Data provided: Tables of distribution functions

The University Of Sheffield.

MAS364

SCHOOL OF MATHEMATICS AND STATISTICS **Autumn Semester 2015–16**

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 In microscopic imaging it is common to model the number of photons arriving at the lens in each frame, X, as $\mathrm{Po}(x \mid \lambda)$, where λ is the rate of photon emission per frame. Given a random sample, $\boldsymbol{x} = \{x_1, \ldots, x_n\}$,

(i) (a) Show that $\pi(\lambda) = \mathrm{Ga}(\lambda \,|\, a, b)$ is a conjugate prior and give explicit expressions for the posterior parameters. ***(5 marks)***

(b) Find the Bayes estimator for λ under 0-1 loss. ***(3 marks)***

(ii) (a) Calculate the predictive distribution of Y, the number of photons captured by the lens in the next random sample of m frames,

$$Y = \sum_{j=n+1}^{n+m} X_j$$

(11 marks)

(b) The scientist a priori believes that $\mathbb{E}[\lambda] = 10/3$ and $\mathbb{V}[\lambda] = 50/9$. Calculate the scientist's probability of observing not more than one photon in the next frame if 3 photons were detected in a sample of $n = 10$ frames. ***(9 marks)***

2 Assume

$$X_i \sim \mathrm{N}\left(x_i \,\middle|\, \mu, \frac{1}{a_i \lambda} \right) ,$$

independent for $i = 1, \ldots, n$, where $\boldsymbol{a} = \{a_1, \ldots, a_n\}$ are known constants with $0 < a_i < 1$ and $\sum_{i=1}^{n} a_i = 1$.

(i) Show that

$$\pi(\mu, \lambda) = \mathrm{N}\left(\mu \,\middle|\, m, \frac{1}{p\lambda} \right) \mathrm{Ga}(\lambda \,|\, a, b)$$

is a conjugate prior and provide explicit expressions for the posterior parameters. ***(15 marks)***

(ii) Show that

$$\mathbb{E}[\mu \mid \boldsymbol{x}] = w\widehat{\mu} + (1 - w)m ,$$

where $0 < w < 1$ and $\widehat{\mu} = \sum_{i=1}^{n} a_i x_i$ is the MLE. ***(5 marks)***

(iii) Find the posterior Bayes estimator of $\sigma^2 = \lambda^{-1}$ under quadratic loss. ***(8 marks)***

3 A chemist is interested in the (relative) molecular weight of a new compound. She sends samples to n different labs and collects the measurements $\boldsymbol{W} = \{W_1, \ldots, W_n\}$. Given that each lab has a different weighing instrument, she thinks it is sensible to assume $W_i \sim \mathrm{N}(w_i \mid \mu, 1/\lambda_i)$, where μ is the actual weight and $\boldsymbol{\lambda} = \{\lambda_1, \ldots, \lambda_n\}$ are the known measuring precisions.

[*Additional information:* Assume $Z \sim \mathrm{N}(z \mid 0, 1)$ and let $\Phi(x) = \mathrm{P}[Z < x]$, then $\Phi(-1.96) = 0.025$, $\Phi(-1.645) = 0.05$, $\Phi(-1.26) = 0.104$ and $\Phi(1.521) = 0.936$.]

(i) Write down the likelihood and show that

$$\widehat{\mu} = \frac{\sum \lambda_i w_i}{\sum \lambda_i}$$

is the MLE. ***(7 marks)***

(ii) From previous stoichiometry analyses she believes $\mathrm{P}[\mu > 0.1] = 0.5$ and $\mathrm{P}[0.02 < \mu < 0.18] = 0.9$ and is willing to use a Gaussian distribution to express her uncertainty.

(a) Use the scientist's prior opinions to elicit her prior. ***(7 marks)***

(b) The new compound could potentially be used in drug production if its molecular weight is within $(0.1, 0.2)$, but it could be risky to use otherwise. After consultation, she thinks that her preferences can be described by

$$\mathcal{L}(a_1, \mu) = \begin{cases} 0 & \mu \in (0.1, 0.2) \\ 9 & \mu \notin (0.1, 0.2) \end{cases} \qquad \mathcal{L}(a_2, \mu) = \begin{cases} 2 & \mu \in (0.1, 0.2) \\ 0 & \mu \notin (0.1, 0.2) \end{cases}$$

where $a_1 =$*use the compound* and $a_2 =$*do not use the compound*. Find her optimal decision if $\widehat{\mu} = 0.2$ and $\sum \lambda_i = 350$ are recorded from a random sample of size $n = 100$. ***(14 marks)***

4 Consider the hierarchical model,

$$X_i \sim \mathrm{Ber}(x_i \mid \theta_i)\ ,\ \text{ind. } i = 1, \ldots, n$$
$$\pi(\theta_i) = \mathrm{Be}(\theta_i \mid a, a)\ ,\ \text{ind. } i = 1, \ldots, n$$
$$\pi(a) = \mathrm{Ga}(a \mid c, d)\ ,\ \text{with } \mathbb{E}[a] = \frac{c}{d}\ .$$

(i) Write down the full conditional distributions for $\boldsymbol{\theta} = \{\theta_1, \ldots, \theta_n\}$ and a. ***(13 marks)***

(ii) Write pseudo-code for a Metropolis-within-Gibbs strategy to sample from $\pi(\boldsymbol{\theta}, a \mid \boldsymbol{x})$. ***(15 marks)***

End of Question Paper

Notation and distributions

Bayesian Statistics 2015–16

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 Normal distribution with mean μ and variance σ^2, its density is denoted by $\mathrm{N}\big(x \,\big|\, \mu, \sigma^2\big)$. 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)	$\mathrm{U}(1,\dots,k)$	$p(x) = 1/k$ $\mathcal{X} = \{1,\dots,k\}, \quad \mathcal{K} = \mathbb{Z}_+$	$\dfrac{k+1}{2}$	$\dfrac{k^2-1}{12}$	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	$\mathrm{Bi}(x \mid n, \theta)$	$p(x) = \binom{n}{x}\theta^x(1-\theta)^{n-x}$ $\mathcal{X} = \{0,1,2,\dots,n\}$ $\Theta = (0,1)$	$n\theta$	$n\theta(1-\theta)$	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	$\mathrm{Ge}(x \mid \theta)$	$p(x) = \theta(1-\theta)^x$ $\mathcal{X} = 0,1,2,\dots$ $\Theta = (0,1)$	$\dfrac{1-\theta}{\theta}$	$\dfrac{1-\theta}{\theta^2}$	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	$\mathrm{NB}(x \mid m, \theta)$	$p(x) = \binom{m+x-1}{x}\theta^m(1-\theta)^x$ $\mathcal{X} = 0,1,2,\dots$ $\Theta = (0,1)$	$\dfrac{m(1-\theta)}{\theta}$	$\dfrac{m(1-\theta)}{\theta^2}$	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	$\mathrm{Po}(x \mid \lambda)$	$p(x) = \dfrac{e^{-\lambda}\lambda^x}{x!}$ $\mathcal{X} = 0,1,2,\dots$ $\Lambda = \mathbb{R}^+$	λ	λ	Counting events occurring 'at random' in space or time	$\mathrm{Bi}(x \mid n, \theta) \equiv \mathrm{Po}(x \mid n\theta)$ if n large, θ small

SOME CONTINUOUS DISTRIBUTIONS

Name	Notation	p.d.f. $f(x \mid \theta)$	$\mathbb{E}[X \mid \theta]$	$\mathbb{V}[X \mid \theta]$	Applications	Comments
Uniform	$\mathrm{Un}(x \mid \alpha, \beta)$	$f(x) = \dfrac{1}{\beta - \alpha}$ $\mathcal{X} = [\alpha, \beta]$ $\Theta = \{(\alpha,\beta) \in \mathbb{R}^2 : \alpha < \beta\}$	$\dfrac{\alpha+\beta}{2}$	$\dfrac{(\beta-\alpha)^2}{12}$	Rounding errors $\mathrm{Un}(x \mid -1/2, 1/2)$. Simulating other distributions from $\mathrm{Un}(x \mid 0, 1)$	
Exponential	$\mathrm{Ex}(x \mid \lambda)$	$f(x) = \lambda e^{-\lambda x}$ $\mathcal{X} = \mathbb{R}_+$ $\Lambda = \mathbb{R}_+$	$\dfrac{1}{\lambda}$	$\dfrac{1}{\lambda^2}$	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	$\mathrm{Ga}(x \mid \alpha, \beta)$	$f(x) = \dfrac{\beta^\alpha x^{\alpha-1} e^{-\beta x}}{\Gamma[\alpha]}$ $\mathcal{X} = \mathbb{R}_+$ $\Theta = \{(\alpha,\beta) \in \mathbb{R}^2 : \alpha > 0, \beta > 0\}$	$\dfrac{\alpha}{\beta}$	$\dfrac{\alpha}{\beta^2}$	Times between k events for Poisson Process. Lifetimes of ageing items.	Also parameterised in terms of $1/\beta$ $\mathrm{Ga}(x \mid 1, \lambda) \equiv \mathrm{Ex}(x \mid \lambda)$, $\mathrm{Ga}(x \mid \nu/2, 1/2) \equiv \chi^2_{(\nu)}(x)$
Beta	$\mathrm{Be}(x \mid \alpha, \beta)$	$f(x) = \dfrac{x^{\alpha-1}(1-x)^{\beta-1}}{\mathrm{B}(\alpha,\beta)}$ $\mathcal{X} = (0, 1)$ $\Theta = \{(\alpha,\beta) \in \mathbb{R}^2 : \alpha > 0, \beta > 0\}$	$\dfrac{\alpha}{\alpha+\beta}$	$\dfrac{\alpha\beta(\alpha+\beta)^{-2}}{(\alpha+\beta+1)}$	Useful model for variables with finite range. Also as a Bayesian conjugate prior.	$\mathrm{Be}(x \mid 1, 1) \equiv \mathrm{Un}(x \mid 0, 1)$ $\mathrm{Be}(x \mid \alpha, \beta)$ 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 $Y = (b-a)X + a$
Normal (Gaussian)	$\mathrm{N}(x \mid \mu, \sigma^2)$	$f(x) = \dfrac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\dfrac{x-\mu}{\sigma}\right)^2\right]$ $\mathcal{X} = \mathbb{R}$ $\Theta = \{(\mu, \sigma^2) \in \mathbb{R}^2 : \sigma^2 > 0\}$	μ	σ^2	Empirically and theoretically (via CLT) a useful model. Often parameterised in terms of the precision $\lambda = 1/\sigma^2$	$Y = aX + b \sim \mathrm{N}(y \mid a\mu + b, a^2\sigma^2)$ $Z = \frac{X-\mu}{\sigma} \sim \mathrm{N}(z \mid 0, 1)$ $\mathrm{P}[X \in (u, v)] = \mathrm{P}\left[Z \in \left(\dfrac{u-\mu}{\sigma}, \dfrac{v-\mu}{\sigma}\right)\right]$
Chi-square	$\chi^2_{(\nu)}(x)$	$f(x) = \dfrac{2^{-\nu/2}}{\Gamma(\nu/2)} x^{\nu/2-1} e^{-x/2}$ $\mathcal{X} = \mathbb{R}_+$; $\Theta = \mathbb{R}_+$	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	$\mathrm{St}(x \mid \mu, \lambda, \nu)$	$f(x) = \frac{\Gamma[(\nu+1)/2]}{\Gamma[\nu/2]}\left(\frac{\lambda}{\nu\pi}\right)^{1/2} \times$ $\left(1 + \lambda(x-\mu)^2/\nu\right)^{-(\nu+1)/2}$ $\mathcal{X} = \mathbb{R}, \quad \mu \in \mathbb{R}, \lambda, \nu > 0$	μ (if $\nu > 1$)	$\lambda^{-1}\dfrac{\nu}{\nu-2}$ (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.** $f(x \mid \theta)$	$\mathbb{E}[X \mid \theta]$	$\mathbb{V}[X \mid \theta]$	**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	$\mathrm{Di}(\boldsymbol{x} \mid \boldsymbol{\alpha})$	$f(\boldsymbol{x}) = \frac{\Gamma(\sum \alpha_l)}{\prod \Gamma(\alpha_l)} \prod_{l=1}^{k} x_l^{\alpha_l}$ $\boldsymbol{x} = \{x_1, \ldots, x_k\}$, $0 < x_l < 1$, $\sum_{l=1}^{k} x_l = 1$ $\boldsymbol{\alpha} = \{\alpha_1, \ldots, \alpha_k\}$, $0 < \alpha_l$	$\mathbb{E}[x_i] = \mu_i = \frac{\alpha_i}{\sum \alpha_l}$	$\mathbb{V}[x_i] = \frac{\mu_i(1-\mu_i)}{1+\sum \alpha_l}$ $\mathrm{Cov}[x_i, x_j] = -\frac{\mu_i\mu_j}{1+\sum \alpha_l}$	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	$f(x) = \mathrm{St}\big(x \mid \mu, \lambda\alpha\beta^{-1}, 2\alpha\big)$
Gaussian	$\mathrm{N}_k(\boldsymbol{x} \mid \boldsymbol{\mu}, \mathit{\Lambda})$	$f(\boldsymbol{x}) = \frac{\lvert\mathit{\Lambda}\rvert^{1/2}}{(2\pi)^{k/2}} \exp\big[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})'\mathit{\Lambda}(\boldsymbol{x}-\boldsymbol{\mu})\big]$ $\mathcal{X} = \boldsymbol{x} \in \mathbb{R}^k$ $\boldsymbol{\mu} \in \mathbb{R}^k$; $\mathit{\Lambda}$ symmetric positive-definite	$\boldsymbol{\mu}$	$\mathit{\Lambda}^{-1}$	See univariate case	Usually parameterised in terms of the covariance matrix $\mathit{\Sigma} = \mathit{\Lambda}^{-1}$
Student	$\mathrm{St_k}(\boldsymbol{x} \mid \boldsymbol{\mu}, \mathit{\Lambda}, \nu)$	$f(\boldsymbol{x}) = \frac{\lvert\mathit{\Lambda}\rvert^{1/2}\Gamma((\nu+k)/2)}{(\nu\pi)^{k/2}\Gamma(\nu/2)} \times \left[1 + \frac{1}{\nu}(\boldsymbol{x}-\boldsymbol{\mu})'\mathit{\Lambda}(\boldsymbol{x}-\boldsymbol{\mu})\right]^{-(\nu+k)/2}$ $\mathcal{X} = \boldsymbol{x} \in \mathbb{R}^k$ $\boldsymbol{\mu} \in \mathbb{R}^k$; $\mathit{\Lambda}$ symmetric positive-definite , $\nu > 0$	$\boldsymbol{\mu}$ (if $\nu > 1$)	$\frac{\nu}{\nu-2}\mathit{\Lambda}^{-1}$ (if $\nu > 2$)	See univariate case	Usually parameterised in terms of the covariance matrix $\mathit{\Sigma} = \mathit{\Lambda}^{-1}$