Data provided: Tables of distribution functions

**MAS364**

The University Of Sheffield.

**SCHOOL OF MATHEMATICS AND STATISTICS** **Autumn Semester 2020–21**

**Bayesian Statistics**

*This is an open book exam.*
*Answer all questions.*
*You can work on the exam during the 24 hour period starting at 10am (GMT), and you must submit your work within 2.5 hours of accessing the exam paper or by the end of the 24 hour period (whichever is earlier).* ***Late submission will not be considered without extenuating circumstances.*** *Unless it is explicitly stated otherwise, it is intended that calculations are performed by hand (possibly with the aid of a calculator). To gain full marks, you will need to show your working. By uploading your solutions you declare that your submission consists entirely of your own work, that any use of sources or tools other than material provided for this module is cited and acknowledged, and that no unfair means have been used.*

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

**1** The Yorkshire education board is assessing whether to implement changes to the exams for a Year 12 module as it is suspected average marks are not within target. They make available the average marks from a sample of 11 schools in the region, $\boldsymbol{x} = \{53.80, 50.43, 55.07, 50.63, 53.84, 53.55, 53.95, 53.90, 53.58, 51.06, 51.96\}$. Since these are average marks, you can assume $x_i \sim \mathrm{N}(x_i \mid \mu, 1/q)$, with $q$ known.

(i) The board director believes the expected average mark is 55 and almost certainly is within $(47, 63)$. Fit a Gaussian distribution that reflects this prior information, show it is conjugate and give explicit expressions for the posterior parameters. ***(6 marks)***

(ii) The board is concerned their beliefs could unduly influence the decision making. To investigate this issue, derive the posterior using the Jeffreys prior, $\pi_{\mathrm{J}}(\mu) \propto 1$, and comment on the differences with the board's posterior using point and interval estimates. ***(7 marks)***

(iii) The board will base their decision on analysing the prediction of the average mark, $\bar{y} = \frac{1}{m}\sum_{j=1}^{m} y_j$, from a second sample of schools, $\boldsymbol{y} = \{y_1, \ldots, y_m\}$, with $m = 7$. Prove that the predictive distribution $f(\bar{y} \mid \boldsymbol{x}) = \mathrm{N}(\bar{y} \mid a, v)$ and provide explicit expressions for its mean, $a$ and variance, $v$. ***(7 marks)***

**2** The University of Sheffield is investigating the number of SARS-CoV-2 cases in campus using a new test developed by the Department of Biochemical Science. Before applying the test to all the university population, a random sample of volunteers are tested to assess the quality of the test. The result from each test, $x_i$, is recorded as positive or negative. After testing 250 volunteers, 2 where found positive.

(i) Briefly explain what are the assumptions behind assuming $x_i \sim \mathrm{Ber}(x_i \mid \theta)$ and what is the interpretation of the unknown parameter $\theta$ in the model.

*(3 marks)*

(ii) (a) Show that $\pi(\theta) = \mathrm{Be}(\theta \mid a, b)$ is a conjugate prior for this model and provide explicit expressions for the posterior parameters.

*(5 marks)*

(b) We conducted an elicitation exercise with the lead virologist in the department about the current infection rate in Sheffield, and obtained a prior median of 1% and quartiles $(0.5\%, 3\%)$. Obtain prior parameters from this information, explain carefully your fitting process.

*(5 marks)*

(c) Calculate the posterior mean and mode. Use these estimates to discuss whether using a Gaussian approximation to the posterior probability interval of probability 0.95 is appropriate and provide the approximate interval. *(7 marks)*

**3** A team of physicists are developing an microscope imaging technique that requires illuminating the target sample with a high power laser beam. It is known that, when hitting the target, the beam will scatter randomly around it, shifting the measured location of the target. Measuring this scattering is key for precise imaging, so a controlled experiment is designed where a single target is placed in the centre of the sample surface, imaged repeatedly and the location of the obtained target on the $(x, y)$ plane recorded in nanometers (nm).

The data recorded $(\boldsymbol{x}, \boldsymbol{y}) = \{(x_i, y_i)\}$, $i = 1, \ldots, n$ is assumed independent and distributed uniformly on the the disk of radius $\theta$,

$$f(x_i, y_i \mid \theta) = \begin{cases} \dfrac{1}{2\pi\theta^2} & 0 < \sqrt{x_i^2 + y_i^2} < \theta \\ 0 & \text{otherwise} \end{cases}$$

(i) To help in the analysis, show that the Pareto,

$$\mathrm{Pa}(\theta \mid a, c) = ac^a\theta^{-(a+1)}; \quad \theta > c, \ a, c > 0,$$

is a conjugate prior for $\theta$ and obtain explicit expressions for the updated parameters. *(7 marks)*


**Question 3 continued on next page**

**3** (continued)

(ii) The experiment is repeated 25 times with the results displayed on the table below

| i | $(x_i, y_i)$ | i | $(x_i, y_i)$ |
|---|---|---|---|
| 1 | $(-2.161, -0.423)$ | 14 | $(-0.133, 0.593)$ |
| 2 | $(1.265, -0.808)$ | 15 | $(-0.706, -2.234)$ |
| 3 | $(-2.008, -0.664)$ | 16 | $(1.751, -1.481)$ |
| 4 | $(1.639, -0.590)$ | 17 | $(-0.046, 2.165)$ |
| 5 | $(-0.101, -0.181)$ | 18 | $(-0.434, -1.659)$ |
| 6 | $(-0.915, 1.683)$ | 19 | $(0.928, -1.059)$ |
| 7 | $(1.598, -0.564)$ | 20 | $(0.305, 0.303)$ |
| 8 | $(1.940, 1.244)$ | 21 | $(0.541, -0.251)$ |
| 9 | $(2.340, 0.371)$ | 22 | $(2.309, 0.235)$ |
| 10 | $(-0.186, -2.421)$ | 23 | $(-0.117, 1.032)$ |
| 11 | $(-1.041, -0.547)$ | 24 | $(-1.565, 0.282)$ |
| 12 | $(0.424, 1.068)$ | 25 | $(-0.375, -1.213)$ |
| 13 | $(0.467, -0.150)$ | | |

(a) Due to physical constraints, it is known $\theta > 1$. After elicitation, the physicists believe the prior variance is $\sqrt{2}$. Use this information to elicit the prior parameters. *(3 marks)*

(b) Calculate the posterior mean, mode and median. Which of these point estimators would you suggest to the physicists and why? *(6 marks)*

(c) Provide a Highest Posterior Density interval of probability 0.9. *(4 marks)*

## End of Question Paper

# Notation and distributions

Bayesian Statistics 2020–21

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 $\Theta$ 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, uppercase matrix valued.

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

Moments and other descriptive measures of probability distributions are denoted 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 | $\mathrm{Ber}(x \mid \theta)$ | $p(x) = \theta^x(1-\theta)^{1-x}$<br>$\mathcal{X} = \{0,1\}$<br>$\Theta = (0,1)$ | $\theta$ | $\theta(1-\theta)$ | Coins, trials. | Constituent of more complex distributions. Experiments with binary outcome: success w.p. $\theta$ and failure w.p. $1-\theta$. |
| Binomial | $\mathrm{Bi}(x \mid n, \theta)$ | $p(x) = \binom{n}{x}\theta^x(1-\theta)^{n-x}$<br>$\mathcal{X} = \{0,1,2,\ldots,n\}$<br>$\Theta = (0,1)$ | $n\theta$ | $n\theta(1-\theta)$ | 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 | $\mathrm{Ge}(x \mid \theta)$ | $p(x) = \theta(1-\theta)^x$<br>$\mathcal{X} = 0,1,2,\ldots$<br>$\Theta = (0,1)$ | $\dfrac{1-\theta}{\theta}$ | $\dfrac{1-\theta}{\theta^2}$ | 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 ($Y = X + 1$) |
| Poisson | $\mathrm{Po}(x \mid \lambda)$ | $p(x) = \dfrac{e^{-\lambda}\lambda^x}{x!}$<br>$\mathcal{X} = 0,1,2,\ldots$<br>$\Lambda = \mathbb{R}^+$ | $\lambda$ | $\lambda$ | Counting (rare) events occurring at random in space or time | Arises empirically or via Poisson Process (PP) for counting events. For PP rate $\nu$ 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, $\theta$ small, and $n\theta = c$. |
| Negative binomial (Pascal) | $\mathrm{NB}(x \mid m, \theta)$ | $p(x) = \binom{m+x-1}{x}\theta^m(1-\theta)^x$<br>$\mathcal{X} = 0,1,2,\ldots$<br>$\Theta = (0,1)$ | $\dfrac{m(1-\theta)}{\theta}$ | $\dfrac{m(1-\theta)}{\theta^2}$ | 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) | $p(x) = \dfrac{\binom{d}{x}\binom{N-d}{n-x}}{\binom{N}{n}}$<br>$\mathcal{X} = \{a, a+1, \ldots, b\}$<br>$a = \max\{0, n+d-N\}$,<br>$b = \min\{n, d\}$ | $\dfrac{nd}{N}$ | $\dfrac{nd}{N}\dfrac{N-n}{N-1}\left(1 - \dfrac{d}{N}\right)$ | 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 $n/N < 0.1$ |

## 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 | $\mathrm{Un}(x \mid \alpha, \beta)$ | $f(x) = \frac{1}{\beta - \alpha}$ <br> $\mathcal{X} = [\alpha, \beta]$ <br> $\Theta = \{(\alpha, \beta) \in \mathbb{R}^2 \,:\, \alpha < \beta\}$ | $\frac{\alpha + \beta}{2}$ | $\frac{(\beta - \alpha)^2}{12}$ | Rounding errors $\mathrm{Un}(x \mid -1/2, 1/2)$. Simulating other distributions from $\mathrm{Un}(x \mid 0, 1)$ | Used as non-informative prior for parameters with bounded support. |
| Pareto | $\mathrm{Pa}(x \mid \alpha, \beta)$ | $f(x) = \alpha\beta^\alpha x^{-(\alpha+1)}$ <br> $\mathcal{X} = (\beta, \infty)$ <br> $\Theta = \{(\alpha, \beta) \in \mathbb{R}^2 \,:\, \alpha > 0, \beta > 0\}$ | $\frac{\alpha\beta}{\alpha - 1}$ (if $\alpha > 1$) | $\frac{\alpha\beta^2}{(\alpha-2)(\alpha-1)^2}$ (if $\alpha > 2$) | Distribution of positive random quantities with heavy tails | Conjugate prior for uniform data with known lower bound |
| Exponential | $\mathrm{Ex}(x \mid \lambda)$ | $f(x) = \lambda e^{-\lambda x}$ <br> $\mathcal{X} = \mathbb{R}_+$ <br> $\Lambda = \mathbb{R}_+$ | $\frac{1}{\lambda}$ | $\frac{1}{\lambda^2}$ | Inter-event times for Poisson Process. Models lifetimes of non-ageing items. | Also parameterised in terms of the mean $\phi = 1/\lambda$. |
| Gamma | $\mathrm{Ga}(x \mid \alpha, \beta)$ | $f(x) = \frac{\beta^\alpha x^{\alpha-1} e^{-\beta x}}{\Gamma[\alpha]}$ <br> $\mathcal{X} = \mathbb{R}_+$ <br> $\Theta = \{(\alpha, \beta) \in \mathbb{R}^2 \,:\, \alpha > 0, \beta > 0\}$ | $\frac{\alpha}{\beta}$ | $\frac{\alpha}{\beta^2}$ | Times between $k$ events for Poisson Process. Lifetimes of ageing items. Conjugate prior for exponential model. | Also parameterised in terms of $1/\beta$ <br> $\mathrm{Ga}(x \mid 1, \lambda) \equiv \mathrm{Ex}(x \mid \lambda)$, <br> $1/x = y \sim \mathrm{IGa}(y \mid \alpha, \beta)$ |
| Beta | $\mathrm{Be}(x \mid \alpha, \beta)$ | $f(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{\mathrm{B}(\alpha, \beta)}$ <br> $\mathcal{X} = (0, 1)$ <br> $\Theta = \{(\alpha, \beta) \in \mathbb{R}^2 \,:\, \alpha > 0, \beta > 0\}$ | $\mu = \frac{\alpha}{\alpha + \beta}$ | $\frac{\mu(1-\mu)}{(\alpha + \beta + 1)}$ | 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) | $\mathrm{N}(x \mid \mu, \sigma^2)$ | $f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x - \mu}{\sigma}\right)^2\right]$ <br> $\mathcal{X} = \mathbb{R}$ <br> $\Theta = \{(\mu, \sigma^2) \in \mathbb{R}^2 \,:\, \sigma^2 > 0\}$ | $\mu$ | $\sigma^2$ | Empirically and theoretically (via CLT) a useful model. <br> Also parameterised in terms of the precision $\lambda = 1/\sigma^2$ | $Y = a + bX \sim \mathrm{N}(y \mid a + b\mu, b^2\sigma^2)$ <br> $Z = \frac{X-\mu}{\sigma} \sim \mathrm{N}(z \mid 0, 1)$ <br> $\mathrm{P}[X \in (u, v)] = \mathrm{P}\left[Z \in \left(\frac{u - \mu}{\sigma}, \frac{v - \mu}{\sigma}\right)\right]$ |
| Student $t$ | $\mathrm{St}(x \mid \mu, \lambda, \nu)$ | $f(x) = \frac{\Gamma[(\nu+1)/2]}{\Gamma[\nu/2]}\left(\frac{\lambda}{\nu\pi}\right)^{1/2} \times$ <br> $\left(1 + \frac{\lambda}{\nu}(x - \mu)^2\right)^{-(\nu+1)/2}$ <br> $\mathcal{X} = \mathbb{R}, \quad \mu \in \mathbb{R}, \lambda, \nu > 0$ | $\mu$ (if $\nu > 1$) | $\lambda^{-1}\frac{\nu}{\nu - 2}$ (if $\nu > 2$) | Useful alternative to Gaussian for random quantities with heavy tails or possible outliers | $Z = \sqrt{\lambda}(x - \mu) \sim t_\nu(z)$ <br> $\mathrm{P}[X \in (u, v)] =$ <br> $\mathrm{P}\left[Z \in \left(\sqrt{\lambda}(u - \mu), \sqrt{\lambda}(v - \mu)\right)\right]$ <br> If $W \sim \mathrm{N}(x \mid 0, 1)$ and $Y \sim \chi^2_{(\nu)}(y)$ ind. then $Z = \frac{W}{\sqrt{Y/\nu}} \sim \mathrm{t}_\nu(z)$. <br> $t_1 \equiv$ Cauchy. $t_\nu^2 \equiv \mathrm{F}_{1,\nu}$. |

## 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 | $\mathrm{Mu}(\boldsymbol{x} \mid \boldsymbol{\theta}, n)$ | $p(\boldsymbol{x}) = \frac{n!}{\prod_{l=1}^k x_l!} \prod_{l=1}^k \theta_l^{x_l}$ <br> $\boldsymbol{x} = \{x_1, \ldots, x_k\}, \quad x_l = 0, 1, \ldots, \sum x_l = n$ <br> $\boldsymbol{\theta} = \{\theta_1, \ldots, \theta_k\}, \quad 0 < \theta_l < 1, \ \sum \theta_l = 1$ | $\mathbb{E}[x_i] = n\theta_i$ | $\mathbb{V}[x_i] = n\theta_i(1-\theta_i)$ <br> $\mathrm{Cov}[x_i, x_j] = -n\theta_i\theta_j$ | Counts of events with more that two possible outcomes | Generalisation of the Binomial distribution |
| Dirichlet | $\mathrm{Di}(\boldsymbol{x} \mid \boldsymbol{\alpha})$ | $f(\boldsymbol{x}) = \frac{\Gamma(\sum \alpha_l)}{\prod \Gamma(\alpha_l)} \prod_{l=1}^k x_l^{\alpha_l - 1}$ <br> $\boldsymbol{x} = \{x_1, \ldots, x_k\}, 0 < x_l < 1, \sum_{l=1}^k x_l = 1$ <br> $\boldsymbol{\alpha} = \{\alpha_1, \ldots, \alpha_k\}, 0 < \alpha_l$ | $\mathbb{E}[x_i] = \mu_i = \frac{\alpha_i}{\sum \alpha_l}$ | $\mathbb{V}[x_i] = \frac{\mu_i(1-\mu_i)}{1 + \sum \alpha_l}$ <br> $\mathrm{Cov}[x_i, x_j] = -\frac{\mu_i \mu_j}{1 + \sum \alpha_l}$ | Distribution of probabilities of exclusive events. | Generalisation of the Beta distribution. Conjugate prior for multinomial data |
| Normal-Gamma | $\mathrm{NG}(x, y \mid \mu, \kappa, \alpha, \beta)$ | $f(x, y) = \mathrm{N}(x \mid \mu, (y\kappa)^{-1}) \mathrm{Ga}(y \mid \alpha, \beta)$ <br> $\mathcal{X} = \{(x, y) : x \in \mathbb{R}, y > 0\}$ <br> $\mu \in \mathbb{R};\ \kappa, \alpha, \beta > 0$ | $\mathbb{E}[x] = \mu$ <br> $\mathbb{E}[y] = \frac{\alpha}{\beta}$ | $\mathbb{V}[x] = \frac{\beta}{\kappa(\alpha - 1)}$ <br> $\mathbb{V}[y] = \frac{\alpha}{\beta^2}$ | 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 | $\mathrm{N}_k(\boldsymbol{x} \mid \boldsymbol{\mu}, \Lambda)$ | $f(\boldsymbol{x}) = \frac{\lvert\Lambda\rvert^{1/2}}{(2\pi)^{k/2}} \exp\left[-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})' \Lambda (\boldsymbol{x} - \boldsymbol{\mu})\right]$ <br> $\mathcal{X} = \mathbb{R}^k$ <br> $\boldsymbol{\mu} \in \mathbb{R}^k;\ \Lambda$ symmetric positive-definite | $\boldsymbol{\mu}$ | $\Lambda^{-1}$ | See univariate case | Usually parameterised in terms of the covariance matrix $\Sigma = \Lambda^{-1}$ |
| (Multivariate) Student | $\mathrm{St_k}(\boldsymbol{x} \mid \boldsymbol{\mu}, \Lambda, \nu)$ | $f(\boldsymbol{x}) = \frac{\lvert\Lambda\rvert^{1/2} \Gamma((\nu + k)/2)}{(\nu\pi)^{k/2} \Gamma(\nu/2)} \times \left[1 + \frac{1}{\nu}(\boldsymbol{x} - \boldsymbol{\mu})' \Lambda (\boldsymbol{x} - \boldsymbol{\mu})\right]^{-(\nu + k)/2}$ <br> $\mathcal{X} = \mathbb{R}^k$ <br> $\boldsymbol{\mu} \in \mathbb{R}^k;\ \Lambda$ symmetric positive-definite , $\nu > 0$ | $\boldsymbol{\mu}$ <br> (if $\nu > 1$) | $\frac{\nu}{\nu - 2} \Lambda^{-1}$ <br> (if $\nu > 2$) | See univariate case | Usually parameterised in terms of the covariance matrix $\Sigma = \Lambda^{-1}$ |
| Wishart | $\mathrm{Wi}_k(X \mid \alpha, \Omega)$ | $f(X) = \frac{(\pi)^{k(k-1)} \lvert\Omega\rvert^{\alpha}}{\prod_{i=1}^k \Gamma[(2\alpha + 1 - i)/2]} \times \lvert X\rvert^{\alpha - (k+1)/2} \exp[-\mathrm{tr}(\Omega X)]$ <br> $\mathcal{X} = (x_{ij})$ symmetric positive-definite <br> $\alpha > (k-1)/2;\ \Omega$ symmetric non-singular | $\alpha\Omega^{-1}$ <br> $\Omega = (\omega_{ij})$ | $\mathbb{V}[X_{ij}] = \alpha\left(\omega_{ij}^2 + \omega_{ii}\omega_{jj}\right)$ | Conjugate prior for the precision matrix in a Gaussian model | Can also be used for the covariance matrix after the appropriate transformation. |