

(a)(i) $f(x \mid \theta) = \prod_{i=1}^{n} f(x_i \mid \theta) = \prod_{i=1}^{n} \theta \exp(-x_i \theta)$, $\theta > 0$

$$= \theta^n \exp\left\{-\left(\sum_{i=1}^{n} x_i\right)\theta\right\}$$

$$\log f(x \mid \theta) = n \log \theta - \left(\sum_{i=1}^{n} x_i\right)\theta$$

$$\frac{\partial}{\partial \theta} \log f(x \mid \theta) = \frac{n}{\theta} - \sum_{i=1}^{n} x_i, \quad \frac{\partial^2}{\partial \theta^2} \log f(x \mid \theta) = -\frac{n}{\theta^2}$$

$$I(\theta) = -E\left(-\frac{n}{\theta^2} \,\middle|\, \theta\right) = \frac{n}{\theta^2}.$$

The Jeffreys prior is $f(\theta) \propto \sqrt{I(\theta)}$
$\propto \theta^{-1}$ (the improper Gamma$(0,0)$)

$$f(\theta \mid x) \propto f(x \mid \theta) f(\theta)$$
$$\propto \theta^n \exp\left\{-\left(\sum_{i=1}^{n} x_i\right)\theta\right\}\theta^{-1}$$
$$= \theta^{n-1} \exp\left\{-\left(\sum_{i=1}^{n} x_i\right)\theta\right\}$$

We recognise this as a kernel of Gamma$(n, \sum_{i=1}^{n} x_i)$ so that $\theta \mid x \sim$ Gamma(α_n, β_n) where $\alpha_n = n$ and $\beta_n = \sum_{i=1}^{n} x_i$. 4

Total

Unit Code MA40189	Unit Title TOPICS IN BAYESIAN STATISTICS	
Academic Year 2019/20	Examiner SIMON SHAW	
Semester II	Question No. 2	Page 2 of 4

Part | Mark

1(ii) $E(\theta|X) = \dfrac{\alpha_n}{\beta_n} = \dfrac{n}{\sum_{i=1}^{n} x_i}$ which is the maximum likelihood estimate of θ. The posterior mean is thus based solely on the data and the prior has no influence: it is noninformative. 2

1(iii) As Z is exchangeable then $Z|\theta \sim Exp(\theta)$ and $(X \perp\!\!\!\perp Z)|\theta$. Thus,

$$\begin{aligned}E(Z|X) &= E(E(Z|\theta)|X) \quad (\text{using } (X \perp\!\!\!\perp Z)|\theta)\\ &= E(\theta^{-1}|X) \quad (\text{as } Z|\theta \sim Exp(\theta))\\ &= \frac{\beta_n}{\alpha_n - 1} = \frac{\sum_{i=1}^{n} x_i}{n-1} \quad (\text{as } \theta|x \sim Gamma(\alpha_n, \beta_n) \text{ so } \theta^{-1}|x \sim Inv\text{-}gamma(\alpha_n, \beta_n))\end{aligned}$$

Similarly,

$$\begin{aligned}Var(Z|X) &= E(Var(Z|\theta)|X) + Var(E(Z|\theta)|X)\\ &= E(\theta^{-2}|X) + Var(\theta^{-1}|X)\\ &= \frac{\beta_n^2}{(\alpha_n-1)(\alpha_n-2)} + \frac{\beta_n^2}{(\alpha_n-1)^2(\alpha_n-2)}\\ &= \frac{\alpha_n\beta_n^2}{(\alpha_n-1)^2(\alpha_n-2)} = \frac{n\left(\sum_{i=1}^{n} x_i\right)^2}{(n-1)^2(n-2)}\end{aligned}$$

(both values well defined for $n > 2$). 3

Total

Unit Code MA40189	Unit Title TOPICS IN BAYESIAN STATISTICS	
Academic Year 2019/20	Examiner SIMON SHAW	
Semester II	Question No. 2	Page 3 of 4

Part | Mark

(iv) $$f(\theta) = \frac{\beta^{\alpha}}{4\Gamma(\alpha)}\theta^{\alpha-1}\exp(-\beta\theta) + \frac{3\beta^{\alpha+1}}{4\Gamma(\alpha+1)}\theta^{\alpha}\exp(-\beta\theta)$$

$$\propto \theta^{\alpha-1}\exp(-\beta\theta) + \frac{3\beta}{\alpha}\theta^{\alpha}\exp(-\beta\theta)$$

$$\propto \alpha\theta^{\alpha-1}\exp(-\beta\theta) + 3\beta\theta^{\alpha}\exp(-\beta\theta).$$

$$f(\theta|x) \propto f(x|\theta)f(\theta)$$

$$\propto \alpha\theta^{(\alpha+n)-1}\exp\left(-(\beta+\sum_{i=1}^{n}x_i)\theta\right) + 3\beta\theta^{(\alpha+1+n)-1}\exp\left(-(\beta+\sum_{i=1}^{n}x_i)\theta\right)$$

We recognise kernels of Gamma$(\alpha+n, \beta+\sum_{i=1}^{n}x_i)$ and Gamma$(\alpha+1+n, \beta+\sum_{i=1}^{n}x_i)$ in this sum. Letting $f_1(\theta)$ be the pdf of Gamma$(\alpha+n, \beta+\sum_{i=1}^{n}x_i)$ and $f_2(\theta)$ the pdf of Gamma$(\alpha+1+n, \beta+\sum_{i=1}^{n}x_i)$ we have:

$$f(\theta|x) \propto \alpha\frac{\Gamma(\alpha+n)}{(\beta+\sum_{i=1}^{n}x_i)^{\alpha+n}}f_1(\theta) + \frac{3\beta\Gamma(\alpha+n+1)}{(\beta+\sum_{i=1}^{n}x_i)^{\alpha+n+1}}f_2(\theta)$$

$$\propto \alpha f_1(\theta) + \frac{3\beta(\alpha+n)}{(\beta+\sum_{i=1}^{n}x_i)}f_2(\theta)$$

$$\propto \alpha(\beta+\sum_{i=1}^{n}x_i)f_1(\theta) + 3\beta(\alpha+n)f_2(\theta)$$

Total

DEPARTMENT OF MATHEMATICAL SCIENCES
Outline solution to Examination Questions

UNIVERSITY OF BATH

Unit Code MA40189	Unit Title TOPICS IN BAYESIAN STATISTICS	
Academic Year 2019/20	Examiner SIMON SHAW	
Semester II	Question No. 2	Page 4 of 4

Part | Mark

Thus, as $\int_\theta f_1(\theta)\,d\theta = 1 = \int_\theta f_2(\theta)\,d\theta$ we have:

$$f(\theta \mid x) = c\,f_1(\theta) + (1-c)\,f_2(\theta) \text{ where } c = \frac{\alpha(\beta + \sum_{i=1}^n x_i)}{\alpha(\beta+\sum_{i=1}^n x_i) + 3\beta(\alpha+n)}$$ 5

(b) (i) The RVs $X_1, \ldots, X_n$ are judged to be finitely exchangeable if their joint density function satisfies

$$f(x_1, \ldots, x_n) = f(x_{\pi(1)}, \ldots, x_{\pi(n)})$$

for all permutations π defined on the set $\{1, \ldots, n\}$. 2

(ii) $$f(x_1, \ldots, x_n \mid \theta) = \prod_{i=1}^n f(x_i \mid \theta) = \prod_{i=1}^n f(x_i \mid \theta_i).$$

Now, $$f(x_1, \ldots, x_n) = \int_\theta f(x \mid \theta) f(\theta)\,d\theta$$
$$= \int_\theta \left\{\prod_{i=1}^n f(x_i \mid \theta_i)\right\}\left\{\prod_{i=1}^n \pi(\theta_i)\right\} d\theta$$
$$= \prod_{i=1}^n \int_{\theta_i} f(x_i \mid \theta_i)\pi(\theta_i)\,d\theta_i = \prod_{i=1}^n f(x_i)$$

as each marginal X_i is identically distributed. Thus, the ordering of the x_is is irrelevant (they're iid) and thus they are finitely exchangeable. 4

Total 20

UNIVERSITY OF BATH

Unit Code MA40189	Unit Title TOPICS IN BAYESIAN STATISTICS	
Academic Year 2019/20	Examiner SIMON SHAW	
Semester II	Question No. 3	Page 1 of 4

Part | Mark

(a) As $X_i \mid \theta \sim Po(s_i \theta_i)$ then $\mathbb{P}(X_i = x_i \mid \theta) = \frac{1}{x_i!}(s_i\theta_i)^{x_i}\exp(-s_i\theta_i)$

$$\propto \theta_i^{x_i}\exp(-s_i\theta_i)$$

Hence, $f(x \mid \theta) = \mathbb{P}(X_1 = x_1, \ldots, X_n = x_n \mid \theta) \propto \prod_{i=1}^{n} \theta_i^{x_i}\exp(-s_i\theta_i)$

$$= \left(\prod_{i=1}^{n}\theta_i^{x_i}\right)\exp\left(-\sum_{i=1}^{n} s_i\theta_i\right)$$

$\theta_i \mid \phi \sim Exp(\phi)$ so that $f(\theta \mid \phi) = \prod_{i=1}^{n} \phi\exp(-\phi\theta_i) = \phi^n\exp\left(-\phi\sum_{i=1}^{n}\theta_i\right)$

$\phi \sim Gamma(\alpha, \beta)$ so that $f(\phi) \propto \phi^{\alpha-1}\exp(-\beta\phi)$

Then, $f(\theta, \phi \mid x) \propto f(x \mid \theta) f(\theta \mid \phi) f(\phi)$

$$\propto \left(\prod_{i=1}^{n}\theta_i^{x_i}\right)\phi^{\alpha+n-1}\exp\left\{-\phi\left(\beta + \sum_{i=1}^{n}\theta_i\right) - \sum_{i=1}^{n} s_i\theta_i\right\}.$$

2

(b) Let $\theta_{-i} = (\theta_1, \ldots, \theta_n) / \theta_i$.

$$f(\theta_i \mid \theta_{-i}, \phi, x) = \frac{f(\theta, \phi \mid x)}{f(\theta_{-i}, \phi \mid x)} \propto f(\theta, \phi \mid x)$$

as a function of θ_i

Total

Unit Code MA40189	Unit Title TOPICS IN BAYESIAN STATISTICS	
Academic Year 2019/20	Examiner SIMON SHAW	
Semester II	Question No. 3	Page 2 of 4

Part | Mark

Thus, $f(\theta_i \mid \theta_{-i}, \phi, x) \propto \theta_i^{x_i} \exp\{-(\phi + s_i)\theta_i\}$

We recognise this as a kernel of $Gamma(x_i+1, \phi+s_i)$ so

$$\theta_i \mid \theta_{-i}, \phi, x \sim Gamma(x_i+1, \phi+s_i).$$

Similarly, $f(\phi \mid \theta, x) \propto f(\theta, \phi \mid x)$ (wrt ϕ)

$$\propto \phi^{\alpha+n-1} \exp\left\{-\phi\left(\beta + \sum_{i=1}^{n} \theta_i\right)\right\}$$

We recognise this as a kernel of $Gamma(\alpha+n, \beta + \sum_{i=1}^{n} \theta_i)$ so

$$\phi \mid \theta, x \sim Gamma(\alpha+n, \beta + \sum_{i=1}^{n} \theta_i).$$

The Gibbs sampler algorithm is:

1. Choose a starting value $(\theta^{(0)}, \phi^{(0)})$ for which $f(\theta^{(0)}, \phi^{(0)} \mid x) > 0$
2. At iteration t generate new values $(\theta^{(t)}, \phi^{(t)})$ as follows
   - draw $\theta_1^{(t)}$ from $Gamma(x_1+1, \phi^{(t-1)}+1)$

     $\vdots$
   - draw $\theta_n^{(t)}$ from $Gamma(x_n+1, \phi^{(t-1)}+1)$
   - draw $\phi^{(t)}$ from $Gamma(\alpha+n, \beta + \sum_{i=1}^{n} \theta_i^{(t)})$
3. Repeat step 2.

Total

DEPARTMENT OF MATHEMATICAL SCIENCES
Outline solution to Examination Questions

UNIVERSITY OF BATH

Unit Code MA40189	Unit Title TOPICS IN BAYESIAN STATISTICS		
Academic Year 2019/20	Examiner SIMON SHAW		
Semester II	Question No. 3	Page 3 of 4	

Part | Mark

The algorithm will produce a Markov chain with stationary distribution $f(\theta, \phi | x)$. After a sufficiently long time to allow for convergence, values $\{\theta^{(t)}, \phi^{(t)}\}$ for $t > b$ may be viewed as a sample from $f(\theta, \phi | x)$ (b denotes the burn-in). 7

(c) The following Metropolis-Hastings algorithm generates a sequence of values $\theta^{(1)}, \theta^{(2)}, \ldots$ which form a Markov chain with stationary distribution $\pi(\theta) = f(\theta | x)$.

1. Choose an arbitrary starting point $\theta^{(0)}$ for which $f(\theta^{(0)} | x) > 0$.
2. At time t,
   a) Sample a candidate point or proposal θ^* from $q(\theta | \theta^{(t-1)})$, the proposal distribution.
   b) Calculate the acceptance probability
   $$\alpha(\theta^{(t-1)}, \theta^*) = \min\left(1, \frac{f(\theta^* | x)\, q(\theta^{(t-1)} | \theta^*)}{f(\theta^{(t-1)} | x)\, q(\theta^* | \theta^{(t-1)})}\right)$$
   c) Generate $U \sim U(0,1)$.
   d) If $U \leq \alpha(\theta^{(t-1)}, \theta^*)$ accept the proposal, $\theta^{(t)} = \theta^*$. Otherwise, reject the proposal, $\theta^{(t)} = \theta^{(t-1)}$.
3. Repeat step 2. 4

Total

Unit Code MA40189	Unit Title TOPICS IN BAYESIAN STATISTICS	
Academic Year 2019/20	Examiner SIMON SHAW	
Semester II	Question No. 3	Page 4 of 4

Part | Mark

(d) $$E(g(\theta)\mid X) = \int_\theta g(\theta) f(\theta\mid x)\, d\theta$$

$$= \int_\theta \frac{g(\theta)f(\theta\mid x)}{q(\theta)}\, q(\theta)\, d\theta = E\left(\frac{g(\theta)f(\theta\mid x)}{q(\theta)} \,\middle|\, \theta \sim q(\theta)\right)$$

Draw a random sample $\theta_1, \ldots, \theta_N$ from $q(\theta)$ and approximate $E(g(\theta)\mid X)$ by $\hat{I} = \frac{1}{N}\sum_{i=1}^{n} \frac{g(\theta_i)f(\theta_i\mid x)}{q(\theta_i)}$. 3

(e) $$Var(\hat{I}\mid \theta \sim q(\theta)) = \frac{1}{N} Var\left(\frac{g(\theta)f(\theta\mid X)}{q(\theta)} \,\middle|\, \theta \sim q(\theta)\right)$$

as θ_i iid from $q(\theta)$. Thus,

$$Var(\hat{I}\mid \theta \sim q(\theta)) = \frac{1}{N}\left(E\left(\frac{g^2(\theta)f^2(\theta\mid X)}{q^2(\theta)} \,\middle|\, \theta \sim q(\theta)\right) - E^2\left(\frac{g(\theta)f(\theta\mid X)}{q(\theta)} \,\middle|\, \theta \sim q(\theta)\right)\right)$$

$$= \frac{1}{N}\left(E\left(\frac{g^2(\theta)f(\theta\mid x)}{q(\theta)} \,\middle|\, X\right) - E^2(g(\theta)\mid X)\right)$$ by changing the measure

Thus, choosing $q(\theta)$ to minimise $E(g^2(\theta)f(\theta\mid x)/q(\theta)\mid X)$ will minimise $Var(\hat{I}\mid \theta \sim q(\theta))$ and any $q(\theta)$ doing so is optimal. 4

Total 20

Unit Code MA40189	Unit Title TOPICS IN BAYESIAN STATISTICS	
Academic Year 2019/20	Examiner SIMON SHAW	
Semester II	Question No. 4	Page 1 of 4

Part | Mark

(a) The risk of decision d is

$$\rho(\pi, d) = E\{g(\theta)(\theta-d)^2\} \quad (\text{wrt } \pi(\theta))$$
$$= E(g(\theta)\theta^2) - 2d\,E(g(\theta)\theta) + d^2 E(g(\theta))$$

We choose d to minimise this:

$$\frac{\partial}{\partial d}\rho(\pi, d) = -2E(g(\theta)\theta) + 2d\,E(g(\theta))$$

$$\text{so} \quad d^* = \frac{E(g(\theta)\theta)}{E(g(\theta))}$$

which is a minimum for $g(\theta) > 0$. The corresponding Bayes risk is

$$\rho^*(\pi) = \rho(\pi, d^*) = E(g(\theta)\theta^2) - 2\frac{E^2(g(\theta)\theta)}{E(g(\theta))} + \frac{E^2(g(\theta)\theta)}{E(g(\theta))}$$
$$= E(g(\theta)\theta^2) - \frac{E^2(g(\theta)\theta)}{E(g(\theta))}$$

When $g(\theta) = 1$, $d^* = E(\theta)$ (the mean) and $\rho^*(\pi) = Var(\theta)$ (the variance). 4

Total

Unit Code MA40189	Unit Title TOPICS IN BAYESIAN STATISTICS	
Academic Year 2019/20	Examiner SIMON SHAW	
Semester II	Question No. 4	Page 2 of 4

Part | Mark

(b) We have $g(\theta) = \theta^{-3}$. For the immediate decision, the Bayes rule is $d^* = \dfrac{E(\theta^{-2})}{E(\theta^{-3})}$

Now as $\theta \sim \text{Gamma}(\alpha, \beta)$,

$$E(\theta^{-k}) = \int_0^\infty \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{(\alpha-k)-1} e^{-\beta\theta} \, d\theta$$

$$= \frac{\beta^k \Gamma(\alpha-k)}{\Gamma(\alpha)} \int_0^\infty \frac{\beta^{\alpha-k}}{\Gamma(\alpha-k)} \theta^{(\alpha-k)-1} e^{-\beta\theta} \, d\theta$$

$$= \frac{\beta^k \Gamma(\alpha-k)}{\Gamma(\alpha)} \text{ provided that } \alpha > k.$$

Hence, $E(\theta^{-3}) = \dfrac{\beta^3 \Gamma(\alpha-3)}{\Gamma(\alpha)}$, $E(\theta^{-2}) = \dfrac{\beta^2 \Gamma(\alpha-2)}{\Gamma(\alpha)}$

so that $\lambda^* = \dfrac{\beta^2 \Gamma(\alpha-2)}{\Gamma(\alpha)} \times \dfrac{\Gamma(\alpha)}{\beta^3 \Gamma(\alpha-3)} = \dfrac{\alpha-3}{\beta}$.

The Bayes risk is $\rho^*(f(\theta)) = E(\theta^{-1}) - \dfrac{E^2(\theta^{-2})}{E(\theta^{-3})}$

$$= \frac{\beta\Gamma(\alpha-1)}{\Gamma(\alpha)} - \beta^2 \frac{\Gamma(\alpha-2)}{\Gamma(\alpha)} \left(\frac{\alpha-3}{\beta}\right)$$

$$= \frac{\beta}{\alpha-1} - \frac{\beta(\alpha-3)}{(\alpha-1)(\alpha-2)} = \frac{\beta}{(\alpha-1)(\alpha-2)}$$ 5

Total

(c) $f(x|\theta) = \prod_{i=1}^{n} f(x_i|\theta) \propto \theta^n \exp\left(-\theta \sum_{i=1}^{n} x_i^{\lambda}\right)$

$$f(\theta) \propto \theta^{\alpha-1} \exp(-\beta\theta)$$

Thus, $f(\theta|x) \propto \theta^{(\alpha+n)-1} \exp\left(-\theta\left(\beta + \sum_{i=1}^{n} x_i^{\lambda}\right)\right)$ so that $\theta|x \sim \text{Gamma}\left(\alpha+n, \beta + \sum_{i=1}^{n} x_i^{\lambda}\right)$.

We may exploit conjugacy to find $d^*(x) = \dfrac{\alpha+n-3}{\beta + \sum_{i=1}^{n} x_i^{\lambda}}$ with risk $\dfrac{\beta + \sum_{i=1}^{n} x_i^{\lambda}}{(\alpha+n-1)(\alpha+n-2)}$. 3

(d) With $\lambda = 1$, the risk of the sampling procedure is $E\left(\dfrac{\beta + \sum_{i=1}^{n} X_i}{(\alpha+n-1)(\alpha+n-2)}\right)$

Now, $E(X_i) = E(E(X_i|\theta)) = E(\theta^{-1}) = \dfrac{\beta}{\alpha-1}$ so that

$$E\left(\frac{\beta + \sum_{i=1}^{n} X_i}{(\alpha+n-1)(\alpha+n-2)}\right) = \frac{\beta + \frac{n\beta}{\alpha-1}}{(\alpha+n-1)(\alpha+n-2)} = \frac{\beta}{(\alpha-1)(\alpha+n-2)}$$

The total risk is $R_n = nc + \dfrac{\beta}{(\alpha-1)(\alpha+n-2)}$

Hence, $\dfrac{dR_n}{dn} = c - \dfrac{\beta}{(\alpha-1)(\alpha+n-2)^2}$

DEPARTMENT OF MATHEMATICAL SCIENCES
Outline solution to Examination Questions

UNIVERSITY OF BATH

Unit Code MA40189	Unit Title TOPICS IN BAYESIAN STATISTICS
Academic Year 2019/20	Examiner SIMON SHAW
Semester II	Question No. 4 — Page 4 of 4

Part | Mark

Solving to 0 gives $(\alpha+n-2)^2 = \dfrac{\beta}{c(\alpha-1)}$

$$\Rightarrow n = \sqrt{\frac{\beta}{c(\alpha-1)}} - (\alpha-2).$$

(Sample if $n>0$ otherwise don't sample). 4.

(c) $E(L_1(\theta,d)) = (\theta_2-\theta_1) + \displaystyle\int_{-\infty}^{\theta_1} \frac{2}{\alpha}(\theta_1-\theta)\pi(\theta)\,d\theta + \int_{\theta_2}^{\infty} \frac{2}{\alpha}(\theta-\theta_2)\pi(\theta)\,d\theta$

$$= (\theta_2-\theta_1) + \frac{2}{\alpha}\theta_1 P(\theta\le\theta_1) - \frac{2}{\alpha}\int_{-\infty}^{\theta_1}\theta\pi(\theta)\,d\theta$$

$$+ \frac{2}{\alpha}\left[E(\theta) - \int_{-\infty}^{\theta_2}\theta\pi(\theta)\,d\theta\right] - \frac{2}{\alpha}\theta_2 \mathbb{P}(\theta\ge\theta_2)$$

$$\frac{\partial}{\partial\theta_1}E(L_1(\theta,d)) = -1 + \frac{2}{\alpha}P(\theta\le\theta_1) + \frac{2}{\alpha}\theta_1\pi(\theta_1) - \frac{2}{\alpha}\theta_1\pi(\theta_1)$$

Hence, $\dfrac{\partial}{\partial\theta_1}E(L_1(\theta,d)) = 0$ when $P(\theta\le\theta_1) = \dfrac{\alpha}{2}$.

$$\frac{\partial}{\partial\theta_2}E(L_1(\theta,d)) = 1 - \frac{2}{\alpha}\theta_2\pi(\theta_2) - \frac{2}{\alpha}P(\theta\ge\theta_2) + \frac{2}{\alpha}\theta_2\pi(\theta_2)$$

Hence, $\dfrac{\partial}{\partial\theta_2}E(L_1(\theta,d)) = 0$ when $1 - \dfrac{2}{\alpha}P(\theta\ge\theta_2) = 0$

i.e. $P(\theta\ge\theta_2) = \dfrac{\alpha}{2}$

Hence, the Bayes rule is $d^* = (\theta_1^*, \theta_2^*)$ when $P(\theta\le\theta_1^*) = \dfrac{\alpha}{2} = P(\theta\ge\theta_2^*)$ 4

Total 20