Data provided: Tables of distribution functions

MAS464

The University Of Sheffield.

SCHOOL OF MATHEMATICS AND STATISTICS **Autumn Semester 2017–18**

Bayesian Statistics **2 hours**

Candidates may bring to the examination a calculator which conforms to University regulations.
Marks will be awarded for your best ***three*** *answers. Total marks 84.*

Standard results from the lecture notes may be used without derivation, but must be clearly stated.

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 A precision weighing device yields unbiased measurements within half a gramme, which can be modelled as $\text{Un}(x \mid \theta - 1/2, \theta + 1/2)$, where θ is the unknown weight. A priori, it is believed $\theta \sim \text{Un}(\theta \mid 10, 20)$.

(i) Find the posterior distribution of θ if a single measurement, $x = 12$, is made. *(7 marks)*

(ii) Using $\boldsymbol{x} = \{11, 11.5, 11.7, 11.1, 11.4, 10.9\}$, a different set of six independent measurements:

(a) Find the posterior distribution of θ. *(10 marks)*

(b) Show that the posterior mean and variance are 11.3 and 0.003, respectively. *(4 marks)*

(c) Provide an equally tailed posterior interval of probability 0.95 and explain why this is a HPD interval. *(7 marks)*

2 Assume that the waiting time, t, of a client in a bank can be modelled with an exponential distribution with unknown parameter λ,

$$f(t \mid \lambda) = \lambda \, \exp[-\lambda\, t] \;, \quad \lambda > 0 \;.$$

and that the prior distribution is Gamma with parameters (a, b):

$$\pi(\lambda) = \frac{b^a}{\Gamma[a]} \, \lambda^{a-1} \exp[-b\,\lambda] \;; \quad a, b > 0 \;.$$

(i) Find the prior parameters if we believe $\mathbb{E}[\lambda] = 0.2$ and $\mathbb{V}[\lambda] = 1$. *(3 marks)*

(ii) An average waiting time, $\bar{t} = 3.8$, is recorded from observing 20 clients at random. Show that the prior is conjugate and provide the posterior parameters. *(7 marks)*

(iii) The coefficient of variation of a random quantity with nonzero mean, μ and standard deviation $\sigma > 0$ is defined as σ/μ. What is the smallest sample size required to reduce the posterior coefficient of variation to 0.1? *(8 marks)*

(iv) Explain why the highest predictive probability interval of the waiting time for a randomly chosen new client is of the form $(0, c)$ and show that $c = 12.286$. *(10 marks)*

3 Assume $\boldsymbol{x} = \{x_1, \ldots, x_n\}$ is a random sample from a Gaussian distribution with mean μ and precision τ, both unknown.

(i) Elicit the parameters of a Normal-Gamma distribution for μ and τ, consistent with the following prior beliefs,

$$\mathbb{E}[\tau] = 1\ , \quad \mathbb{V}[\tau] = \frac{1}{3}\ , \quad \mathrm{P}[\mu > 3] = \frac{1}{2}\ , \quad \mathrm{P}[\mu > 0.12] = 0.9\ .$$

Hint: Let t_n follow a standard Student-t distribution with n degrees of freedom, then $\mathrm{P}[t_4 > -1.533] = 0.9$, $\mathrm{P}[t_6 > -1.440] = 0.9$, $\mathrm{P}[t_8 > -1.397] = 0.9$. ***(8 marks)***

(ii) From a random sample of size 8, $\sum_{i=1}^{8} x_i = 16$ and $\sum_{i=1}^{8} x_i^2 = 48$, were recorded.

(a) Determine the HPD interval of probability 0.99 for the mean.
Hint: Let t_n follow a standard Student-t distribution with n degrees of freedom, then $\mathrm{P}[t_8 < 3.355] = 0.995$, $\mathrm{P}[t_{14} < 2.977] = 0.995$, $\mathrm{P}[t_{16} < 2.921] = 0.995$. ***(10 marks)***

(b) Calculate the posterior Bayes estimate of the precision using a square loss function. ***(4 marks)***

(c) Calculate the posterior Bayes estimate of the variance, $\sigma^2 = 1/\tau$, using a 0–1 loss function. ***(6 marks)***

4 Consider the regression model,

$$y_i = \alpha_i + \beta x_i + \varepsilon_i \ ; \quad i = 1, \ldots, n$$

with $\varepsilon_i \sim \mathrm{N}(\varepsilon_i \,|\, 0, 1/\lambda)$, i.i.d., and prior structure

$$\alpha_i \sim \mathrm{N}(\alpha_i \,|\, \mu, 1/p) \ ; \quad \text{independent for } i = 1, \ldots, n$$
$$\mu \sim \mathrm{N}(\mu \,|\, a, 1/r) \ , \quad \beta \sim \mathrm{N}(\beta \,|\, b, 1/q) \quad \text{and} \quad \lambda \sim \mathrm{Ga}(\lambda \,|\, c, d)$$

(i) Show that the full conditional of:

(a) Each of the individual intercepts, α_i, is Gaussian and provide explicit expressions for the parameters. *(5 marks)*

(b) The mean intercept, μ, is Gaussian and provide explicit expressions for the parameters. *(5 marks)*

(c) The regression slope, β, is Gaussian and provide explicit expressions for the parameters. *(5 marks)*

(d) The regression precision, λ, is Gamma and provide explicit expressions for the parameters. *(5 marks)*

(ii) Write pseudo-code for an MCMC sampling scheme for exploring the posterior distribution. *(8 marks)*

End of Question Paper

Notation and distributions

Bayesian Statistics 2017–18

Throughout the course it is assumed that the probabilistic behaviour of available data, $\boldsymbol{x}$, is described by a parametric model; hence all inferences will be conditional to the selected model.

Each model is composed by a family of probability distributions, indexed by a parameter vector, $\boldsymbol{\theta}$, which in turn can be described by their appropriate density functions. We will denote a specific model by

$$\mathcal{M} = \{f(\boldsymbol{x} \mid \boldsymbol{\theta}),\ \boldsymbol{x} \in \mathcal{X},\ \boldsymbol{\theta} \in \Theta\},$$

where $f(\boldsymbol{x} \mid \boldsymbol{\theta}) \geq 0$ and $\int_{\mathcal{X}} f(\boldsymbol{x} \mid \boldsymbol{\phi})\, \mathrm{d}\boldsymbol{x} = 1$; when there is no risk of confusion, we will refer to a model simply as $f(\boldsymbol{x} \mid \boldsymbol{\theta})$. We call $\mathcal{X}$ the support of the distribution and Θ the parameter space.

We will use $f(\boldsymbol{x} \mid \boldsymbol{\phi})$ and $f(\boldsymbol{y} \mid \boldsymbol{\psi})$ to refer to probability densities of $\boldsymbol{x}$ and $\boldsymbol{y}$, without necessarily meaning that both quantities share a common distribution. In general, the Greek alphabet is reserved for non-observables (typically, parameters) and the Latin alphabet for observations (data). Bold typeface denotes vector valued quantities.

Specific density functions are referred by appropriate names; e.g. if the observable x follows a Gaussian distribution with mean μ and variance σ^2, its density is denoted by $\mathrm{N}(x \mid \mu, \sigma^2)$. Tables below present some density functions used throughout the course.

Moments and other descriptive measures of probability distributions are described by appropriate symbols. Thus,

$$\begin{aligned}
\mathbb{E}[\boldsymbol{x} \mid \boldsymbol{\theta}] &= \int_{\mathcal{X}} \boldsymbol{x}\ f(\boldsymbol{x} \mid \boldsymbol{\theta})\, \mathrm{d}\boldsymbol{x}, \\
\mathbb{V}[x \mid \boldsymbol{\theta}] &= \int_{\mathcal{X}} (x - \mathbb{E}[x \mid \boldsymbol{\theta}])^2\ f(x \mid \boldsymbol{\theta})\, \mathrm{d}x, \\
\mathrm{Cov}[\boldsymbol{x} \mid \boldsymbol{\theta}] &= \int_{\mathcal{X}} (\boldsymbol{x} - \mathbb{E}[\boldsymbol{x} \mid \boldsymbol{\theta}])^t (\boldsymbol{x} - \mathbb{E}[\boldsymbol{x} \mid \boldsymbol{\theta}]) f(\boldsymbol{x} \mid \boldsymbol{\theta})\, \mathrm{d}\boldsymbol{x}\ ,
\end{aligned}$$

respectively stand for the expected value, variance and covariance of the given quantity, while $\mathrm{Med}[x \mid \theta]$ and $\mathrm{Mode}[\boldsymbol{x} \mid \boldsymbol{\theta}]$ denote the median and mode, respectively. Sums are used instead of integrals when the support of the random quantity is discrete.

We use, $\boldsymbol{t} = \boldsymbol{t}(\boldsymbol{x})$ to denote a generic statistic (typically sufficient) derived from observed data, $\boldsymbol{x} = \{x_1, \ldots, x_n\}$; standard symbols are used for common statistics; thus,

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \quad \text{and} \quad s_x^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

denote the sample mean and variance, respectively; while $x_{(p)}$ stands for the p^{th} order statistic; in particular $x_{(1)}$ and $x_{(n)}$ respectively denote the minimum and maximum observed values.

SOME DISCRETE DISTRIBUTIONS

Name	**Context**	**Notation**	**p.f.** $p(x \mid \theta)$	$\mathbb{E}[X \mid \theta]$	$\mathbb{V}[X \mid \theta]$	**Applications**	**Comments**
Uniform	Set of k equally likely outcomes (usually, not necessarily, the integers)	$\mathrm{U}(1, \ldots, k)$	$p(x) = 1/k$ $\mathcal{X} = \{1, \ldots, k\}, \quad \mathcal{K} = \mathbb{Z}_+$	$\frac{k+1}{2}$	$\frac{k^2-1}{12}$	Dice	
Bernoulli	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	$\mathrm{Bi}(x \mid n, \theta) \approx \mathrm{Po}(x \mid n\theta)$ if n large, θ small, and $n\theta = c$.

SOME CONTINUOUS DISTRIBUTIONS

Name	**Notation**	**p.d.f.** $f(x \mid \theta)$	$\mathbb{E}[X \mid \boldsymbol{\theta}]$	$\mathbb{V}[X \mid \boldsymbol{\theta}]$	**Applications**	**Comments**
Uniform	$\mathrm{Un}(x \mid \alpha, \beta)$	$f(x) = \frac{1}{\beta - \alpha}$ $\mathcal{X} = [\alpha, \beta]$ $\Theta = \{(\alpha, \beta) \in \mathbb{R}^2 \,:\, \alpha < \beta\}$	$\frac{\alpha + \beta}{2}$	$\frac{(\beta - \alpha)^2}{12}$	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. Conjugate prior for exponential model.	Also parameterised in terms of $1/\beta$ $\mathrm{Ga}(x \mid 1, \lambda) \equiv \mathrm{Ex}(x \mid \lambda)$, $\mathrm{Ga}(x \mid \nu/2, 1/2) \equiv \chi^2_{(\nu)}(x)$ $1/x = y \sim \mathrm{IGa}(y \mid \alpha, \beta)$
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. Conjugate prior for Binomial model.	$\mathrm{Be}(x \mid 1, 1) \equiv \mathrm{Un}(x \mid 0, 1)$ $\mathrm{Be}(x \mid \alpha, \beta)$ is reflection about $\frac{1}{2}$ of $\mathrm{Be}(x \mid \beta, \alpha)$. Can re-scale $\mathrm{Be}(x \mid \alpha, \beta)$ to any finite range $[a, b]$ by $Y = (b - a)X + a$
Gaussian (Normal)	$\mathrm{N}(x \mid \mu, \sigma^2)$	$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x - \mu}{\sigma}\right)^2\right]$ $\mathcal{X} = \mathbb{R}$ $\Theta = \{(\mu, \sigma^2) \in \mathbb{R}^2 \,:\, \sigma^2 > 0\}$	μ	σ^2	Empirically and theoretically (via CLT) a useful model. 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	$\chi^2_{(\nu)}(x)$	$f(x) = \frac{2^{-\nu/2}}{\Gamma(\nu/2)} x^{\nu/2 - 1} e^{-x/2}$ $\mathcal{X} = \mathbb{R}_+$; $\Theta = \mathbb{R}_+$	ν	2ν	Sum of squares of ν independent standard Gaussians	$\chi^2_{(\nu)}(x) \equiv \mathrm{Ga}(x \mid \nu/2, 1/2)$
Student t	$\mathrm{St}(x \mid \mu, \lambda, \nu)$	$f(x) = \frac{\Gamma[(\nu+1)/2]}{\Gamma[\nu/2]}\left(\frac{\lambda}{\nu\pi}\right)^{1/2} \times$ $\left(1 + \lambda(x - \mu)^2/\nu\right)^{-(\nu+1)/2}$ $\mathcal{X} = \mathbb{R}$, $\mu \in \mathbb{R}, \lambda, \nu > 0$	μ (if $\nu > 1$)	$\lambda^{-1}\frac{\nu}{\nu - 2}$ (if $\nu > 2$)	Useful alternative to Gaussian for variables with heavy tails.	If $X \sim \mathrm{N}(x \mid 0, 1)$ and $Y \sim \chi^2_{(\nu)}(y)$ independent then $\frac{X}{\sqrt{Y/\nu}} \sim \mathrm{t}_\nu$. If $Y = \sqrt{\lambda}(x - \mu)$ then $Y \sim t_\nu(y)$ $\mathrm{t}_1 \equiv$ Cauchy. $\mathrm{t}^2_\nu \equiv \mathrm{F}_{1,\nu}$.

SOME MULTIVARIATE DISTRIBUTIONS

Name	Notation	p.d.f. $f(\boldsymbol{x} \mid \boldsymbol{\theta})$	$\mathbb{E}[\boldsymbol{X} \mid \boldsymbol{\theta}]$	$\mathbb{V}[\boldsymbol{X} \mid \boldsymbol{\theta}]$	Applications	Comments
Multinomial	$\mathrm{Mu}(\boldsymbol{x} \mid \boldsymbol{\theta}, n)$	$p(\boldsymbol{x}) = \frac{n!}{\prod_{l=1}^{k} x_l!} \prod_{l=1}^{k} \theta_l^{x_l}$ $\boldsymbol{x} = \{x_1, \ldots, x_k\}$, $x_l = 0, 1, \ldots$, $\sum x_l = n$ $\boldsymbol{\theta} = \{\theta_1, \ldots, \theta_k\}$, $0 < \theta_l < 1$, $\sum \theta_l = 1$	$\mathbb{E}[x_i] = n\theta_i$	$\mathbb{V}[x_i] = n\theta_i(1-\theta_i)$ $\mathrm{Cov}[x_i, x_j] = -n\theta_i\theta_j$	Counts of events with more that two possible outcomes	Generalisation of the Binomial distribution
Dirichlet	$\mathrm{Di}(\boldsymbol{x} \mid \boldsymbol{\alpha})$	$f(\boldsymbol{x}) = \frac{\Gamma(\sum \alpha_l)}{\prod \Gamma(\alpha_l)} \prod_{l=1}^{k} x_l^{\alpha_l - 1}$ $\boldsymbol{x} = \{x_1, \ldots, x_k\}$, $0 < x_l < 1$, $\sum_{l=1}^{k} x_l = 1$ $\boldsymbol{\alpha} = \{\alpha_1, \ldots, \alpha_k\}$, $0 < \alpha_l$	$\mathbb{E}[x_i] = \mu_i = \frac{\alpha_i}{\sum \alpha_l}$	$\mathbb{V}[x_i] = \frac{\mu_i(1-\mu_i)}{1 + \sum \alpha_l}$ $\mathrm{Cov}[x_i, x_j] = -\frac{\mu_i \mu_j}{1 + \sum \alpha_l}$	Distribution of points in a simplex	Generalisation of the Beta distribution
Normal-Gamma	$\mathrm{NG}(x, y \mid \mu, \lambda, \alpha, \beta)$	$f(x, y) = \mathrm{N}\big(x \mid \mu, (y\lambda)^{-1}\big)\mathrm{Ga}(y \mid \alpha, \beta)$ $\mathcal{X} = \{(x, y) \,:\, x \in \mathbb{R}, y > 0\}$ $\mu \in \mathbb{R}$; $\lambda, \alpha, \beta > 0$	$\mathbb{E}[x] = \mu$ $\mathbb{E}[y] = \alpha\beta^{-1}$	$\mathbb{V}[x] = \frac{\beta}{\lambda(\alpha - 1)}$ $\mathbb{V}[y] = \alpha\beta^{-2}$	Conjugate prior for Gaussian data	$f(x) = \mathrm{St}\big(x \mid \mu, \lambda\alpha\beta^{-1}, 2\alpha\big)$
(Multivariate) Gaussian	$\mathrm{N}_k(\boldsymbol{x} \mid \boldsymbol{\mu}, \Lambda)$	$f(\boldsymbol{x}) = \frac{\lvert\Lambda\rvert^{1/2}}{(2\pi)^{k/2}} \exp\big[-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})' \Lambda (\boldsymbol{x} - \boldsymbol{\mu})\big]$ $\mathcal{X} = \boldsymbol{x} \in \mathbb{R}^k$ $\boldsymbol{\mu} \in \mathbb{R}^k$; Λ symmetric positive-definite	$\boldsymbol{\mu}$	Λ^{-1}	See univariate case	Usually parameterised in terms of the covariance matrix $\Sigma = \Lambda^{-1}$
(Multivariate) Student	$\mathrm{St_k}(\boldsymbol{x} \mid \boldsymbol{\mu}, \Lambda, \nu)$	$f(\boldsymbol{x}) = \frac{\lvert\Lambda\rvert^{1/2}\Gamma((\nu + k)/2)}{(\nu\pi)^{k/2}\Gamma(\nu/2)} \times \left[1 + \frac{1}{\nu}(\boldsymbol{x} - \boldsymbol{\mu})' \Lambda (\boldsymbol{x} - \boldsymbol{\mu})\right]^{-(\nu + k)/2}$ $\mathcal{X} = \boldsymbol{x} \in \mathbb{R}^k$ $\boldsymbol{\mu} \in \mathbb{R}^k$; Λ symmetric positive-definite , $\nu > 0$	$\boldsymbol{\mu}$ (if $\nu > 1$)	$\frac{\nu}{\nu - 2}\Lambda^{-1}$ (if $\nu > 2$)	See univariate case	Usually parameterised in terms of the covariance matrix $\Sigma = \Lambda^{-1}$