MATH 559: Bayesian Theory and Methods

Practice Final Exam: Solutions
3 hours

MATH 559: Specific Instructions.

Please write your answers on the exam in the boxes provided; only answers written inside the boxes will be marked. Please write clearly. You may request a blank answer booklet for working, but any answers written on the booklet will not be marked.

This paper contains five questions. Each question carries 20 marks. Credit will be given for all questions attempted. The total mark available is 100 but rescaling of the final mark may occur.

A formula sheet is provided. No crib sheet is allowed.

Only non-programmable calculators may be used. Dictionaries and Translation dictionaries are permitted.

The usual abbreviations are used: pmf indicates the probability mass function; pdf indicates the probability density function; cdf indicates the cumulative distribution function; i.i.d. indicates independent and identically distributed. The function $\mathbb{1}_A(x)$ is the indicator function for set A that takes the value 1 if $x \in A$ and zero otherwise.

If the intention of a question is unclear to you, please state your interpretation and provide an answer accordingly. Your comment will be taken into account in marking.

This page intentionally left blank.

1. Random variables $Y_1, Y_2, \ldots$ are infinitely exchangeable.

   (a) Suppose for each i, Y_i is binary.

(i) Write down the de Finetti representation for the joint pmf for $Y_1, \ldots, Y_n$, $n \geq 1$ explaining the meaning of each of the terms, including the definition of the *parameter*, denoted θ, its *true value*, denoted θ_0, and the *prior distribution* π_0. 6 MARKS

Q1(a)(i): The representation is

$$p_{Y_1,\ldots,Y_n}(y_1, \ldots, y_n) = \int_0^1 \prod_{i=1}^n f_Y(y_i; \theta)\pi_0(d\theta)$$

where

- $f_Y(y; \theta) = \theta^y(1-\theta)^{1-y}$ is the conditional distribution of each Y_i
- θ is a generic parameter that defines the presumed conditional distribution, where

$$\Pr[Y = 1; \theta] = \theta$$

- The term

$$\prod_{i=1}^n f_Y(y_i; \theta)$$

  is the likelihood that describes the conditional joint pmf of $Y_1, \ldots, Y_n$ given θ, and which confirms that the observable quantities are conditionally independent.

- The true value θ_0 is defines in this case as

$$\theta_0 = \lim_{n \longrightarrow \infty} \frac{1}{n} \sum_{i=1}^n Y_i$$

  that is, the (almost sure) limit of the sample mean of the Y_is.

- $\pi_0(d\theta)$ is the prior distribution for θ, representing prior uncertainty over the value of θ_0.

- $p_{Y_1,\ldots,Y_n}(y_1, \ldots, y_n)$ is the prior predictive distribution, which is the presumed (i.e. modelled) joint distribution of $Y_1, \ldots, Y_n$.

- If $S_n = \sum_{i=1}^n Y_i$, then the implied distribution for S_n is

$$p_{S_n}(s) = \int_0^1 \theta^s(1-\theta)^{n-s}\pi_0(d\theta)$$

  and $\pi_0(d\theta)$ can be shown to be the limit of the distribution of the observable

$$T_n = \frac{1}{n} S_n$$

  as $n \longrightarrow \infty$.

(ii) Suppose that the prior distribution on θ is $Uniform(0,1)$. Find the *prior predictive distribution* for Y_1. 3 MARKS

Q1(a)(ii): We have by definition

$$p_{Y_1}(y_1) = \int_0^1 \theta^{y_1}(1-\theta)^{1-y_1}\, d\theta$$

which, by recognizing the integrand as the kernel of a $Beta(y_1+1, 1-y_1+1)$ distribution, we can immediately write for $y_1 = 0, 1$

$$p_{Y_1}(y_1) = \frac{\Gamma(y_1+1)\Gamma(1-y_1+1)}{\Gamma(3)} = \begin{cases} \frac{1}{2} & y_1 = 0 \\ \frac{1}{2} & y_1 = 1 \end{cases}$$

(iii) Find the *posterior predictive distribution* for Y_2 given $Y_1 = y_1$ under the $Uniform(0,1)$ prior.

3 MARKS

Q1(a)(iii): By de Finetti, for $(y_1, y_2) \in \{0,1\}^2$,

$$p_{Y_1,Y_2}(y_1,y_2) = \int_0^1 \theta^{y_1+y_2}(1-\theta)^{2-y_1-y_2}\, d\theta$$

which, by the same method as in (ii), yields

$$p_{Y_1,Y_2}(y_1,y_2) = \frac{\Gamma(y_1+y_2+1)\Gamma(2-y_1-y_2+1)}{\Gamma(4)} = \begin{cases} \frac{2}{6} & y_1=0, y_2=0 \\ \frac{1}{6} & y_1=1, y_2=0 \\ \frac{1}{6} & y_1=0, y_2=1 \\ \frac{2}{6} & y_1=1, y_2=1 \end{cases}$$

so that the posterior predictive is

$$p_{Y_2|Y_1}(y_2|y_1) = \frac{p_{Y_1,Y_2}(y_1,y_2)}{p_{Y_1}(y_1)} = \begin{cases} \frac{1}{3} & y_2=0 \\ \frac{2}{3} & y_2=1 \end{cases}$$

for $y_1 = 1$, and

$$p_{Y_2|Y_1}(y_2|y_1) = \frac{p_{Y_1,Y_2}(y_1,y_2)}{p_{Y_1}(y_1)} = \begin{cases} \frac{2}{3} & y_2=0 \\ \frac{1}{3} & y_2=1 \end{cases}$$

for $y_1 = 0$.

Note: typo, with the two cases reversed.

(b) Suppose that data $y_1, \ldots, y_n$ are observed values of random variables that are assumed to be conditionally independent $Poisson(\theta)$ random variables. Show that the posterior distribution for θ depends only on a summary, s_n, of the observed values, and identify the limiting behaviour of this posterior as $n \longrightarrow \infty$ if the data generating model is $Poisson(\theta_0)$ with $\theta_0 = 1$. 8 MARKS

Q1(b): Assuming a prior $\pi_0(\theta)$ is continuous on $\mathbb{R}^+$, we have that the posterior is

$$\pi_n(\theta) \propto \prod i = 1^n \theta^{y_i} \exp\{-\theta\}\pi_0(\theta)$$

which simplifies to

$$\pi_n(\theta) \propto \theta^{s_n} \exp\{-n\theta\}\pi_0(\theta)$$

where $s_n = \sum_{i=1}^{n} y_i$. Therefore the posterior depends only on s_n.

If the support of the prior includes θ_0, then the posterior distribution converges to a point mass at the limiting value of the statistic

$$\frac{1}{n}\sum_{i=1}^{n} Y_i$$

which, by the law of large numbers, is $\theta_0 = 1$. If the support of the prior does not contain θ_0, then the posterior converges to a point mass at some value θ_0^* in the support.

2. Random variables $Y_1, \ldots, Y_n$ are presumed conditionally independent with pdf

$$f_Y(y;\theta) = \mathbb{1}_{(\theta,\infty)}(y)\lambda \exp\{-\lambda(y-\theta)\} \qquad y \in R$$

where $\theta \in \Theta \equiv (-1, 1)$, and $\lambda > 0$.

(a) Find the form of the posterior distribution for θ if λ is a fixed constant, $\lambda = \lambda_0$, and the prior on θ is a Uniform distribution. 6 MARKS

Q2(a): The likelihood is

$$\prod_{i=1}^{n} \mathbb{1}_{(\theta,\infty)}(y_i)\lambda \exp\{-\lambda(y_i-\theta)\} = \mathbb{1}_{(-1,m_n)}(\theta)\lambda^n \exp\left\{-\lambda \sum_{i=1}^{n}(y_i-\theta)\right\}$$

where $m_n = \min\{y_1, \ldots, y_n\}$ as the product of indicators is only non-zero if $y_i > \theta$ for each i. Given that $\pi_0(\theta) \equiv Uniform(-1, 1)$, we have that

$$\pi_n(\theta) \propto \mathbb{1}_{(-1,t_n)}(\theta) \exp\{n\lambda\theta\}$$

where $t_n = \min\{m_n, 1\}$. Therefore

$$\pi_n(\theta) = \frac{\mathbb{1}_{(-1,t_n)}(\theta) \exp\{n\lambda\theta\}}{\displaystyle\int_{-1}^{t_n} \exp\{n\lambda u\}\, du} = \frac{\mathbb{1}_{(-1,t_n)}(\theta) \exp\{n\lambda\theta\}}{\dfrac{1}{n\lambda}\left[\exp\{n\lambda t_n\} - \exp\{-n\lambda\}\right]}$$

where we set $\lambda = \lambda_0$.

Note: in general, you should always be looking to compute the normalizing constant by performing the integration. In this case (and the subsequent cases in this question), all the integrals are straightforward in one step either by using the fact that the integrand is proportional to a pdf, or by elementary calculus.

(b) Find the form of the *joint posterior* for (θ, λ) if θ and λ have independent prior distributions, if the prior on θ is Uniform and the prior on λ is $Gamma(2, 1)$. 6 MARKS

Q2(b): For the joint posterior, the support is $(-1, t_n) \times \mathbb{R}^+$ and if $s_n = \sum_{i=1}^{n} y_i$, we have

$$\begin{aligned}\pi_n(\theta, \lambda) &\propto \lambda^n \exp\{-\lambda(s_n - n\theta)\} \times \lambda \exp\{-\lambda\} \\ &\propto \lambda^{n+1} \exp\{-\lambda(s_n + 1 - n\theta)\} \\ &= c\lambda^{n+1} \exp\{-\lambda(s_n + 1 - n\theta)\}\end{aligned}$$

say. To get the normalizing constant c, we first integrate out λ, and using the fact that the integrant is proportional to a Gamma pdf, we have

$$\pi_n(\theta) = c\frac{\Gamma(n+2)}{(s_n + 1 - n\theta)^{n+2}}$$

Therefore, by direct integration

$$\frac{1}{c} = \int_{-1}^{t_n} \frac{\Gamma(n+2)}{(s_n + 1 - n\theta)^{n+2}}\, d\theta = \frac{\Gamma(n+2)}{n(n+1)} \left[\frac{1}{(s_n + 1 - nt_n)^{n+1}} - \frac{1}{(s_n + 1 + n)^{n+1}}\right]$$

which yields c.

(c) From the joint posterior in (b), find the marginal posteriors for θ and λ. 8 MARKS

Q2(c): By the result above, we have

$$\pi_n(\theta) = c\frac{\Gamma(n+2)}{(s_n+1-n\theta)^{n+2}} \qquad \theta \in (-1, t_n).$$

For $\lambda > 0$,

$$\begin{aligned}
\pi_n(\lambda) &= c\lambda^{n+1}\int_{-1}^{t_n} \exp\{-\lambda(s_n+1-n\theta)\}\, d\theta \\
&= c\lambda^{n+1}\exp\{-\lambda(s_n+1)\}\int_{-1}^{t_n} \exp\{n\lambda\theta\}\, d\theta \\
&= c\lambda^{n+1}\exp\{-\lambda(s_n+1)\}\frac{1}{n\lambda}\left[\exp\{n\lambda t_n\} - \exp\{-n\lambda\}\right] \\
&= \frac{c}{n}\lambda^{n}\exp\{-\lambda(s_n+1)\}\left[\exp\{n\lambda t_n\} - \exp\{-n\lambda\}\right]
\end{aligned}$$

3. Suppose that $Y_1, \ldots, Y_n$ are presumed to be conditionally independent and distributed as $Normal(\mu, \sigma^2)$. A sample of data $y_1, \ldots, y_n$ are observed.

   (a) Derive the joint posterior for μ and σ^2 under a conjugate prior specification. 10 MARKS

Q3(a): Likelihood: using the usual decomposition

$$\mathcal{L}_n(\mu, \sigma^2) \propto \left(\frac{1}{\sigma^2}\right)^{n/2} \exp\left\{-\frac{1}{2\sigma^2}\left[n(\overline{y} - \theta)^2 + nv_n\right]\right\} \qquad v_n = \frac{1}{n}\sum_{i=1}^{n}(y_i - \overline{y})^2$$

Conjugate prior: $\pi_0(\mu, \sigma^2) = \pi_0(\sigma^2)\pi_0(\mu|\sigma^2)$, where

$$\pi_0(\sigma^2) \equiv InvGamma(a_0/2, b_0/2) \qquad \pi_0(\mu|\sigma^2) \equiv Normal(\eta_0, \sigma^2/\lambda_0)$$

say, where $a_0, b_0, \eta_0, \lambda_0$ are constants. Thus, up to proportionality, the joint posterior is

$$\left(\frac{1}{\sigma^2}\right)^{(n+a_0+1)/2+1} \exp\left\{-\frac{1}{2\sigma^2}\left[n(\overline{y} - \mu)^2 + nv_n + \lambda_0(\mu - \eta_0)^2 + b_0\right]\right\}$$

Using the complete the square formula

$$n(\overline{y} - \theta)^2 + \lambda_0(\theta - \eta_0)^2 = \lambda_n(\mu - m_n)^2 + d_n$$

where

$$m_n = \frac{n\overline{y} + \lambda_0\eta_0}{n + \lambda_0} \qquad \lambda_n = n + \lambda_0 \qquad d_n = \frac{n\lambda_0}{n + \lambda_0}(\overline{y} - \eta_0)^2.$$

Thus, up to proportionality, the joint posterior is

$$\left(\frac{1}{\sigma^2}\right)^{(n+a_0+1)/2+1} \exp\left\{-\frac{1}{2\sigma^2}\left[\lambda_n(\mu - m_n)^2 + d_n + nv_n + b_0\right]\right\}$$

from which we can deduce directly that

$$\pi_n(\mu|\sigma^2) \equiv Normal(m_n, \sigma^2/\lambda_n)$$

and on integrating out μ, using the fact that the integrand is a Normal pdf, we discover that

$$\pi_n(\sigma^2) \propto \left(\frac{1}{\sigma^2}\right)^{(n+a_0)/2+1} \exp\left\{-\frac{1}{2\sigma^2}\left[d_n + nv_n + b_0\right]\right\}$$

that

$$\pi_n(\sigma^2) \equiv InvGamma((n + a_0)/2, (d_n + nv_n + b_0)/2).$$

(b) For the posterior in (a), derive the Bayesian estimate of μ, $\widehat{\mu}_{nB}$ under *quadratic loss*, defined by

$$\widehat{\mu}_{nB} = \arg\min_t \int_0^\infty \int_{-\infty}^\infty L(t,\mu)\pi_n(\mu,\sigma^2)\, d\mu\, d\sigma^2$$

where $L(t,\mu) = (t-\mu)^2$. 5 MARKS

Q3(b): To perform the minimization, we may differentiate under the integrals, and deduce that we need to solve

$$\int_0^\infty \int_{-\infty}^\infty (t-\mu)\pi_n(\mu,\sigma^2)\, d\mu\, d\sigma^2 = 0.$$

Thus

$$\widehat{\mu}_{nB} = \mathbb{E}_{\pi_n(\mu,\sigma^2)}[\mu]$$

where the expectation is taken with respect to the joint posterior. However, by iterated expectation

$$\mathbb{E}_{\pi_n}[\mu] = \mathbb{E}_{\pi_n(\sigma^2)}\left[\mathbb{E}_{\pi_n(\mu|\sigma^2)}[\mu]\right]$$

where the interior expectation is simply m_n. Therefore, as this does not depend on σ^2, we have that

$$\widehat{\mu}_{nB} = m_n.$$

(c) Suppose that the prior for σ^2 is *degenerate* at $\sigma^2 = 1$. Give an expression for a 95 % credible interval for μ in terms of the standard Normal cdf, $\Phi(\cdot)$ and its inverse. 5 MARKS

Q3(c): With $\sigma^2 = 1$, the posterior for $\pi_n(\mu)$ is $Normal(m_n, 1/\lambda_n)$. Therefore a 95 % credible interval is any interval in $\mathbb{R}$ that contains probability 0.95 under the posterior. Therefore, for example, the interval (l_n, u_n)

$$l_n = m_n - \frac{\Phi^{-1}(0.975)}{\sqrt{\lambda_n}} \qquad u_n = m_n + \frac{\Phi^{-1}(0.975)}{\sqrt{\lambda_n}}$$

contains probability 0.95 in light of the data.

4. Suppose that $Y_1, \ldots, Y_n$ are conditionally independent discrete random variables with conditional mass function

$$f_Y(y; \boldsymbol{\theta}) = \theta_y \qquad y = 1, 2, \ldots, N$$

for integer $N > 0$, and zero otherwise, where for each y, $0 \leq \theta_y \leq 1$, and

$$\sum_{y=1}^{N} \theta_y = 1.$$

A sample of data $y_1, \ldots, y_n$ is collected.

(a) Write down the likelihood for this model. 4 MARKS

Q4(a): The likelihood is simply

$$\mathcal{L}_n(\boldsymbol{\theta}) = \prod_{k=1}^{N} \theta_k^{s_k}$$

where s_k is the number of times $y_i = k$, $k = 1, \ldots, N$.

(b) Compute the joint posterior for $(\theta_1, \ldots, \theta_N)$ under a *conjugate prior*. 6 MARKS

Q4(b): A conjugate prior here is

$$\pi_0(\boldsymbol{\theta}) \propto \prod_{k=1}^{N} \theta_k^{\alpha_k - 1}$$

that is, $Dirichlet(N-1; \alpha_1, \ldots, \alpha_N)$. Under this prior

$$\pi_n(\boldsymbol{\theta}) \propto \prod_{k=1}^{N} \theta_k^{s_k + \alpha_k - 1}$$

that is $Dirichlet(N-1; s_1 + \alpha_1, \ldots, s_N + \alpha_N)$.

(c) In a reparameterization of the model, suppose that $\phi_1 = \theta_1$ and, for $k = 2, \ldots, N$,

$$\phi_k = \frac{\theta_k}{1 - \theta_1 - \cdots - \theta_{k-1}}$$

that is

$$\phi_2 = \frac{\theta_2}{1 - \theta_1} \qquad \phi_3 = \frac{\theta_3}{1 - \theta_1 - \theta_2}$$

etc. Find a conjugate prior for the new parameterization, and the posterior under this prior.
10 MARKS

Q4(c): Here, we have for the inverse transformation, by direct computation in sequence

$$\begin{aligned}
\theta_1 &= \phi_1 \\
\theta_2 &= (1 - \phi_1)\phi_2 \\
\theta_3 &= (1 - \phi_1)(1 - \phi_2)\phi_3 \\
&\vdots \\
\theta_{N-1} &= (1 - \phi_1)(1 - \phi_2) \cdots (1 - \phi_{N-2})\phi_{N-1}
\end{aligned}$$

Note that, in this $N-1$ dimensional model, we have by taking out common factors sequentially that

$$\begin{aligned}
\theta_N &= (1 - \theta_1 - \theta_2 - \cdots - \theta_{N-1}) \\
&= (1 - \phi_1 - (1 - \phi_1)\phi_2 - (1 - \phi_1)(1 - \phi_2)\phi_3 - (1 - \phi_1)(1 - \phi_2) \cdots (1 - \phi_{N-2})\phi_{N-1}) \\
&= (1 - \phi_1)(1 - \phi_2)(1 - \phi_3) \cdots (1 - \phi_{N-1})
\end{aligned}$$

Therefore, the likelihood is

$$\phi_1^{s_1} \times \prod_{k=2}^{N-1} \left\{ \prod_{j=1}^{k-1} (1 - \phi_j) \right\}^{s_k} \phi_k^{s_k} \times \left\{ \prod_{k=1}^{N-1} (1 - \phi_k) \right\}^{n_N} = \prod_{k=1}^{N-1} \phi_k^{a_k} (1 - \phi_k)^{b_k}$$

where for $k = 1, 2, \ldots, N-1$, $a_k = s_k, b_k = \sum_{j=k+1}^{N} s_j$. A conjugate prior is therefore

$$\prod_{k=1}^{N-1} Beta(a_{0k}, b_{0k})$$

as the likelihood is the product of independent $Binomial(a_k + b_k, \phi_k)$ terms. Under this prior, the posterior is

$$\prod_{k=1}^{N-1} Beta(a_{0k} + a_k, b_{0k} + b_k)$$

that is, the product of independent Beta random variables.

Note: here, you are not asked to use the SAME prior as in (b) after transformation, you are asked to find a conjugate prior for the new parameterization. Therefore, you do not anywhere need to compute the Jacobian, as re-parameterizing in the likelihood part is not a re-parameterization in a pdf: the likelihood is not a pdf in $\boldsymbol{\theta}$.

5. (a) Consider the pdf

$$f(x) = c\mathbb{1}_{(0,1)}(x)x(1-x)(1-\sqrt{x})^2 \quad x \in \mathbb{R}$$

for some constant $c > 0$. It is proposed to estimate the expectation $\mathbb{E}_f[X]$ using Monte Carlo.

(i) Give details of the Monte Carlo procedure that uses *importance sampling* based on the proposal pdf $f_0(x) \equiv Uniform(0,1)$. 3 MARKS

Q5(a)(i): The integral we are after is

$$\int_0^1 xf(x)\,dx = c\int_0^1 x^2(1-x)(1-\sqrt{x})^2\,dx$$

which can in fact be computed analytically.

If we take $g_1(x) = x^2(1-x)(1-\sqrt{x})^2$, then we can write the integral

$$\int_0^1 g_1(x)f_0(x)\,dx$$

and estimate it by

$$\frac{1}{N}\sum_{i=1}^{N} g_1(X_i) \qquad X_1,\ldots,X_n \sim f_0$$

However, we also need to estimate c, which we can do by noting that

$$c^{-1} = \int_0^1 g_0(x)\,dx$$

where

$$g_0(x) = x(1-x)(1-\sqrt{x})^2.$$

Hence

$$\widehat{c} = \left\{\frac{1}{N}\sum_{i=1}^{N} g_0(X_i)\right\}^{-1}$$

Thus

$$\widehat{\mathbb{E}}_f[X] = \frac{\sum_{i=1}^{N} g_1(X_i)}{\sum_{i=1}^{N} g_0(X_i)}.$$

Note: remember that here $f_0(x) = 1$ on $0 < x < 1$, so

$$\frac{xf(x)}{f_0(x)} = xf(x) = cg_1(x) \quad \text{and} \quad \frac{f(x)}{f_0(x)} = f(x) = cg_0(x).$$

(ii) Give details of a Monte Carlo procedure that uses *importance sampling* based on a proposal $f_0(x)$ designed to produce an estimator whose variance is smaller than the variance of the Monte Carlo estimator from (a) (you do not have to verify that the variance is smaller). 4 MARKS

Q5(a)(ii): The optimal proposal is

$$f_0^\star(x) \propto x f(x) \propto x^2(1-x)(1-\sqrt{x})^2$$

so making the approximation $\sqrt{x} \doteq x$, we may try $f_{01}(x) \equiv Beta(3,4)$. To get the normalizing constant, we need to integrate the unnormalized $f(x)$, and to do that we would use $f_{02}(x) \equiv Beta(2,4)$, by the same principle. Thus we sample $X_{11}, \ldots, X_{1N} \sim f_{01}(x)$ and $X_{21}, \ldots, X_{1N} \sim f_{02}(x)$, and use the estimator

$$\widehat{E}_f[X] = \frac{\sum_{i=1}^{N} g_1(X_{1i})/f_{01}(X_{1i})}{\sum_{i=1}^{N} g_0(X_{2i})/f_{02}(X_{2i})}$$

Note: you could also add in an antithetic variable step if you decide to use a symmetric proposal on $(0,1)$, that is say $Beta(\alpha,\alpha)$ for some α, as if $X_1 \sim Beta(\alpha,\alpha)$, then also $1 - X_1 \sim Beta(\alpha,\alpha)$. This should further reduce the variance. However, the main purpose of IS is to induce variance reduction in the first place.

(iii) Describe how to carry out *rejection sampling* from $f(x)$ based on proposals from a $Beta(2,2)$ distribution. 3 MARKS

Q5(a)(iii): The $Beta(2,2)$ pdf is

$$f_0(x) = \frac{\Gamma(4)}{\Gamma(2)\Gamma(2)} x(1-x) = 6x(1-x) \qquad 0 < x < 1.$$

Our first task is to bound

$$\frac{f(x)}{f_0(x)} = \frac{cx(1-x)(1-\sqrt{x})^2}{6x(1-x)} = \frac{c}{6}(1-\sqrt{x})^2 < \frac{c}{6}$$

as the ratio is monotone decreasing on $(0,1)$. Thus the ratio is clearly bounded, but the bound is indeterminate as c is unknown. This does not affect the algorithm, however, and we can implement it with the *unnormalized* densities, as

$$x(1-x)(1-\sqrt{x})^2 \leq x(1-x) \quad 0 < x < 1$$

Thus the algorithm is

(i) sample $U \sim Uniform(0,1)$ and $X \sim Beta(2,2)$

(ii) if

$$U \leq (1-\sqrt{x})^2$$

accept X as a draw from $f(x)$; if not, reject and return to 1.

(iii) repeat until the required Monte Carlo sample size is met.

(b) Consider a linear regression model for observations $y_1, \ldots, y_n$ that are considered conditionally independent with

$$Y_i|x_i; \beta_0, \beta_1, \sigma^2 \sim Normal(\beta_0 + \beta_1 x_i, \sigma^2)$$

for $i = 1, \ldots, n$. In a Bayesian analysis, an *independence* prior

$$\pi_0(\beta_0, \beta_1, \sigma^2) = \pi_0(\beta_0)\pi_0(\beta_1)\pi_0(\sigma^2)$$

is specified, where

$$\pi_0(\beta_0) \equiv \pi_0(\beta_1) \equiv Normal(0, M_0) \qquad \pi_0(\sigma^2) \equiv InvGamma(a_0/2, b_0/2)$$

for hyperparameters M_0, a_0, b_0.

(i) Describe how to implement a *Gibbs sampler* for this model, giving details of the full conditional posterior distributions. 6 MARKS

Q5(b)(i): The Gibbs sampler samples recursively from the three full conditional posteriors

$$\pi_n(\beta_0|\beta_1, \sigma^2) \qquad \pi_n(\beta_1|\beta_2, \sigma^2) \qquad \pi_n(\sigma^2|beta_0, \beta_1)$$

and in this case all of the distributions are available in closed form. To see this: the likelihood is proportional to

$$\left(\frac{1}{\sigma^2}\right)^{n/2} \exp\left\{-\frac{1}{2\sigma^2}(y_i - \beta_0 - \beta_1 x_i)^2\right\}$$

which reveals immediately that $\pi_n(\sigma^2|beta_0, \beta_1) \equiv InvGamma(a_n/2, b_n/2$ where

$$a_n = n + a_0 \qquad b_n = \sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2 + b_0.$$

For $\pi_n(\beta_0|\beta_1, \sigma^2)$, we note the decomposition

$$\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2 = \sum_{i=1}^{n}(y_i - \overline{d}_1)^2 + n(\beta_0 - \overline{d}_0)^2$$

where

$$\overline{d}_0 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \beta_1 x_i).$$

Therefore,

$$\pi_n(\beta_0|\beta_1, \sigma^2) \propto \exp\left\{-\frac{1}{2}\left[\frac{n}{\sigma^2}(\beta_0 - \overline{d}_0)^2 + \frac{1}{M_0}\beta_0^2\right]\right\}$$

and using the complete-the-square formula, we have that

Q5(b)(i) continued:

$$\pi_n(\beta_0|\beta_1,\sigma^2) \equiv Normal(m_{n0}, M_{n0})$$

where

$$m_{n0} = \frac{n\overline{d}_0/\sigma^2}{n/\sigma^2 + 1/M_0} \qquad M_{n0} = \left(\frac{n}{\sigma^2} + \frac{1}{M_0}\right)^{-1}.$$

Finally for $\pi_n(\beta_1|\beta_2,\sigma^2)$, we have by the same approach

$$\pi_n(\beta_1|\beta_2,\sigma^2) \propto \exp\left\{-\frac{1}{2}\left[\frac{v_{n1}}{\sigma^2}(\beta_1 - \overline{d}_1)^2 + \frac{1}{M_0}\beta_1^2\right]\right\}$$

where

$$v_{n1} = \sum_{i=1}^{n} x_i^2 \qquad \overline{d}_2 = \frac{1}{n}\sum_{i=1}^{n}\frac{y_i - \beta_0}{x_i}$$

which follows as

$$(y_i - \beta_0 - \beta_1 x_i)^2 = x_i^2(\beta_1 - (y_i - \beta_0)/x_i)^2.$$

Therefore

$$\pi_n(\beta_1|\beta_0,\sigma^2) \equiv Normal(m_{n1}, M_{n1})$$

where

$$m_{n1} = \frac{v_{n1}\overline{d}_1/\sigma^2}{v_{n1}/\sigma^2 + 1/M_0} \qquad M_{n1} = \left(\frac{v_{n1}}{\sigma^2} + \frac{1}{M_0}\right)^{-1}.$$

(ii) Suppose for $k = 1, 2$, independent priors

$$\pi_0(\beta_k) = \frac{1}{2}\exp\{-|\beta_k|\} \quad \beta_k \in \mathbb{R}$$

are used. Describe how to implement a *Metropolis-within-Gibbs* MCMC algorithm for this model.
4 MARKS

Q5(b)(iii): With this new prior specification, the full conditional posteriors for β_0 and β_1 are not available analytically, but we may use the Metropolis-Hastings (MH) algorithm to propose and accept or reject new points to continue the updating for these parameters. Here is the general strategy: we have

$$\pi_n(\beta_0|\beta_1, \sigma^2) \propto \mathcal{L}_n(\beta_0, \beta_1, \beta_2)\pi_0(\beta_0)$$

which we are able to compute pointwise (up to proportionality). For the Metropolis-within-Gibbs, we initialize the chain at some starting values, and then update the parameters in turn as follows.

(i) Current point at iteration t is $\beta_0^{(t)}$

(ii) Propose β from a proposal density $q(\beta_0^{(t)}, \cdot)$ which ensures the theoretical properties hold; choosing $q(\beta_0^{(t)}, \cdot) \equiv Normal(\beta_0^{(t)}, \sigma_q^2)$ is sufficient.

(iii) Set $\beta_0^{(t+1)} = \beta$ with probability

$$\alpha_{0t} = \min\left\{1, \frac{\pi_n(\beta|\beta_1, \sigma^2)q(\beta, \beta_0^{(t)})}{\pi_n(\beta_0^{(t)}|\beta_1, \sigma^2)q(\beta_0^{(t)}, \beta)}\right\}$$

otherwise set $\beta_0^{(t+1)} = \beta_0^{(t)}$. We perform the acceptance step by generating $U \sim Uniform(0, 1)$, and accepting if $U < \alpha_{0t}$.

The same algorithm is used to update β_1 from its full conditional. However, the full conditional for σ^2 is still available analytically, so can be updated using direct simulation. We cycle through updating the three parameters in turn.

Note: typo in the acceptance probability formula.