Data provided: Tables of distribution functions

MAS6004

The University Of Sheffield.

SCHOOL OF MATHEMATICS AND STATISTICS **Spring Semester 2019–2020**

Inference **2 hours**

This is an open book exam. Answer ***all*** *questions.*
The submission deadline is 10 am (BST), twenty-four hours after it is released. Late submission will not be considered without extenuating circumstances. It is expected that you will be able to complete this exam in approximately two hours and it is recommended that you submit the work within four hours. You will not be penalised for taking longer, however.
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. You will not get full marks if you simply write down output from a computer package.
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.
For reference *on completing the square from a quadratic equation:*

$$ax^2 - 2bx + c = a(x+d)^2 + e, \quad \text{where} \quad d = \frac{b}{a} \quad \text{and} \quad e = c - \frac{b^2}{a}.$$

On simplifying a sum of squares:

$$\sum_{j=1}^{m}(z_i - a)^2 = m\left(s_z^2 + (\bar{z} - a)^2\right), \quad \text{where} \quad s_z^2 = \frac{1}{m}\sum_{j=1}^{m}(z_i - \bar{z})^2, \quad \bar{z} = \frac{1}{m}\sum_{j=1}^{m} z_i.$$

1 (i) Five observations of the random variable X are recorded:

$$\{0.02, 1.04, 3.63, 1.04, 1.67\},$$

and we are interested in the mean of X, given by $m(X)$.

(a) Sketch the estimated cumulative distribution function (ECDF) of X based on the observed sample. *(3 marks)*

(b) Produce a single bootstrap sample, m_1^*, of $m(X)$ by sampling from the ECDF of X, using the following five random draws from the $Unif(0, 1)$ distribution:

$$\{0.04, 0.52, 0.97, 0.17, 0.63\}.$$

(3 marks)

(c) Explain the approach you would then take to perform a bootstrap estimate of the standard error of $m(X)$. *(4 marks)*

(ii) Let X be the random variable from a $Beta(2, 3)$ distribution, with probability density function

$$f(x) = \frac{x(1-x)^2}{B(2,3)},$$

for $x \in [0, 1]$ and where $B(\cdot, \cdot)$ is the Beta function.

(a) Describe how you would use importance sampling to estimate $\mathbb{E}(h(X))$, where $h(x) = \exp(-x)$, using a $Gam(2, 1)$ proposal distribution.

Hint: the probability density function of a $Gam(2, 1)$ random variable is $g(x) = x \exp(-x)$, for $x > 0$. *(4 marks)*

(b) This importance sampling approach is implemented to generate 100 samples $X_1, \ldots, X_{100}$. The importance weights, $w(X_i)$, are summarised by the sample variance being equal to 10^4, and

$$\sum_{i=1}^{100} w(X_i)^2 \approx 10^6.$$

Comment on the suitability of the proposal distribution in this sampling approach. *(1 mark)*

 Question 1 continued on next page

1 (continued)

(iii) Given a dataset of 10 observation pairs $(x_1, y_1), \ldots, (x_{10}, y_{10})$, the following analysis was carried out in R.

```
> head(obs)
      x     y
1  1.00 17.04
2  6.44 26.96
3 11.89 28.10
4 17.33 53.02
5 22.78 77.92
6 28.22 79.35
>
> error.lm1 <- rep(NA, 10)
> error.lm2 <- rep(NA, 10)
>
> for(i in 1:10){
+   obs.reduced <- obs[-i,]
+
+   lm1 <- lm(y ~ x + I(log(x)), data = obs.reduced)
+   lm2 <- lm(y ~ x + I(x^2), data = obs.reduced)
+
+   lm1.pred <- predict(lm1, newdata = obs[i, ])
+   lm2.pred <- predict(lm2, newdata = obs[i, ])
+
+   error.lm1[i] <- (lm1.pred - obs[i, ]$y)^2
+   error.lm2[i] <- (lm2.pred - obs[i, ]$y)^2
+ }
>
> mean(error.lm1)
[1] 148.4239
> mean(error.lm2)
[1] 179.8139
```

(a) Identify the name of the above approach and describe the general method implemented. *(2 marks)*

(b) What is the purpose of this method and the motivation for using it in this scenario? *(2 marks)*

(c) Given only the information in this analysis, write the mathematical model that should be preferred for this dataset. *(1 mark)*

2 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, $\phi_g = \theta_g/(1 - \theta_g)$, $g = t, c$.

(a) Write down the conjugate prior for each group with mean 1/2 and variance 1/8. *(2 marks)*

(b) Show that the posterior expected value of the odds ratio, $\mathbb{E}\left[\frac{\phi_t}{\phi_c} \,\middle|\, n_t, n_c, x_t, x_c\right] = 2.072$. Explain if it can be concluded, solely on the basis of this number, that the treatment is effective. *(7 marks)*

(c) Provide an expression for the predictive probability of five or more patients improving in a new treatment group of size 10. The expression may involve factorials, beta and gamma functions, but should not contain integrals. *(6 marks)*

3 Consider the simple regression model with random effects,

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

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

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

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

(i) Show that the full conditional posterior distribution of the slope, β, is Gaussian and provide explicit expressions for the parameters. *(3 marks)*

(ii) Considering that the full conditional distribution of

- the location hyper-parameter is $\mathrm{N}(\mu \,|\, m', 1/q')$, with $m' = (np\bar{\alpha} + qm)/(np + q)$ and $q' = np + q$;
- the individual intercepts is $\mathrm{N}(\alpha_i \,|\, a', 1/p')$, with $p' = \lambda + p$ and $a' = (\lambda(y_i - \beta x_i) + p\mu)/(\lambda + p)$; and
- the precision is $\mathrm{Ga}(\lambda \,|\, c', d')$, with $c' = c + n/2$ and $d' = d + \sum (y_i - \alpha_i - \beta x_i)^2/2$.

Write pseudo-code for an MCMC sampling scheme for exploring the posterior distribution. *(7 marks)*

(iii) Assuming you have run the code above, explain how would you use the output to estimate a posterior interval of probability 0.95 for the slope, β. *(4 marks)*

(iv) Write pseudo-code for exploring the predictive distribution $f(y_{n+1} | \boldsymbol{x}, x_{n+1})$. *(6 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, 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,

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

respectively stand for the 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. Experiments 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	$\mathrm{Ge}(x \mid \theta)$	$p(x) = \theta(1-\theta)^x$ $\mathcal{X} = 0, 1, 2, ...$ $\Theta = (0, 1)$	$\dfrac{1-\theta}{\theta}$	$\dfrac{1-\theta}{\theta^2}$	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	$\mathrm{Po}(x \mid \lambda)$	$p(x) = \dfrac{e^{-\lambda}\lambda^x}{x!}$ $\mathcal{X} = 0, 1, 2, ...$ $\Lambda = \mathbb{R}^+$	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 $n\theta = c$.
Negative binomial (Pascal)	$\mathrm{NB}(x \mid m, \theta)$	$p(x) = \binom{m+x-1}{x}\theta^m (1-\theta)^x$ $\mathcal{X} = 0, 1, 2, ...$ $\Theta = (0, 1)$	$\dfrac{m(1-\theta)}{\theta}$	$\dfrac{m(1-\theta)}{\theta^2}$	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)	$p(x) = \dfrac{\binom{d}{x}\binom{N-d}{n-x}}{\binom{N}{n}}$ $\mathcal{X} = \{a, a+1, \ldots, b\}$ $a = \max\{0, n+d-N\}$, $b = \min\{n, d\}$	$\dfrac{nd}{N}$	$\dfrac{nd}{N}\dfrac{N-n}{N-1}\left(1 - \dfrac{d}{N}\right)$	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	$\mathrm{Un}(x \mid \alpha, \beta)$	$f(x) = \frac{1}{\beta - \alpha}$ $\mathcal{X} = [\alpha, \beta]$ $\Theta = \{(\alpha, \beta) \in \mathbb{R}^2 \,:\, \alpha < \beta\}$	$\frac{\alpha + \beta}{2}$	$\frac{(\beta - \alpha)^2}{12}$	Rounding errors $\mathrm{Un}(x \mid -1/2, 1/2)$. Simulating other distributions from $\mathrm{Un}(x \mid 0, 1)$	Used as non-informative prior for parameters with bounded support.
Pareto	$\mathrm{Pa}(x \mid \alpha, \beta)$	$f(x) = \alpha\beta^\alpha x^{-(\alpha+1)}$ $\mathcal{X} = (\beta, \infty)$ $\Theta = \{(\alpha, \beta) \in \mathbb{R}^2 \,:\, \alpha > 0, \beta > 0\}$	$\frac{\alpha\beta}{\alpha - 1}$ (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	$\mathrm{Ex}(x \mid \lambda)$	$f(x) = \lambda e^{-\lambda x}$ $\mathcal{X} = \mathbb{R}_+$ $\Lambda = \mathbb{R}_+$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	Inter-event times for Poisson Process. Models lifetimes of non-ageing items.	Also parameterised in terms of the mean $\phi = 1/\lambda$. $\mathrm{Ga}(x \mid 1, \lambda) \equiv \mathrm{Ex}(x \mid \lambda)$
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 $1/\beta$ $\mathrm{Ga}(x \mid 1, \lambda) \equiv \mathrm{Ex}(x \mid \lambda)$, $1/x = y \sim \mathrm{IGa}(y \mid \alpha, \beta)$
Beta	$\mathrm{Be}(x \mid \alpha, \beta)$	$f(x) = \frac{x^{\alpha-1}(1 - x)^{\beta-1}}{\mathrm{B}(\alpha, \beta)}$ $\mathcal{X} = (0, 1)$ $\Theta = \{(\alpha, \beta) \in \mathbb{R}^2 \,:\, \alpha > 0, \beta > 0\}$	$\mu = \frac{\alpha}{\alpha + \beta}$	$\frac{\mu(1 - \mu)}{(\alpha + \beta + 1)}$	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)	$\mathrm{N}(x \mid \mu, \sigma^2)$	$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x - \mu}{\sigma}\right)^2\right]$ $\mathcal{X} = \mathbb{R}$ $\Theta = \{(\mu, \sigma^2) \in \mathbb{R}^2 \,:\, \sigma^2 > 0\}$	μ	σ^2	Empirically and theoretically (via CLT) a useful model. Also parameterised in terms of the precision $\lambda = 1/\sigma^2$	$Y = a + bX \sim \mathrm{N}(y \mid a + b\mu, b^2\sigma^2)$ $Z = \frac{X-\mu}{\sigma} \sim \mathrm{N}(z \mid 0, 1)$ $\mathrm{P}[X \in (u, v)] = \mathrm{P}\left[Z \in \left(\frac{u - \mu}{\sigma}, \frac{v - \mu}{\sigma}\right)\right]$
Student t	$\mathrm{St}(x \mid \mu, \lambda, \nu)$	$f(x) = \frac{\Gamma[(\nu + 1)/2]}{\Gamma[\nu/2]}\left(\frac{\lambda}{\nu\pi}\right)^{1/2} \times$ $\left(1 + \frac{\lambda}{\nu}(x - \mu)^2\right)^{-(\nu+1)/2}$ $\mathcal{X} = \mathbb{R}, \quad \mu \in \mathbb{R}, \lambda, \nu > 0$	μ (if $\nu > 1$)	$\lambda^{-1}\frac{\nu}{\nu - 2}$ (if $\nu > 2$)	Useful alternative to Gaussian for random quantities with heavy tails	$Z = \sqrt{\lambda}(x - \mu) \sim t_\nu(z)$ If $X \sim \mathrm{N}(x \mid 0, 1)$ and $Y \sim \chi^2_{(\nu)}(y)$ ind. then $Z = \frac{X}{\sqrt{Y/\nu}} \sim \mathrm{t}_\nu(z)$. $t_1 \equiv$ 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	$\mathrm{Mu}(\boldsymbol{x} \mid \boldsymbol{\theta}, n)$	$p(\boldsymbol{x}) = \frac{n!}{\prod_{l=1}^k x_l!} \prod_{l=1}^k \theta_l^{x_l}$ $\boldsymbol{x} = \{x_1, \ldots, x_k\}, \quad x_l = 0, 1, \ldots, \sum x_l = n$ $\boldsymbol{\theta} = \{\theta_1, \ldots, \theta_k\}, \ 0 < \theta_l < 1, \ \sum \theta_l = 1$	$\mathbb{E}[x_i] = n\theta_i$	$\mathbb{V}[x_i] = n\theta_i(1-\theta_i)$ $\mathrm{Cov}[x_i, x_j] = -n\theta_i\theta_j$	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 probabilities of exclusive events.	Generalisation of the Beta distribution. Conjugate prior for multinomial data
Normal-Gamma	$\mathrm{NG}(x, y \mid \mu, \kappa, \alpha, \beta)$	$f(x, y) = \mathrm{N}(x \mid \mu, (y\kappa)^{-1}) \mathrm{Ga}(y \mid \alpha, \beta)$ $\mathcal{X} = \{(x, y) \,:\, x \in \mathbb{R}, y > 0\}$ $\mu \in \mathbb{R};\ \kappa, \alpha, \beta > 0$	$\mathbb{E}[x] = \mu$ $\mathbb{E}[y] = \frac{\alpha}{\beta}$	$\mathbb{V}[x] = \frac{\beta}{\kappa(\alpha - 1)}$ $\mathbb{V}[y] = \frac{\alpha}{\beta^2}$	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	$\mathrm{N}_k(\boldsymbol{x} \mid \boldsymbol{\mu}, \Lambda)$	$f(\boldsymbol{x}) = \frac{\lvert\Lambda\rvert^{1/2}}{(2\pi)^{k/2}} \exp\left[-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})' \Lambda (\boldsymbol{x} - \boldsymbol{\mu})\right]$ $\mathcal{X} = \mathbb{R}^k$ $\boldsymbol{\mu} \in \mathbb{R}^k$; Λ symmetric positive-definite	$\boldsymbol{\mu}$	Λ^{-1}	See univariate case	Usually parameterised in terms of the covariance matrix $\Sigma = \Lambda^{-1}$
(Multivariate) Student	$\mathrm{St_k}(\boldsymbol{x} \mid \boldsymbol{\mu}, \Lambda, \nu)$	$f(\boldsymbol{x}) = \frac{\lvert\Lambda\rvert^{1/2} \Gamma((\nu + k)/2)}{(\nu\pi)^{k/2} \Gamma(\nu/2)} \times$ $\left[1 + \frac{1}{\nu}(\boldsymbol{x} - \boldsymbol{\mu})' \Lambda (\boldsymbol{x} - \boldsymbol{\mu})\right]^{-(\nu+k)/2}$ $\mathcal{X} = \mathbb{R}^k$ $\boldsymbol{\mu} \in \mathbb{R}^k$; 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 $\Sigma = \Lambda^{-1}$
Wishart	$\mathrm{Wi}_k(X \mid \alpha, \Omega)$	$f(X) = \frac{(\pi)^{k(k-1)} \lvert\Omega\rvert^{\alpha}}{\prod_{i=1}^k \Gamma[(2\alpha + 1 - i)/2]} \times$ $\lvert X\rvert^{\alpha - (k+1)/2} \exp[-\mathrm{tr}(\Omega X)]$ $\mathcal{X} = (x_{ij})$ symmetric positive-definite $\alpha > (k-1)/2$; Ω symmetric non-singular	$\alpha\Omega^{-1}$ $\Omega = (\omega_{ij})$	$\mathbb{V}[X_{ij}] = \alpha\left(\omega_{ij}^2 + \omega_{ii}\omega_{jj}\right)$	Conjugate prior for the precision matrix in a Gaussian model	Can also be used for the covariance matrix after the appropriate transformation.