Data provided: Tables of distribution functions

MAS61006

The University Of Sheffield.

SCHOOL OF MATHEMATICS AND STATISTICS **Spring Semester 2020–2021**

Bayesian Statistics and Computational Methods

This is an open book exam.
Answer all questions.
You can work on the exam during the 24 hour period starting from 10am (BST), and you must submit your work within 3.5 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.*** *Calculations should be performed by hand. A university-approved calculator may be used. The use of any other calculational device, software or service is not permitted. 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 A polling company is hired by one of the two candidates for the South Yorkshire Police and Crime Commissioner post to measure voting intention, θ in the elections next year. From previous surveys, the expected vote intention for this candidate was 0.33 and it is believed to be within $(0.24, 0.42)$ with probability 0.5.

(i) Fit a suitable Beta distribution for θ to represent this prior information. You may use R to answer this question. *(4 marks)*

(ii) The company will design a simple survey to ask a sample of the population which of the two candidates they will vote for and record the answer,

$$x_i = \begin{cases} 1 & \text{voting for this candidate} \\ 0 & \text{otherwise} \end{cases}.$$

(a) 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 θ in the model. *(3 marks)*

(b) Assume $\boldsymbol{x} = \{x_1, \ldots, x_n\}$ is recorded. 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)*

(c) What is the smallest sample size, n, needed so that the posterior variance is no more than 50% of the prior variance? *(8 marks)*

2 A new silicon chip production line is known to produce faulty chips at a constant daily rate, λ. A quality control engineer selects n fractions of different days, $\{c_1, \dots, c_n\}$, at random for inspection and records the number of faulty chips, x_i.

(i) Assuming $x_i \sim \mathrm{Po}(x_i \mid c_i\lambda)$, show that $\mathrm{Ga}(\lambda \mid a, b)$ is a conjugate prior and provide explicit expressions for the posterior parameters. *(3 marks)*

(ii) The data recorded are,

c_i	0.1	0.1	0.25	0.05	0.4	0.5	0.1	0.15	0.2	0.1
x_i	7	10	15	3	30	32	8	5	6	7

and the engineer's prior mean, 3, and variance, 10. You may use R to answer this question.

(a) Provide an equally tailed posterior interval of probability 0.9. *(3 marks)*

(b) Provide a measure of the strength of evidence in favour of $\lambda > 60$. *(7 marks)*

(c) Provide an approximate predictive interval of probability 0.95 for the number of faulty chips in a whole day. *(7 marks)*

[HINT: For $Z \sim \mathrm{N}(z \mid 0, 1)$, $\mathrm{P}[Z < -0.674] = 0.25$, $\mathrm{P}[Z < 1] = 0.841$, $\mathrm{P}[Z < 1.645] = 0.95$, $\mathrm{P}[Z < 1.96] = 0.975$, $\mathrm{P}[Z < 2.58] = 0.995$.]

3 In order to be effective, surgical masks should be able to filter out small droplets in the air which may contain pathogens. A key parameter in the design process is the mean size, μ, of the infectious cells —measured in nm. Based on preliminary lab results and knowledge of related cells, it is thought that μ lies almost certainly between 50 and 140.

(i) Find a suitable Gaussian distribution to represent this prior distribution for μ. *(3 marks)*

(ii) A single measurement of the size of the pathogen, x, is taken using equipment known to have Gaussian errors with mean zero and standard deviation 4. State the posterior distribution for μ after observing x. *(4 marks)*

(iii) If $x = 135$, calculate the posterior mean and variance for μ. Find values z_L and z_U such that the posterior probabilities for $\mu < z_L$ and $\mu > z_U$ are both equal to 0.25. How do these values compare with the corresponding quantiles of the prior distribution? You may use R to answer this question. *(8 marks)*

(iv) Using the current information, provide a posterior predictive interval of probability 0.9 for a future independent measurement, y. You may use R to answer this question. *(5 marks)*

[HINT: For $Z \sim \mathrm{N}(z \mid 0, 1)$, $\mathrm{P}[Z < -0.674] = 0.25$, $\mathrm{P}[Z < 1] = 0.841$, $\mathrm{P}[Z < 1.645] = 0.95$, $\mathrm{P}[Z < 1.96] = 0.975$, $\mathrm{P}[Z < 2.58] = 0.995$.]

4 (i) Consider a dataset from a study on 62 mammals concerning the relationship between sleeping patterns (`sws`) and three ecological variables (`bw`, `mls`, `odi`). The output below shows the structure of the data:

```
> str(sleep)
`data.frame': 62 obs. of  4 variables:
 $ sws: num  NA 6.3 NA NA 2.1 9.1 15.8 5.2 10.9 8.3 ...
 $ bw : num  6654 1 3.38 0.92 2547 ...
 $ mls: num  38.6 4.5 14 NA 69 27 19 30.4 28 50 ...
 $ odi: int  3 3 1 3 4 4 1 4 1 1 ...

> head(sleep)
  sws       bw  mls odi
1  NA 6654.000 38.6   3
2 6.3    1.000  4.5   3
3  NA    3.385 14.0   1
4  NA    0.920   NA   3
5 2.1 2547.000 69.0   4
6 9.1   10.550 27.0   4
```

The following R command is used:

```
sleep_mice <- mice(sleep,
                   m=5, maxit = 10,
                   method = c(`norm'))
```

(a) Describe, in detail, the statistical procedure that has been implemented here. Your description should reference the parameters in the above R command and its output, along with how uncertainty is handled. *(4 marks)*

 Question 4 continued on next page

4 (continued)

(b) Interest lies in the regression model

$$\texttt{sws} = \beta_0 + \beta_1 \log_{10}(\texttt{bw}) + \beta_2 \texttt{mls} + \beta_4 \texttt{odi}.$$

Use the information below to estimate the coefficient of `odi`, given by β_4. Provide this estimate in the form of an approximate 95% interval.

```
> fit_mice <- with(sleep_mice,
                   lm(sws ~ log10(bw) + mls + odi))

> (coefs <- sapply(fit_mice$analyses, coef))
              [,1]   [,2]   [,3]  [,4]   [,5]
(Intercept) 13.000 12.000 12.000 12.00 12.000
log10(bw)   -0.940 -0.760 -0.860 -0.85 -1.000
mls         -0.038 -0.034 -0.043 -0.03 -0.039
odi         -1.100 -0.960 -0.840 -0.87 -0.930

> sapply(fit_mice$analyses, function(x) vcov(x)[4,4])
[1] 0.0899 0.0911 0.1010 0.1020 0.0893
```

(3 marks)

(c) What proportion of the uncertainty in the estimate of β_4 is due to the missing data in the `sleep` dataset? Give your answer as a percentage. *(1 mark)*

(d) What can you conclude about the effect of the danger index of a mammal from other animals (`odi`) on their amount of sleep (`sws`)? *(2 marks)*

(ii) An analysis was carried out in R on the data `X`, involving 100 observations, shown in Figure 1.

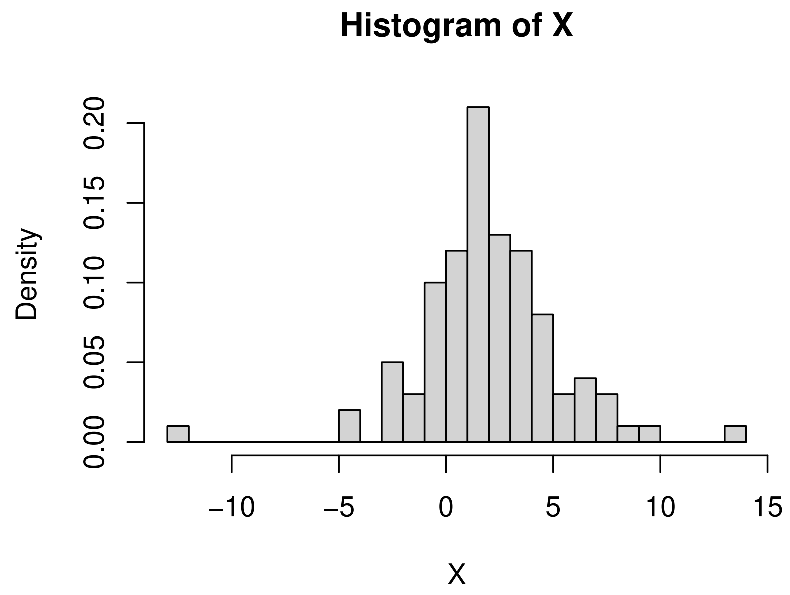

Figure 1: Histogram of the variable X.

4 (continued)

(a) The following R commands are implemented:

```
> fn_1 <- function(B, X, trim){
+   replicate(B, {
+     Xsim <- sample(X, replace = T)
+     mean(Xsim, trim = trim)
+   })
+ }
> result_1 <- fn_1(10^4, X, 0)
> quantile(result_1, c(0.025, 0.975))
    2.5%    97.5%
1.367235 2.613829
```

Name and carefully describe this procedure. *(5 marks)*

4 (continued)

(b) A second implementation is run:

```
> fn_2 <- function(B, X, trim){
+   replicate(B, {
+     Xsim <- rnorm(length(X), mean(X), sd(X))
+     mean(Xsim, trim = trim)
+   })
+ }
> result_2 <- fn_2(10^2, X, 0)
> quantile(result_2, c(0.025, 0.975))
    2.5%    97.5%
1.433099 2.686094
```

Explain how the implementation in `fn_2` differs from that in `fn_1`. Which implementation is more appropriate for the data `X`?

(3 marks)

(c) Why might the following implementation be considered more robust?

```
> result_3 <- fn_1(10^4, X, 0.25)
> quantile(result_3, c(0.025, 0.975))
    2.5%    97.5%
1.403487 2.409583
```

(2 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 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, uppercase matrix valued.

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 $x \sim \mathrm{N}\big(x \bigm| \mu, \sigma^2\big)$. 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	$\mathrm{Ber}(x \mid \theta)$	$p(x) = \theta^x(1-\theta)^{1-x}$ $\mathcal{X} = \{0,1\}$ $\Theta = (0,1)$	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	$\mathrm{Un}(x \mid \alpha, \beta)$	$f(x) = \dfrac{1}{\beta - \alpha}$ $\mathcal{X} = [\alpha, \beta]$ $\Theta = \{(\alpha, \beta) \in \mathbb{R}^2 \,:\, \alpha < \beta\}$	$\dfrac{\alpha + \beta}{2}$	$\dfrac{(\beta - \alpha)^2}{12}$	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 the mean $\phi = 1/\lambda$.
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 or possible outliers	$Z = \sqrt{\lambda}(x - \mu) \sim t_\nu(z)$ $\mathrm{P}[X \in (u, v)] =$ $\mathrm{P}\left[Z \in \left(\sqrt{\lambda}(u-\mu), \sqrt{\lambda}(v-\mu)\right)\right]$ If $W \sim \mathrm{N}(x \mid 0, 1)$ and $Y \sim \chi^2_{(\nu)}(y)$ ind. then $Z = \frac{W}{\sqrt{Y/\nu}} \sim \mathrm{t}_\nu(z)$. $t_1 \equiv$ 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	$\mathrm{Mu}(\boldsymbol{x} \mid \boldsymbol{\theta}, n)$	$p(\boldsymbol{x}) = \frac{n!}{\prod_{l=1}^k x_l!} \prod_{l=1}^k \theta_l^{x_l}$ $\boldsymbol{x} = \{x_1, \ldots, x_k\}, \quad x_l = 0, 1, \ldots, \sum x_l = n$ $\boldsymbol{\theta} = \{\theta_1, \ldots, \theta_k\}, \quad 0 < \theta_l < 1, \ \sum \theta_l = 1$	$\mathbb{E}[x_i] = n\theta_i$	$\mathbb{V}[x_i] = n\theta_i(1-\theta_i)$ $\mathrm{Cov}[x_i, x_j] = -n\theta_i\theta_j$	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	$\mathrm{NG}(x, y \mid \mu, \kappa, \alpha, \beta)$	$f(x, y) = \mathrm{N}\big(x \mid \mu, (y\kappa)^{-1}\big)\mathrm{Ga}(y \mid \alpha, \beta)$ $\mathcal{X} = \{(x, y) : x \in \mathbb{R}, y > 0\}$ $\mu \in \mathbb{R};\ \kappa, \alpha, \beta > 0$	$\mathbb{E}[x] = \mu$ $\mathbb{E}[y] = \frac{\alpha}{\beta}$	$\mathbb{V}[x] = \frac{\beta}{\kappa(\alpha - 1)}$ $\mathbb{V}[y] = \frac{\alpha}{\beta^2}$	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	$\mathrm{N}_k(\boldsymbol{x} \mid \boldsymbol{\mu}, \Lambda)$	$f(\boldsymbol{x}) = \frac{\lvert\Lambda\rvert^{1/2}}{(2\pi)^{k/2}} \exp\big[-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})'\Lambda(\boldsymbol{x} - \boldsymbol{\mu})\big]$ $\mathcal{X} = \mathbb{R}^k$ $\boldsymbol{\mu} \in \mathbb{R}^k;\ \Lambda$ 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	$\mathrm{Wi}_k(X \mid \alpha, \Omega)$	$f(X) = \frac{(\pi)^{k(k-1)}\lvert\Omega\rvert^{\alpha}}{\prod_{i=1}^k \Gamma[(2\alpha + 1 - i)/2]} \times$ $\lvert X\rvert^{\alpha - (k+1)/2} \exp[-\mathrm{tr}(\Omega X)]$ $\mathcal{X} = \big(x_{ij}\big)$ symmetric positive-definite $\alpha > (k-1)/2;\ \Omega$ symmetric non-singular	$\alpha\Omega^{-1}$ $\Omega = \big(\omega_{ij}\big)$	$\mathbb{V}[X_{ij}] = \alpha\Big(\omega_{ij}^2 + \omega_{ii}\omega_{jj}\Big)$	Conjugate prior for the precision matrix in a Gaussian model	Can also be used for the covariance matrix after the appropriate transformation.