Data provided: Tables of distribution functions

The University Of Sheffield.

MAS464

SCHOOL OF MATHEMATICS AND STATISTICS **Autumn Semester 2016–17**

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 biologist is interested in the prevalence of a genetic disorder, $0 < \theta < 1$, in the bacterium *Streptomyces coelicolor*. She has access to results of a recent experiment in which a random sample of n bacteria were tested and x were positive for the disorder.

(i) Show that the Jeffreys prior is conjugate for this model and give explicit expressions for the posterior parameters. ***(6 marks)***

(ii) Using the same population, a new experiment is carried out using negative binomial sampling in which bacteria are screened until m without the disorder are observed, with m fixed in advance.

(a) Denoting by y the number of bacteria in the sample with the disorder and assuming both samples are stochastically independent, argue why the likelihood function using all the data can be written as

$$\mathrm{L}(\theta\,;\,x,y,m) \propto \theta^{x}(1-\theta)^{n-x}\theta^{y}(1-\theta)^{m}\,.$$

(4 marks)

(b) Show that the Jeffreys prior for the whole data is not conjugate, but the Beta distribution still is. ***(8 marks)***

(c) A third biologist is trying to set up a similar experiment. In order to prepare for it, she needs to know the probability of observing two or more healthy bacteria before observing 5 with the disorder. Using all the data available, calculate her predictive probability if she specifies $\pi(\theta) = \mathrm{Be}(\theta\,|\,1,1)$ as her prior and $n = 10, x = 3, m = 5$ and $y = 3$ were recorded in the previous two experiments. ***(10 marks)***

2 A branch manager is interested in the rate of clients served in a day, θ. Through a typical period he records a random sample of clients served by day $\boldsymbol{x} = \{x_1, \ldots, x_n\}$ and assumes $x_i \sim \mathrm{Po}(x_i \mid \theta)$. He decides to use $\pi(\theta) = \mathrm{Ga}(\theta \mid a, b)$ as a prior.

(i) Show that his posterior distribution is $\mathrm{Ga}(\theta \mid a^\star, b^\star)$ and provide explicit expressions for the posterior parameters. ***(5 marks)***

(ii) Show that the posterior mean is the optimal decision under square error loss. ***(5 marks)***

(iii) Using past records of similar branches the manager elicits $\mathbb{E}[\theta] = 10/3$ and $\mathbb{V}[\theta] = 50/9$ and obtains $n = 40$ and $\sum_{i=1}^{40} x_i = 425.3$ from the sample.

(a) Calculate his prior and posterior point estimates under a quadratic loss function,

$$\mathcal{L}(\theta, \widehat{\theta}) = \left(\theta - \widehat{\theta}\right)^2 ,$$

and the associated expected loss. ***(6 marks)***

(b) Calculate his posterior point estimate under the absolute loss function,

$$\mathcal{L}(\theta, \widehat{\theta}) = \left|\theta - \widehat{\theta}\right| ,$$

assuming the posterior distribution can be approximated by a Gaussian. ***(6 marks)***

(c) Using a zero-one loss function,

$$\mathcal{L}(\theta, \widehat{\theta}) = \begin{cases} 0 & |\theta - \widehat{\theta}| < c \\ 1 & |\theta - \widehat{\theta}| \geq c \end{cases} ,$$

and assuming $c \to 0$, calculate his prior and posterior point estimates. ***(6 marks)***

3 A common model used in experimental design to investigate whether the mean of several populations is the same or not can be written as

$$y_{ij} = \mu_j + \varepsilon_i \ ; \qquad i = 1, \ldots, n_j \ ,$$
$$\mu_j \sim \mathrm{N}\big(\mu_j \,\big|\, \eta, 1/t\big) \ , \quad j = 1, \ldots, k$$

and

$$\varepsilon_i \sim \mathrm{N}(\varepsilon_i \,|\, 0, 1/\lambda) \ , \quad \text{independent} \ ,$$

where $y_{ij} \in \mathbb{R}$, are the observations; $\mu_j \in \mathbb{R}$, the mean of the j-th population, $\boldsymbol{\mu} = \{\mu_1, \ldots, \mu_k\}$, $\eta \in \mathbb{R}$ and $\lambda > 0$ are unknown parameters. Let the prior be

$$\pi(\eta, \lambda) = \mathrm{N}(\eta \,|\, m, 1/p) \, \mathrm{Ga}(\lambda \,|\, a, b) \ .$$

with $\{t, m, p, a, b\}$ known.

(i) (a) Show that the full conditional distribution of each μ_j is $\mathrm{N}\big(\mu_j \,\big|\, m_j^\star, 1/t_j^\star\big)$ and give explicit expressions for the parameters. ***(6 marks)***

(b) Show that the full conditional distribution of η is $\mathrm{N}(\eta \,|\, m^\star, 1/p^\star)$ and give explicit expressions for the parameters. ***(7 marks)***

(c) Show that the full conditional distribution of λ is $\mathrm{Ga}(\lambda \,|\, a^\star, b^\star)$ and give explicit expressions for the parameters. ***(5 marks)***

(ii) Write down pseudo-code for an MCMC scheme to explore the posterior distribution $\pi(\boldsymbol{\mu}, \eta, \lambda \mid \boldsymbol{y})$. ***(10 marks)***

4 An engineer is testing a new precision weighing device. In her experimental design n pieces of titanium of identical known weight are measured and the relative discrepancy, $\boldsymbol{y} = \{y_1, \ldots, y_n\}$ is recorded and it is assumed $y_i \sim \mathrm{Un}(y_i \,|\, 0, \theta)$, where θ represents the maximum technical discrepancy of the device.

(i) Sketch the likelihood function and show that $\widehat{\theta} = y_{(n)} = \max\{y_1, \ldots, y_n\}$ is the MLE. ***(4 marks)***

(ii) The engineer decides to use

$$\mathrm{Pa}(\theta \,|\, a, b) = a b^a \theta^{-(a+1)} \ , \qquad \theta > b \ , \quad a, b > 0 \ ,$$

as a prior distribution.

(a) Sketch the engineer's prior distribution. ***(4 marks)***

(b) Show that her posterior distribution is $\mathrm{Pa}(\theta \,|\, a^\star, b^\star)$, with $a^\star = n + a$ and $b^\star = \max\Big\{b, \widehat{\theta}\Big\}$. ***(10 marks)***

(c) Discuss the implications on the Bayesian learning process if $b > \widehat{\theta}$. ***(6 marks)***

(iii) Provide the HPD interval of size 0.95 if $n = 10$, $\widehat{\theta} = 0.5$, $a = 3$ and $b = 0.4$. ***(4 marks)***

End of Question Paper

Notation and distributions

Bayesian Statistics 2016–17

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 $f(\boldsymbol{x}|\boldsymbol{\phi})$ and $f(\boldsymbol{y}|\boldsymbol{\psi})$ 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 $\mathrm{N}(x \mid \mu, \sigma^2)$. Tables below present some density functions used throughout the course.

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

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

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

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

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)	$\mathrm{U}(1, ..., k)$	$p(x) = 1/k$ $\mathcal{X} = \{1, ..., k\}$, $\mathcal{K} = \mathbb{Z}_+$	$\dfrac{k+1}{2}$	$\dfrac{k^2-1}{12}$	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	$\mathrm{Bi}(x \mid n, \theta)$	$p(x) = \binom{n}{x}\theta^x (1-\theta)^{n-x}$ $\mathcal{X} = \{0, 1, 2, ..., n\}$ $\Theta = (0, 1)$	$n\theta$	$n\theta(1-\theta)$	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	$\mathrm{Ge}(x \mid \theta)$	$p(x) = \theta(1-\theta)^x$ $\mathcal{X} = 0, 1, 2, ...$ $\Theta = (0, 1)$	$\dfrac{1-\theta}{\theta}$	$\dfrac{1-\theta}{\theta^2}$	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	$\mathrm{NB}(x \mid m, \theta)$	$p(x) = \binom{m+x-1}{x}\theta^m (1-\theta)^x$ $\mathcal{X} = 0, 1, 2, ...$ $\Theta = (0, 1)$	$\dfrac{m(1-\theta)}{\theta}$	$\dfrac{m(1-\theta)}{\theta^2}$	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	$\mathrm{Po}(x \mid \lambda)$	$p(x) = \dfrac{e^{-\lambda}\lambda^x}{x!}$ $\mathcal{X} = 0, 1, 2, ...$ $\Lambda = \mathbb{R}^+$	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	$\mathrm{Un}(x \mid \alpha, \beta)$	$f(x) = \dfrac{1}{\beta - \alpha}$ $\mathcal{X} = [\alpha, \beta]$ $\Theta = \{(\alpha, \beta) \in \mathbb{R}^2 \,:\, \alpha < \beta\}$	$\dfrac{\alpha + \beta}{2}$	$\dfrac{(\beta - \alpha)^2}{12}$	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. 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	$\mathrm{Be}(x \mid \alpha, \beta)$	$f(x) = \dfrac{x^{\alpha-1}(1-x)^{\beta-1}}{\mathrm{B}(\alpha, \beta)}$ $\mathcal{X} = (0, 1)$ $\Theta = \{(\alpha, \beta) \in \mathbb{R}^2 \,:\, \alpha > 0, \beta > 0\}$	$\dfrac{\alpha}{\alpha + \beta}$	$\dfrac{\alpha\beta(\alpha + \beta)^{-2}}{(\alpha + \beta + 1)}$	Useful model for variables with finite range. 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 $Y = (b - a)X + a$
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	$\chi^2_{(\nu)}(x)$	$f(x) = \dfrac{2^{-\nu/2}}{\Gamma(\nu/2)} x^{\nu/2-1} e^{-x/2}$ $\mathcal{X} = \mathbb{R}_+$; $\quad \Theta = \mathbb{R}_+$	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 $\nu > 1$)	$\lambda^{-1}\dfrac{\nu}{\nu - 2}$ (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.** $f(\boldsymbol{x} \mid \boldsymbol{\theta})$	$\mathbb{E}[\boldsymbol{X} \mid \boldsymbol{\theta}]$	$\mathbb{V}[\boldsymbol{X} \mid \boldsymbol{\theta}]$	**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	$\mathrm{NG}(x, y \mid \mu, \lambda, \alpha, \beta)$	$f(x, y) = \mathrm{N}\big(x \,\big\vert\, \mu, (y\lambda)^{-1}\big)\mathrm{Ga}(y \mid \alpha, \beta)$ $\mathcal{X} = \{(x, y) \,:\, x \in \mathbb{R}, y > 0\}$ $\mu \in \mathbb{R};\ \lambda, \alpha, \beta > 0$	$\mathbb{E}[x] = \mu$ $\mathbb{E}[y] = \alpha\beta^{-1}$	$\mathbb{V}[x] = \frac{\beta}{\lambda(\alpha - 1)}$ $\mathbb{V}[y] = \alpha\beta^{-2}$	Conjugate prior for Gaussian data	$f(x) = \mathrm{St}\big(x \,\big\vert\, \mu, \lambda\alpha\beta^{-1}, 2\alpha\big)$
Gaussian	$\mathrm{N}_k(\boldsymbol{x} \mid \boldsymbol{\mu}, \Lambda)$	$f(\boldsymbol{x}) = \frac{\lvert\Lambda\rvert^{1/2}}{(2\pi)^{k/2}} \exp\big[-\tfrac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})'\Lambda(\boldsymbol{x}-\boldsymbol{\mu})\big]$ $\mathcal{X} = \boldsymbol{x} \in \mathbb{R}^k$ $\boldsymbol{\mu} \in \mathbb{R}^k;\ \Lambda$ symmetric positive-definite	$\boldsymbol{\mu}$	Λ^{-1}	See univariate case	Usually parameterised in terms of the covariance matrix $\Sigma = \Lambda^{-1}$
Student	$\mathrm{St_k}(\boldsymbol{x} \mid \boldsymbol{\mu}, \Lambda, \nu)$	$f(\boldsymbol{x}) = \frac{\lvert\Lambda\rvert^{1/2}\Gamma((\nu+k)/2)}{(\nu\pi)^{k/2}\Gamma(\nu/2)} \times \left[1 + \frac{1}{\nu}(\boldsymbol{x}-\boldsymbol{\mu})'\Lambda(\boldsymbol{x}-\boldsymbol{\mu})\right]^{-(\nu+k)/2}$ $\mathcal{X} = \boldsymbol{x} \in \mathbb{R}^k$ $\boldsymbol{\mu} \in \mathbb{R}^k;\ \Lambda$ 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 $\Sigma = \Lambda^{-1}$