UNIVERSITY OF OSLO

Faculty of mathematics and natural sciences

Exam in:	STK4021/9021 — Applied Bayesian Analysis
Day of examination:	Monday, 25 November 2024
Examination hours:	15.00 – 19.00
This problem set consists of 4 pages.	
Appendices:	None
Permitted aids:	Approved calculator

Please make sure that your copy of the problem set is complete before you attempt to answer anything.

Problem 1

Let $y_1, \ldots, y_n$ be independent observations, each with the probability density

$$f_\theta(y) = \frac{\theta}{y^2}\left(\frac{y}{1+y}\right)^{1+\theta}, \quad \text{for } y > 0,$$

where $\theta > 0$ is a parameter.

(a) Verify that $f_\theta(y)$ is a valid probability density function. [Hint: use the substitution $s = 1/y$.]

(b) Show that y has infinite expectation. [You may assume that $\int_1^\infty \mathrm{d}y/y = \infty$.]

(c) Write down the log-likelihood and find the maximum likelihood estimator $\widehat{\theta}$ as a function of $w = \sum_{i=1}^n \log(1 + 1/y_i)$.

(d) Find the Jeffreys prior for θ. Is it proper or improper?

We will use a conjugate prior for the rest of the question. Let $\theta \sim \text{Gamma}(a, b)$, so that its prior density is

$$\pi(\theta) = \frac{b^a}{\Gamma(a)}\theta^{a-1}\exp\{-b\theta\}, \quad \text{for } \theta > 0,$$

where $a, b > 0$ are parameters. Here, $\Gamma(a) = \int_0^\infty t^{a-1}e^{-t}\,\mathrm{d}t$ is the Gamma function.

(e) Verify that $\mathbb{E}\,\theta = a/b$. [You may use that $\Gamma(x+1) = x\Gamma(x)$ for all $x > 0$.]

(Continued on page 2.)

(f) Show that the posterior distribution is also a Gamma distribution, $\theta \,|\, y_1, \ldots, y_n \sim \text{Gamma}(a_n, b_n)$, and specify the new parameters a_n, b_n.

(g) Show that the posterior mean can be written as

$$\mathbb{E}[\theta \,|\, y_1, \ldots, y_n] = \lambda \mathbb{E}\, \theta + (1 - \lambda)\widehat{\theta},$$

for some constant $0 \leq \lambda \leq 1$ (allowed to depend on the data) which you should specify.

(h) Now assume that x has a $\text{Beta}(\theta, 1)$ distribution, so its density is

$$g_\theta(x) = \theta x^{\theta - 1}, \quad \text{for } 0 < x < 1.$$

Show that $z = x/(1 - x)$ has density f_θ.

(i) Assume that you are able to simulate from the Beta and Gamma distributions. Give a recipe for how to generate a marginal sample $y_1, \ldots, y_n$.

Problem 2

Suppose we have observed a sequence of pairs of data points $(x_1, y_1), \ldots, (x_n, y_n)$, where $x_i \in \mathbb{R}^d$ and $y_i \in \mathbb{R}$ for $i = 1, \ldots, n$. We map each input x_i to a vector of p features,

$$\boldsymbol{\phi}(x_i) = (\phi_0(x_i), \phi_1(x_i), \ldots, \phi_{p-1}(x_i))^\top,$$

where by convention $\phi_0(x) = 1$. The vector of outcomes $\mathbf{y} = (y_1, \ldots, y_n)^\top$ is modelled using linear regression on the the features with Gaussian noise,

$$\mathbf{y} \,|\, \mathbf{w}, \beta \sim \text{N}(\Phi \mathbf{w}, \beta^{-1} I_n),$$

where $\mathbf{w} = (w_0, w_1, \ldots, w_{p-1})^\top$ is the vector of coefficients, $\beta > 0$ is the precision (inverse variance) of the model, I_n is the $n \times n$ identity matrix and

$$\Phi = \begin{pmatrix} \phi_0(x_1) & \phi_1(x_1) & \cdots & \phi_{p-1}(x_1) \\ \phi_0(x_2) & \phi_1(x_2) & \cdots & \phi_{p-1}(x_2) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_0(x_n) & \phi_1(x_n) & \cdots & \phi_{p-1}(x_n) \end{pmatrix}$$

is the $n \times p$ design matrix. Recall that a multivariate Gaussian random variable $\mathbf{x}$ with mean $\boldsymbol{\mu} \in \mathbb{R}^n$ and (symmetric and positive definite) covariance matrix $\Sigma \in \mathbb{R}^{n \times n}$ has density

$$f(\mathbf{x}; \boldsymbol{\mu}, \Sigma) = (2\pi)^{-n/2} |\Sigma|^{-1/2} \exp\left\{ -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^\top \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right\}, \quad \text{for } \mathbf{x} \in \mathbb{R}^n.$$

Recall also that the maximum likelihood estimator for $\mathbf{w}$ is given by $\widehat{\mathbf{w}} = (\Phi^\top \Phi)^{-1} \Phi^\top \mathbf{y}$.

We will impose a prior distribution on both $\mathbf{w}$ and β. For $\mathbf{w}$, we use a Gaussian prior, $\mathbf{w} \sim \text{N}(\mathbf{m}, S)$, with parameters $\mathbf{m} \in \mathbb{R}^p$ and $S \in \mathbb{R}^{p \times p}$. For β, we use a Gamma prior, $\beta \sim \text{Gamma}(a, b)$, with parameters $a, b > 0$.

(Continued on page 3.)

(a) In lectures we showed that when we condition on β, then the posterior distribution of $\mathbf{w} \,|\, \mathbf{y}, \beta$ is also Gaussian, $\mathbf{w} \,|\, \mathbf{y}, \beta \sim \mathrm{N}(\mathbf{m}_n, S_n)$, with

$$\mathbf{m}_n = S_n(S^{-1}\mathbf{m} + \beta\Phi^\top \mathbf{y}), \quad S_n^{-1} = S^{-1} + \beta\Phi^\top\Phi.$$

Show that we can write

$$\mathbf{m}_n = Q\mathbf{m} + (I_p - Q)\widehat{\mathbf{w}},$$

where you need to specify the $p \times p$ matrix Q.

(b) Show that the posterior distribution of $\beta \,|\, \mathbf{y}, \mathbf{w}$ is another Gamma distribution, $\beta \,|\, \mathbf{y}, \mathbf{w} \sim \mathrm{Gamma}(a_n, b_n)$, and specify its parameters a_n, b_n. [You may use that if A is an $n \times n$ matrix and $\lambda > 0$ is a positive scalar, then $|\lambda A| = \lambda^n |A|$.]

(c) Use the above results to construct a Gibbs sampling algorithm which targets the joint posterior $\pi(\mathbf{w}, \beta \,|\, \mathbf{y})$.

(d) Suppose we run your algorithm to obtain a sample $(\mathbf{w}^{(1)}, \beta^{(1)}), \ldots, (\mathbf{w}^{(S)}, \beta^{(S)})$, for some large number S. Explain briefly how you would assess the quality of this sample.

(e) How would you use your sample to estimate $\mathrm{median}\{\|\mathbf{w}\| \,|\, \mathbf{y}\}$? That is, the posterior median value of the norm $\|\mathbf{w}\| = \left\{\sum_{j=0}^{p-1} w_j^2\right\}^{1/2}$, given the data.

Problem 3

For $n > 1$, let $\mathbf{y} = (y_1, \ldots, y_n)^\top$ be a multivariate Gaussian random variable such that each coordinate y_i has mean $\mu \in \mathbb{R}$ and variance $\sigma^2 > 0$. Let also $\mathrm{cov}(y_i, y_j) = \rho\sigma^2$ for $i \neq j$, where $0 < \rho < 1$. In matrix notation, this means that $\mathbf{y}$ has mean $\boldsymbol{\mu} = (\mu, \ldots, \mu)^\top = \mu\mathbf{u}$, where $\mathbf{u} = (1, \ldots, 1)^\top$ is an n-dimensional vector of ones, and covariance matrix

$$\Sigma = \begin{pmatrix} \sigma^2 & \rho\sigma^2 & \cdots & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 & \cdots & \rho\sigma^2 \\ \vdots & \vdots & \ddots & \vdots \\ \rho\sigma^2 & \rho\sigma^2 & \cdots & \sigma^2 \end{pmatrix} = \alpha I_n + \beta\mathbf{u}\mathbf{u}^\top,$$

where $\alpha = \sigma^2 - \rho\sigma^2$ and $\beta = \rho\sigma^2$.

(a) Show that there exists a random variable $\theta \sim \mathrm{N}(m, s^2)$ such that the variables y_i can be realised as conditionally independent given θ. [You may assume that pairwise independence implies independence for jointly Gaussian random variables.]

(b) Deduce that the sequence $y_1, \ldots, y_n$ is exchangeable.

(Continued on page 4.)

(c) Consider the $n \times n$ matrix

$$T = \begin{pmatrix} \sigma^2 & -\rho\sigma^2 & \cdots & -\rho\sigma^2 \\ -\rho\sigma^2 & \sigma^2 & \cdots & -\rho\sigma^2 \\ \vdots & \vdots & \ddots & \vdots \\ -\rho\sigma^2 & -\rho\sigma^2 & \cdots & \sigma^2 \end{pmatrix} = \gamma I_n - \beta \mathbf{u}\mathbf{u}^\top,$$

where $\gamma = \sigma^2 + \rho\sigma^2$ and like before, $\beta = \rho\sigma^2$. Show that T is positive definite if and only if $\rho < 1/(n-1)$. [You may use that the eigenvalues of $\mathbf{u}\mathbf{u}^\top$ are precisely 0 and $\mathbf{u}^\top\mathbf{u}$. You may also assume standard facts about real symmetric matrices.]

(d) Verify that the inverse of T is given by

$$T^{-1} = \frac{1}{\gamma} I_n + \frac{\beta}{\gamma^2 - n\beta\gamma} \mathbf{u}\mathbf{u}^\top.$$

(e) Now let $\mathbf{z} = (z_1, \ldots, z_n)$ be a multivariate Gaussian random variable such that each coordinate z_i has mean μ and variance σ^2, and that the pairwise covariance is $\operatorname{cov}(z_i, z_j) = -\rho\sigma^2$ for $i \neq j$, where $0 < \rho < 1/(n-1)$. Show that the sequence $z_1, \ldots, z_n$ is exchangeable. [You may use that if $\mathbf{x}$ is an n-dimensional vector and $A = (a_{ij})$ is an $n \times n$ matrix, then $\mathbf{x}^\top A \mathbf{x} = \sum_{i,j=1}^n x_i a_{ij} x_j$.]

(f) Prove that there does *not* exist a random variable θ such that the variables z_i can be realised as conditionally independent given θ.

Remark. In this version, three typos have been corrected. In Problem 1(d), it originally said "for $y > 0$". This has been changed to "for $\theta > 0$". In Problem 2, it originally said "$\alpha \sim \text{Gamma}(a, b)$". This has been changed to "$\beta \sim \text{Gamma}(a, b)$". In problem 2, it originally said "$\widehat{\mathbf{w}} = (\Phi^\top\Phi)^{-1}\Phi\mathbf{y}$". This has been changed to "$\widehat{\mathbf{w}} = (\Phi^\top\Phi)^{-1}\Phi^\top\mathbf{y}$". Everyone automatically got full marks on question 2a because of this last typo.

THE END - GOOD LUCK!