UNIVERSITY OF BRISTOL

School of Mathematics

MATH 30015
BAYESIAN MODELLING
(Paper Code MATH-30015)

Summer 2018

2 hour and 30 minutes

This paper contains **FOUR** questions.
ALL answers will be used for assessment.

Calculators of an approved type (non-programmable, no text facility) are permitted.

On this examination, the marking scheme is indicative and is intended only as a guide to the relative weighting of the questions.

Do not turn over until instructed.

1. **(25 marks)** Let $x_1, \ldots, x_n$ be n observations that we model as independent Binomial(θ) random variables, with $\theta \in \Theta := (0,1)$. The likelihood of observation $x_1 \in \{0,1\}$ given θ is therefore given by

$$\tilde{f}(x_1|\theta) = \theta^{x_1}(1-\theta)^{1-x_1}.$$

Below we denote by $f(x^{(n)}|\theta)$ the likelihood of the observation $x^{(n)} := (x_1, ..., x_n)$ given θ; that is

$$f(x^{(n)}|\theta) = \prod_{k=1}^{n} \tilde{f}(x_k|\theta),$$

and assume that the prior distribution of θ is the Beta(α_0, β_0) distribution, i.e.,

$$\pi(\theta) = \frac{\Gamma(\alpha_0+\beta_0)}{\Gamma(\alpha_0)\Gamma(\beta_0)}\theta^{\alpha_0-1}(1-\theta)^{\beta_0-1}, \quad \theta \in \Theta$$

where $\alpha_0 \in (2, +\infty)$ and $\beta_0 \in (2, +\infty)$ are known hyper-parameters.

Hint: We recall the following result that may be useful below. For any $t > 0$ the relationship $\Gamma(t+1) = t\,\Gamma(t)$ holds.

(a) **(2 marks)** Show that the posterior density of θ given $x^{(n)}$ is given by

$$\pi(\theta|x^{(n)}) = \frac{\Gamma(\alpha_n+\beta_n)}{\Gamma(\alpha_n)\Gamma(\beta_n)}\theta^{\alpha_n-1}(1-\theta)^{\beta_n-1}, \quad \theta \in \Theta$$

with

$$\alpha_n = \alpha_0 + \sum_{k=1}^{n} x_k, \quad \beta_n = \beta_0 + n - \sum_{k=1}^{n} x_k.$$

(b) **(5 marks)** Let $\theta_n^* : \{0,1\}^n \to \Theta$ be such that $\theta_n^*(z) = 0.2$ for any observation $z \in \{0,1\}^n$. Assuming a quadratic loss function, show carefully that the estimator θ_n^* is admissible. Assume for this question that $n = 1$.

(c) **(10 marks)** Let $L : \Theta \times \Theta \to [0, +\infty)$ be the loss function defined by

$$L(\theta, d) = \frac{(\theta - d)^2}{\theta(1-\theta)}, \quad (\theta, d) \in \Theta \times \Theta.$$

Compute $\delta^\pi(x^{(n)})$, the minimizer of the resulting posterior expected loss.

(d) **(8 marks)** Show that $\delta^\pi : \{0,1\}^n \to \Theta$, the estimator defined in part (c), is admissible. Assume for this question that $n = 1$.

2. **(25 marks)** Let $X_1, \dots, X_n$ be n independent and identically distributed random variables such that $X_1 \sim \text{Binomial}(\theta_0)$ for some $\theta_0 \in \Theta := (0,1)$. The likelihood of $x_1 \in \{0,1\}$ given θ is therefore given by

$$\tilde{f}(x_1|\theta) = \theta^{x_1}(1-\theta)^{1-x_1}$$

and we assume that the prior distribution of θ is the $\text{Beta}(\alpha_0, \beta_0)$ distribution, i.e.,

$$\pi(\theta) = \frac{\Gamma(\alpha_0+\beta_0)}{\Gamma(\alpha_0)\Gamma(\beta_0)}\theta^{\alpha_0-1}(1-\theta)^{\beta_0-1}, \quad \theta \in \Theta$$

where $\alpha_0 \in (0,+\infty)$ and $\beta_0 \in (0,+\infty)$ are known hyper-parameters. The corresponding posterior mean of θ given $X^{(n)}$ is given by

$$\mathbb{E}_\pi[\theta|X^{(n)}] = \frac{\alpha_0 + \sum_{k=1}^n X_k}{\alpha_0+\beta_0+n}$$

while the maximum likelihood estimator of θ is given by $\hat{\theta}_n = \frac{1}{n}\sum_{k=1}^n X_k$.

(a) **(5 marks)** Show that, as $n \to +\infty$,

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \overset{\text{dist.}}{\Longrightarrow} \mathcal{N}_1(0, \theta_0(1-\theta_0))$$

and

$$\sqrt{n}\big(\mathbb{E}_\pi[\theta|X^{(n)}] - \theta_0\big) \overset{\text{dist.}}{\Longrightarrow} \mathcal{N}_1(0, \theta_0(1-\theta_0)).$$

(b) **(5 marks)** Show that

$$\limsup_{n\to+\infty} n\big|\mathbb{E}_\pi[\theta|X^{(n)}] - \hat{\theta}_n\big| < +\infty \quad \text{(almost surely)}.$$

(c) **(5 marks)** Using the results in part (a) and in part (b) explain why a frequentist statistician may be willing to use the estimator $\mathbb{E}_\pi[\theta|X^{(n)}]$.

(d) **(10 marks)** It can be shown that the Bayesian model we are considering in this question is such that

- there exists a constant $c > 0$ for which

$$\mathbb{P}_{\theta_0}\bigg(\int_\Theta \prod_{k=1}^n \frac{\tilde{f}(X_k|\theta)}{\tilde{f}(X_k|\theta_0)}\pi(\theta)\mathrm{d}\theta \le c\,\epsilon\, e^{-2\,n\epsilon^2}\bigg) \le \frac{1}{n\epsilon^2}, \quad \forall\epsilon > 0, \quad \forall n \ge 1,$$

- for any $\epsilon > 0$ there exists a sequence of tests $(\phi_n)_{n\ge1}$ such that

$$\mathbb{E}_{\theta_0}[\phi_n(X^{(n)})] \le 2\,e^{-n\frac{\epsilon^2}{2}}, \qquad \sup_{\theta\in\{\theta'\in\Theta:\,|\theta'-\theta_0|\ge\epsilon\}} \mathbb{E}_\theta[1-\phi_n(X^{(n)})] \le 2\,e^{-n\frac{\epsilon^2}{2}}, \quad \forall n \ge 1.$$

Using these results show that

$$\lim_{n\to+\infty} \mathbb{E}_{\theta_0}\Big[\pi\big(\{\theta\in\Theta:\ |\theta-\theta_0|\ge\epsilon\}|X^{(n)}\big)\Big] = 0.$$

3. **(25 marks)** Let $X_1, \ldots, X_n$ be n random variables that we model as independent $\mathcal{N}_1(\theta, 1)$ random variables, with $\theta \in \Theta := \mathbb{R}$. We assign to θ a $\mathcal{N}_1(\mu_0, \sigma_0^2)$ prior distribution so that the posterior distribution $\pi(\theta|X^{(n)})$ is given by

$$\pi(\theta|X^{(n)}) = \frac{1}{\sqrt{2\pi\sigma_n^2}} \exp\Big(-\frac{(\theta-\mu_n)^2}{2\sigma_n^2}\Big), \quad \theta \in \Theta$$

where

$$\mu_n = \frac{1}{1+n\sigma_0^2}\mu_0 + \frac{\sigma_0^2}{1+n\sigma_0^2}\sum_{k=1}^{n} X_k, \qquad \sigma_n^2 = \frac{\sigma_0^2}{1+n\sigma_0^2}.$$

The maximum likelihood estimator of θ is given by $\hat{\theta}_n = \frac{1}{n}\sum_{k=1}^n X_k$.

(a) **(5 marks)** Give the Highest Posterior Density (HPD) region at level $(1-\alpha)$, with $\alpha \in (0,1)$.

(b) We first assume that the model is well-specified; that is, $X_1, \ldots, X_n$ are independent and identically distributed random variables such that $X_1 \sim \mathcal{N}_1(\theta_0, 1)$ for a $\theta_0 \in \Theta$.

i. **(10 marks)** Let $S_n = \sqrt{n}(\theta - \hat{\theta}_n)$ and $\pi^*(s|X^{(n)})$ be the probability density function of the posterior distribution of S_n given $X^{(n)}$. Show that, for any $s \in \mathbb{R}$,

$$\lim_{n\to+\infty} \pi^*(s|X^{(n)}) = \frac{1}{\sqrt{2\pi}} e^{-\frac{s^2}{2}} \quad \text{(almost surely)}$$

so that, as $n \to +\infty$, we have $\sqrt{n}(\theta - \hat{\theta}_n)|X^{(n)} \overset{\text{dist.}}{\Longrightarrow} \mathcal{N}_1(0,1)$ (almost surely).

ii. **(2 marks)** Give the confidence interval at level $1-\alpha$ centred at $\hat{\theta}_n$.

iii. **(2 marks)** Use these results to compare, as $n \to +\infty$, the credible interval at level $1-\alpha$ computed in part (a) and the confidence interval at level $1-\alpha$ computed in part (b)ii.

(c) We now assume that the model is misspecified. More precisely, we assume that $X_1, \ldots, X_n$ are independent and identically distributed random variables such that $X_1 \sim t_6(\theta_0, 1)$ for a $\theta_0 \in \Theta$. Here, $t_6(\theta_0, 1)$ denotes the Student-t distribution with location parameter θ_0, scale parameter 1 and 6 degrees of freedom. Note that $\mathbb{E}[X_1] = \theta_0$ while $\mathrm{Var}(X_1) = 1.5$.

i. **(2 mark)** Using the result in part (b)i deduce that, as $n \to +\infty$, we have $\sqrt{n}(\theta - \hat{\theta}_n)|X^{(n)} \overset{\text{dist.}}{\Longrightarrow} \mathcal{N}_1(0,1)$ (almost surely).

ii. **(2 mark)** Show that, as $n \to +\infty$, we have $\sqrt{n}\big(\hat{\theta}_n - \theta_0\big) \overset{\text{dist.}}{\Longrightarrow} \mathcal{N}_1(0, 1.5)$ and give the confidence interval at level $1-\alpha$ centred at $\hat{\theta}_n$.

iii. **(2 marks)** Use these results to compare, as $n \to +\infty$, the credible interval at level $1-\alpha$ computed in part (a) and the confidence interval at level $1-\alpha$ computed in part (c)ii.

Continued...

4. **(25 marks)** Let $x_0, \dots, x_t$ be $t+1$ observations that we model as a Markov($\lambda, P_{\alpha,\beta}$) process, where $\lambda = (\lambda_1, 1-\lambda_1)$ for some $\lambda_1 \in [0,1]$, while, for $\alpha, \beta \in [0,1]$, the transition matrix $P_{\alpha,\beta}$ is given by

$$P_{\alpha,\beta} = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix}.$$

Let $\theta = (\lambda_1, \alpha, \beta) \in \Theta := [0,1]^3$ be the parameter, $\mathcal{X} = \{1,2\}$ be the state space and assume that our prior information is as follows:

- The prior probability that $P_{\alpha,\beta}$ is irreducible and aperiodic is 0.7;
- The prior probability that $\lim_{t\to+\infty} P^t_{\alpha,\beta}$ does not exist is 0.05;
- The prior probability that $P_{\alpha,\beta}$ is not irreducible is 0.25.

(a) **(2 marks)** Write down the likelihood function for the observation $x^{(t+1)} := (x_0, \dots, x_t)$.

(b) Before specifying a prior distribution $\pi(\theta)$ that incorporates the above information we need to understand the behaviour of $P_{\alpha,\beta}$ for any pair (α, β) in $[0,1]^2$. To this end it may be useful to note that, when $\alpha + \beta > 0$, we have

$$P^t_{\alpha,\beta} = \begin{pmatrix} \frac{\beta}{\alpha+\beta} + \frac{\alpha}{\alpha+\beta}(1-\alpha-\beta)^t & \frac{\alpha}{\alpha+\beta} - \frac{\alpha}{\alpha+\beta}(1-\alpha-\beta)^t \\ \frac{\beta}{\alpha+\beta} - \frac{\beta}{\alpha+\beta}(1-\alpha-\beta)^t & \frac{\alpha}{\alpha+\beta} + \frac{\beta}{\alpha+\beta}(1-\alpha-\beta)^t \end{pmatrix}, \quad \forall t \geq 1.$$

i. **(2 marks)** Show that $\lim_{t\to+\infty} P^t_{\alpha,\beta}$ does not exist if and only if $\alpha = \beta = 1$.

ii. **(2 marks)** Show that $P_{\alpha,\beta}$ is irreducible if and only if $\alpha > 0$ and $\beta > 0$.

iii. **(2 marks)** Show that $P_{\alpha,\beta}$ is aperiodic if and only if $(\alpha, \beta) \in \{(0,0)\} \cup (0,1)^2$.

(c) **(5 marks)** Using the results of part (b), propose a prior distribution $\pi(\theta)$ that captures the above prior information.

(d) Let $\pi(\theta|x^{(t+1)})$ be the posterior distribution associated with the likelihood function and prior distribution specified in part (a) and in part (c) respectively.

i. **(4 marks)** Write down a Metropolis-Hastings algorithm that is implementable in practice to approximate $\pi(\theta|x^{(t+1)})$ and specify explicitly all its ingredients. (Remark that you are not asked to propose a "good" Metropolis-Hastings algorithm, just an implementable one.)

ii. **(8 marks)** Explain why and in which sense the Metropolis-Hastings algorithm you propose in part (d)i is valid to approximate $\pi(\theta|x^{(t+1)})$. Explain how to use its output to approximate the posterior expectation $\mathbb{E}_\pi[\theta|x^{(t+1)}]$.

End of examination.