Data provided: Tables of distribution functions

MAS464

The University Of Sheffield.

SCHOOL OF MATHEMATICS AND STATISTICS **Autumn Semester 2019–20**

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 pharmaceutical company is developing a drug for alleviating some symptoms of migraine, specifically sensitivity to light. As an early (Phase I) clinical trial, n_t volunteers are given the treatment, n_c are given a placebo (a sugar pill) and the number of patients declaring an improvement to sensitivity to light are recorded as x_t for the treatment group and x_c for the placebo group.

(i) Assuming θ_t and θ_c are the improvement probabilities from the treatment and placebo, respectively, show that $\pi(\theta_i) = \mathrm{Be}(\theta_i \mid a, b)$, $i = t, c$ is a conjugate prior for the likelihood from each dataset and provide explicit expressions for the posterior parameters. *(5 marks)*

(ii) From the trial, $n_t = 20$, $n_c = 100$, $x_t = 11$ and $x_c = 40$ were obtained. The company will assess the effectiveness of the drug by comparing the odds from each group, $\phi_i = \theta_i/(1 - \theta_i)$. Assuming a conjugate prior with mean 1/2 and variance 1/8,

(a) Find $\mathbb{E}[\phi_t \mid n_t, x_t]$, $\mathbb{E}[\phi_c \mid n_c, x_c]$ and an approximate interval of probability 0.95 for each. Based on these intervals, is there evidence in the data in favour of the treatment? *(12 marks)*

(b) Show that the posterior expected value of the odds product, $\mathbb{E}[\phi_t\phi_c \mid n_t, n_c, x_t, x_c] = \mathbb{E}[\phi_t \mid n_t, n_c, x_t, x_c]\ \mathbb{E}[\phi_c \mid n_t, n_c, x_t, x_c] = 0.921$. *(5 marks)*

(c) Show that the posterior expected value of the odds ratio, $\mathbb{E}\left[\frac{\phi_t}{\phi_c} \,\middle|\, n_t, n_c, x_t, x_c\right] = 2.072$. *(6 marks)*

2 The local council is considering changes to the traffic system in a stretch of road near the University and must provide evidence of whether the change is needed. To this end, on each day $i = 1, \ldots, 40$, the number of accidents, y_i, in a specific hour are recorded. After consultation with their statisticians, it is agreed to treat the data as a random sample from a Poisson distribution, $\mathrm{Po}(x_i \mid \lambda)$.

(i) Show that $\pi(\lambda) = \mathrm{Ga}(\lambda \mid a, b)$ is conjugate and provide explicit expressions for the posterior parameters. *(4 marks)*

(ii) The council's elicited prior mean and variance are 10/3 and 50/9, respectively; and after the observation period a total of $s = \sum_{i=1}^{40} y_i = 11$ accidents were recorded. One of the key indicators the council will use in their decision making is the time between two consecutive accidents, t, which follows an Exponential distribution with rate λ, $t \sim \mathrm{Ex}(t \mid \lambda)$.

(a) Using the conjugate prior with the elicited values, show that the predictive distribution of the time between two consecutive accidents,

$$f(t \mid \boldsymbol{y}) = \frac{c}{d}\left(1 + \frac{t}{d}\right)^{-(c+1)},$$

and provide explicit values for c, d. *(5 marks)*

(b) Show $\mathbb{E}[t \mid \boldsymbol{y}] = \mathbb{E}\left[\lambda^{-1} \mid \boldsymbol{y}\right] = 3.383$. *(7 marks)*

(c) Explain why the predictive HPD interval of probability 0.975 for t is of the form $(0, q)$ and provide the value for q. *(5 marks)*

(d) The council will not introduce any changes to the traffic system unless there is evidence that the predictive time between two consecutive accidents is less than 5 hours. Is there evidence in the data for introducing a change? *(7 marks)*

3 Measurements of external water pressure against the bottom of a dam wall, μ, are taken with two different instruments. One is very precise but not accurate, the other is accurate but not very precise. Formally, for the first set of measurements we can assume the observations $x_i \sim \mathrm{N}(x_i \,|\, a\mu, 1/p)$, $i = 1, \ldots, n$ and for the second set $y_j \sim \mathrm{N}\big(y_j \,\big|\, \mu, 1/(bp)\big)$, $j = 1, \ldots, m$; with $p > 0$, $|a| > 1$, and $b < 1$ known constants. Assuming $\pi(\mu) \propto 1$,

(i) Show the posterior distribution for μ is Gaussian and find explicit expressions for the parameters. *(7 marks)*

(ii) Show that the posterior distribution converges to a point mass at the origin as the bias, $|a|$, increases. *(4 marks)*

(iii) Using the transformed observations, $z_i = x_i/a$, show that the posterior mean is a weighted average of the sample means and show the weight associated with $\bar{y}$ is $\dfrac{bm}{bm + a^2 n}$. *(7 marks)*

(iv) From a pilot experiment with $p = 0.1, a = 3, b = 4$, the following data were collected: $n = 12, m = 13, \bar{y} = 5.25, \bar{z} = 7.75$. The wall would be at risk of collapsing if the mean pressure, $\mu > 6.5$. Based on the posterior odds, do the data provide evidence of risk?. *(10 marks)*

HINT: If $Z \sim \mathrm{N}(z \,|\, 0, 1)$, then $\mathrm{P}[Z \leq 1] = 0.841$, $\mathrm{P}[Z \leq 1.25] = 0.894$, $\mathrm{P}[Z \leq 1.5] = 0.933$, $\mathrm{P}[Z \leq 1.75] = 0.960$, $\mathrm{P}[Z \leq 2] = 0.977$.

4 Consider the simple regression model with random effects,

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

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

$$\begin{aligned}
\pi(\lambda) &= \mathrm{Ga}(\lambda \,|\, a, b), \\
\pi(\alpha_i) &= \mathrm{N}\left(\alpha_i \,\middle|\, \mu, \frac{1}{p}\right); \quad i = 1, \dots, n, \\
\pi(\mu) &= \mathrm{N}\left(\mu \,\middle|\, m, \frac{1}{q}\right), \\
\pi(\beta) &= \mathrm{N}\left(\beta \,\middle|\, b, \frac{1}{t}\right),
\end{aligned}$$

where $a, b, p, q, t > 0$ and $m \in \mathbb{R}$ are known constants.

(i) Show that the full conditional posterior distribution of:

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

(b) The individual intercepts, α_i, is Gaussian and provide explicit expressions for the parameters. *(5 marks)*

(c) The location hyper-parameter, μ, is Gaussian and provide explicit expressions for the parameters. *(5 marks)*

(d) The regression slope, β, is Gaussian 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 2019–20

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 **probability density function** (pdf). 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

$$f(\boldsymbol{x} \mid \boldsymbol{\theta}) \geq 0 \qquad \text{and} \qquad \int_{\mathcal{X}} f(\boldsymbol{x} \mid \boldsymbol{\theta})\, \mathrm{d}\boldsymbol{x} = 1;$$

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,

$$\mathbb{E}[\boldsymbol{x} \mid \boldsymbol{\theta}] = \int_{\mathcal{X}} \boldsymbol{x}\ f(\boldsymbol{x} \mid \boldsymbol{\theta})\, \mathrm{d}\boldsymbol{x},$$

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

$$\mathrm{Cov}[\boldsymbol{x} \mid \boldsymbol{\theta}] = \int_{\mathcal{X}} (\boldsymbol{x} - \mathbb{E}[\boldsymbol{x} \mid \boldsymbol{\theta}])'(\boldsymbol{x} - \mathbb{E}[\boldsymbol{x} \mid \boldsymbol{\theta}]) f(\boldsymbol{x} \mid \boldsymbol{\theta})\, \mathrm{d}\boldsymbol{x}\ ,$$

respectively stand for the mean, 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.

DISCRETE DISTRIBUTIONS

Name	Notation	p.f. $p(x \mid \theta)$	$\mathbb{E}[X \mid \theta]$	$\mathbb{V}[X \mid \theta]$	Applications	Comments
Bernoulli	$\mathrm{Ber}(x \mid \theta)$	$p(x) = \theta^x (1-\theta)^{1-x}$ $\mathcal{X} = \{0, 1\}$ $\Theta = (0, 1)$	θ	$\theta(1-\theta)$	Coins, trials.	Constituent of more complex distributions. Expt. with binary outcome: success w.p. θ and failure w.p. $1-\theta$.
Binomial	$\mathrm{Bi}(x \mid n, \theta)$	$p(x) = \binom{n}{x}\theta^x (1-\theta)^{n-x}$ $\mathcal{X} = \{0, 1, 2, ..., n\}$ $\Theta = (0, 1)$	$n\theta$	$n\theta(1-\theta)$	Sampling with replacement	$X \equiv$ no. successes in n ind. $\mathrm{Ber}(x \mid \theta)$ trials. $\mathrm{Bi}(x \mid 1, \theta) \equiv \mathrm{Ber}(x \mid \theta)$
Geometric	$\mathrm{Ge}(x \mid \theta)$	$p(x) = \theta(1-\theta)^x$ $\mathcal{X} = 0, 1, 2, ...$ $\Theta = (0, 1)$	$\frac{1-\theta}{\theta}$	$\frac{1-\theta}{\theta^2}$	Waiting times (for single events)	$X \equiv$ no. failures until 1st success in sequence of ind. $\mathrm{Ber}(x \mid \theta)$ trials. Alternative formulation in terms of $Y \equiv$ no. of trials to 1st success ($Y = X + 1$)
Negative binomial (Pascal)	$\mathrm{NB}(x \mid m, \theta)$	$p(x) = \binom{m+x-1}{x}\theta^m (1-\theta)^x$ $\mathcal{X} = 0, 1, 2, ...$ $\Theta = (0, 1)$	$\frac{m(1-\theta)}{\theta}$	$\frac{m(1-\theta)}{\theta^2}$	Waiting times (for compound events)	$X \equiv$ no. failures to m-th success in sequence of ind. $\mathrm{Ber}(x \mid \theta)$ trials. Generalisation of Geometric. $\mathrm{NB}(x \mid 1, \theta) \equiv \mathrm{Ge}(x \mid \theta)$
Hypergeometric	$\mathrm{Hy}(x \mid N, d, n)$ (not standard, esp. order of arguments)	$p(x) = \frac{\binom{d}{x}\binom{N-d}{n-x}}{\binom{N}{n}}$ $\mathcal{X} = \{a, a+1, \ldots, b\}$ $a = \max\{0, n+d-N\}$, $b = \min\{n, d\}$	$\frac{nd}{N}$	$\frac{nd}{N}\frac{N-n}{N-1}\left(1 - \frac{d}{N}\right)$	Sampling without replacement	$X \equiv$ no. of defectives in sample of size n taken without replacement from population of size N of which d are defective. $\mathrm{Bi}(x \mid n, d/N)$ —a suitable approx if $n/N < 0.1$
Poisson	$\mathrm{Po}(x \mid \lambda)$	$p(x) = \frac{e^{-\lambda}\lambda^x}{x!}$ $\mathcal{X} = 0, 1, 2, ...$ $\Lambda = \mathbb{R}^+$	λ	λ	Counting (rare) events occurring at random in space or time	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. $\mathrm{Bi}(x \mid n, \theta) \approx \mathrm{Po}(x \mid n\theta)$ if n large, θ small, and $n\theta = c$.

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	$\mathrm{Un}(x \mid \alpha, \beta)$	$f(x) = \dfrac{1}{\beta - \alpha}$ $\mathcal{X} = [\alpha, \beta]$ $\Theta = \{(\alpha, \beta) \in \mathbb{R}^2 \,:\, \alpha < \beta\}$	$\dfrac{\alpha + \beta}{2}$	$\dfrac{(\beta - \alpha)^2}{12}$	Rounding errors $\mathrm{Un}(x \mid -1/2, 1/2)$. Simulating other distributions from $\mathrm{Un}(x \mid 0, 1)$	Used as non-informative prior for parameters with bounded support.
Pareto	$\mathrm{Pa}(x \mid \alpha, \beta)$	$f(x) = \alpha\beta^\alpha x^{-(\alpha+1)}$ $\mathcal{X} = (\beta, \infty)$ $\Theta = \{(\alpha, \beta) \in \mathbb{R}^2 \,:\, \alpha > 0, \beta > 0\}$	$\dfrac{\alpha\beta}{\alpha - 1}$ (if $\alpha > 1$)	$\dfrac{\alpha\beta^2}{(\alpha-2)(\alpha-1)^2}$ (if $\alpha > 2$)	Distribution of positive random quantities with heavy tails	Conjugate prior for uniform data with known lower bound
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. Conjugate prior for exponential model.	Also parameterised in terms of $1/\beta$ $\mathrm{Ga}(x \mid 1, \lambda) \equiv \mathrm{Ex}(x \mid \lambda)$, $1/x = y \sim \mathrm{IGa}(y \mid \alpha, \beta)$
Beta	$\mathrm{Be}(x \mid \alpha, \beta)$	$f(x) = \dfrac{x^{\alpha-1}(1-x)^{\beta-1}}{\mathrm{B}(\alpha, \beta)}$ $\mathcal{X} = (0, 1)$ $\Theta = \{(\alpha, \beta) \in \mathbb{R}^2 \,:\, \alpha > 0, \beta > 0\}$	$\mu = \dfrac{\alpha}{\alpha + \beta}$	$\dfrac{\mu(1-\mu)}{(\alpha + \beta + 1)}$	Useful model for variables with finite range. Conjugate prior for Binomial model.	$\mathrm{Be}(x \mid 1, 1) \equiv \mathrm{Un}(x \mid 0, 1)$ Can re-scale $\mathrm{Be}(x \mid \alpha, \beta)$ to any finite range (a, b) by $Y = (b - a)X + a$
Gaussian (Normal)	$\mathrm{N}(x \mid \mu, \sigma^2)$	$f(x) = \dfrac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\dfrac{x - \mu}{\sigma}\right)^2\right]$ $\mathcal{X} = \mathbb{R}$ $\Theta = \{(\mu, \sigma^2) \in \mathbb{R}^2 \,:\, \sigma^2 > 0\}$	μ	σ^2	Empirically and theoretically (via CLT) a useful model. Also parameterised in terms of the precision $\lambda = 1/\sigma^2$	$Y = a + bX \sim \mathrm{N}(y \mid a + b\mu, b^2\sigma^2)$ $Z = \frac{X-\mu}{\sigma} \sim \mathrm{N}(z \mid 0, 1)$ $\mathrm{P}[X \in (u, v)] = \mathrm{P}\left[Z \in \left(\dfrac{u - \mu}{\sigma}, \dfrac{v - \mu}{\sigma}\right)\right]$
Student t	$\mathrm{St}(x \mid \mu, \lambda, \nu)$	$f(x) = \dfrac{\Gamma[(\nu+1)/2]}{\Gamma[\nu/2]}\left(\frac{\lambda}{\nu\pi}\right)^{1/2} \times$ $\left(1 + \frac{\lambda}{\nu}(x - \mu)^2\right)^{-(\nu+1)/2}$ $\mathcal{X} = \mathbb{R}, \quad \mu \in \mathbb{R}, \lambda, \nu > 0$	μ (if $\nu > 1$)	$\lambda^{-1}\dfrac{\nu}{\nu - 2}$ (if $\nu > 2$)	Useful alternative to Gaussian for random quantities 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. $\mathrm{t}^2_\nu \equiv \mathrm{F}_{1,\nu}$.

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	$\mathrm{Mu}(\boldsymbol{x} \mid \boldsymbol{\theta}, n)$	$p(\boldsymbol{x}) = \frac{n!}{\prod_{l=1}^{k} x_l!} \prod_{l=1}^{k} \theta_l^{x_l}$ $\boldsymbol{x} = \{x_1, \ldots, x_k\}, \quad x_l = 0, 1, \ldots, \sum x_l = n$ $\boldsymbol{\theta} = \{\theta_1, \ldots, \theta_k\}, \quad 0 < \theta_l < 1, \ \sum \theta_l = 1$	$\mathbb{E}[x_i] = n\theta_i$	$\mathbb{V}[x_i] = n\theta_i(1-\theta_i)$ $\mathrm{Cov}[x_i, x_j] = -n\theta_i\theta_j$	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 probabilities of exclusive events.	Generalisation of the Beta distribution. Conjugate prior for multinomial data
Normal-Gamma	$\mathrm{NG}(x, y \mid \mu, \kappa, \alpha, \beta)$	$f(x, y) = \mathrm{N}\big(x \,\big\vert\, \mu, (y\kappa)^{-1}\big)\mathrm{Ga}(y \mid \alpha, \beta)$ $\mathcal{X} = \{(x, y) \ : \ x \in \mathbb{R}, y > 0\}$ $\mu \in \mathbb{R}; \ \kappa, \alpha, \beta > 0$	$\mathbb{E}[x] = \mu$ $\mathbb{E}[y] = \frac{\alpha}{\beta}$	$\mathbb{V}[x] = \frac{\beta}{\kappa(\alpha - 1)}$ $\mathbb{V}[y] = \frac{\alpha}{\beta^2}$	Conjugate prior for Gaussian data, both parameters unknown	The marginal distribution of x is $\mathrm{St}(x \mid \mu, \kappa\alpha/\beta, 2\alpha)$
(Multivariate) Gaussian	$\mathrm{N}_k(\boldsymbol{x} \mid \boldsymbol{\mu}, \Lambda)$	$f(\boldsymbol{x}) = \frac{\lvert\Lambda\rvert^{1/2}}{(2\pi)^{k/2}} \exp\big[-\tfrac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})'\Lambda(\boldsymbol{x} - \boldsymbol{\mu})\big]$ $\mathcal{X} = \mathbb{R}^k$ $\boldsymbol{\mu} \in \mathbb{R}^k; \ \Lambda$ symmetric positive-definite	$\boldsymbol{\mu}$	Λ^{-1}	See univariate case	Usually parameterised in terms of the covariance matrix $\Sigma = \Lambda^{-1}$
(Multivariate) Student	$\mathrm{St_k}(\boldsymbol{x} \mid \boldsymbol{\mu}, \Lambda, \nu)$	$f(\boldsymbol{x}) = \frac{\lvert\Lambda\rvert^{1/2}\Gamma((\nu + k)/2)}{(\nu\pi)^{k/2}\Gamma(\nu/2)} \times$ $\left[1 + \frac{1}{\nu}(\boldsymbol{x} - \boldsymbol{\mu})'\Lambda(\boldsymbol{x} - \boldsymbol{\mu})\right]^{-(\nu+k)/2}$ $\mathcal{X} = \mathbb{R}^k$ $\boldsymbol{\mu} \in \mathbb{R}^k; \ \Lambda$ 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 $\Sigma = \Lambda^{-1}$
Wishart	$\mathrm{Wi}_k(X \mid \alpha, \Omega)$	$f(X) = \frac{(\pi)^{k(k-1)}\lvert\Omega\rvert^{\alpha}}{\prod_{i=1}^{k} \Gamma[(2\alpha + 1 - i)/2]} \times$ $\lvert X\rvert^{\alpha - (k+1)/2} \exp[-\operatorname{tr}(\Omega X)]$ $\mathcal{X} =$ symmetric positive-definite $\alpha > (k-1)/2; \ \Omega$ symmetric non-singular	$\alpha\Omega^{-1}$	$\mathbb{V}[X_{ij}] = \alpha\Big(\omega_{ij}^2 + \omega_{ii}\omega_{jj}\Big)$	Conjugate prior for the precision matrix in a Gaussian model	Can also be used for the covariance matrix after the appropriate transformation.