Data provided: Tables of distribution functions

MAS6062

The University Of Sheffield.

SCHOOL OF MATHEMATICS AND STATISTICS **Spring Semester 2016–2017**

Bayesian Methods and Clinical Trials **3 hours**

Candidates may bring to the examination a calculator which conforms to University regulations.
*Marks will be awarded for your best **five** answers. Total marks 100.*

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 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. ***(2 marks)***

(ii) Show that the posterior mean is the optimal decision under square error loss. ***(3 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. ***(5 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. ***(5 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. ***(5 marks)***

2 For these questions **Table 1** can be used to assist in the calculations.

(i) A two arm parallel group superiority randomised controlled trial is to be undertaken to investigate the effect of a new treatment for Migraine compared to a placebo control treatment. The primary endpoint is participant self-reported pain measured by a 100-point Visual Analogue Pain scale at 6 month post-randomisation. For the primary endpoint a target difference of 10 points or more on the pain scale is thought to be of clinical and practical importance. The standard deviation of the pain outcome at 6 month post-randomisation, is anticipated to be 25 points. Assuming a two tailed Type I error of 5% and 90% power using the data and equal numbers per group what is the sample size per arm? ***(2 marks)***

(ii) Take the result from (i) as the sample size but consider this only as the evaluable sample size i.e. the number of subjects required for the required power and statistical significance. Suppose there is an expectation at the start of the study that 15% of the subjects will drop out and not give valid 6 month pain outcome data for analysis. What sample size would be required to ensure the requisite number of evaluable subjects? With 20% drop outs what would the sample size per arm be allowing for drop outs? Comment on your results. ***(2 marks)***

(iii) An alternative design such as cross-over trial is being considered for the same study. In the cross-over design participants will be treated with the new treatment for six months followed by a 1 month wash-out period and the placebo treatment for six months; the order of treatments to placebo or active to be randomised. The target is the same, 10 points, and an estimate of the within subject standard deviation is 20 points. Taking the sample size per arm for a parallel group trial from Table 1 as an estimate for the sample size what is the total sample size this alternative design would require? ***(2 marks)***

(iv) The sample size in (iii) is the evaluable sample size - taken as the number of patients anticipated to complete the cross-over trial. With only 80% of patients anticipated to complete what sample size is required to ensure an evaluable sample size estimated in (iii)? ***(1 mark)***

(v) A third treatment arm of an active control is being considered. Ignoring the effect of multiplicity by having three treatment arms what would be the impact on the sample size in (i) the parallel group design and (iii) the cross-over design. ***(2 marks)***

(vi) What would be the treatment sequences for a three period cross-over study with three treatments? ***(2 marks)***

(vii) If the study design was to be designed as a balanced incomplete block design (BIBD) with 3 treatments in 2 periods what would the treatment sequences be? ***(2 marks)***

2 (continued)

(viii) What would be the impact on the sample size of a the 3 treatment 2 period BIBD design in (vii) compared to the sample size for the design in (vi). ***(1 mark)***

(ix) Comment on the relative merits of the study designs from (i), (iii), (v) and (vi). ***(3 marks)***

(x) Returning back to a two arm parallel group trial. What are the null and alternative hypotheses for a superiority, non-inferiority and equivalence trial? ***(3 marks)***

Table 1. Sample sizes for one group, n_A in a parallel group study for different standardised differences and allocation ratios for 90% power and a two sided Type I error of 5% using a non-central t-distribution

	Allocation ratios			
$\delta = d/s$	1	2	3	4
0.05	8407	6306	5605	5255
0.10	2103	1577	1402	1314
0.15	935	702	624	585
0.20	527	395	351	329
0.25	338	253	225	211
0.30	235	176	157	147
0.35	173	130	115	108
0.40	133	100	89	83
0.45	105	79	70	66
0.50	86	64	57	53
0.55	71	53	47	44
0.60	60	45	40	37
0.65	51	38	34	32
0.70	44	33	30	28
0.75	39	29	26	24
0.80	34	26	23	21
0.85	31	23	20	19
0.90	27	21	18	17
0.95	25	19	17	15
1.00	23	17	15	14

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. ***(4 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. ***(6 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. ***(3 marks)***

(ii) Write down pseudo-code for an MCMC scheme to explore the posterior distribution $\pi(\boldsymbol{\mu}, \eta, \lambda \mid \boldsymbol{y})$. ***(7 marks)***

4 A cluster randomised controlled trial has been undertaken in East Africa where 10 schools were randomised to receive an intervention ($n = 5$) or control ($n = 5$) to help them prevent school absences amongst school girls associated with their periods. The intervention consisted of a menstrual health education programme. The primary outcome was the number of school days missed in a month.

Table 2. Simple analysis of school absence data.

Mean Difference in number of school days missed	p-value	95% Confidence Interval
1.69	0.001	(0.71, 2.67)

A naïve analysis ignoring the effect of school (cluster) is given in Table 2 of the mean difference in number of school days missed in a month. A positive sign means the control group has missed more days.

(i) Comment on the results in Table 2. What is the best estimate of the difference in the number of days missed with a plausible range? ***(3 marks)***

(ii) The analysis was repeated but with the effect of school (cluster) allowed for in the analysis. The results are given in Table 3. Comment on the results in Table 3. How do they compare to the results in Table 2? ***(3 marks)***

Table 3. Simple analysis of school absence data.

Mean Difference in number of school days missed	p-value	95% Confidence Interval
1.50	0.077	(−0.21, 3.17)

(iii) Why are the results in (ii) different to those in (i)? ***(1 mark)***

(iv) Define an intra-class correlation (ICC) in words and explain how it affects the design and analysis of a cluster randomised controlled trial. How has the ICC impacted on the results in (ii)? ***(3 marks)***

(v) What are the main reasons for undertaking a cluster randomised controlled trial compared with an individually randomised controlled trial? ***(2 marks)***

(vi) What are the possible sources of bias in a cluster randomised controlled trial? ***(2 marks)***

(vii) To design the study to detect a target difference of 1.5 days with 90% power and a two sided significance level of 5%; 260 school girls (130 per arm) are required in the study for an individually randomised controlled trial design. A cluster randomised controlled trial design requires 1560 school girls in total. What is the design effect for the cluster randomised controlled trial? ***(2 marks)***

4 (continued)

(viii) If the study in (vii) was designed as an individually randomised controlled trial it would be anticipated that there would be contamination in the control arm such that the effect size is reduced by 40%. What would the sample size be for individually randomised controlled trial with allowance for contamination? ***(2 marks)***

(ix) Following on from (viii), if a further issue is that only 40% of eligible school girls would be expected to consent to enter an individually randomised trial, how does this impact on recruitment and consent considerations? ***(2 marks)***

5 (i) What is the difference between a critical and non-critical path study? ***(2 marks)***

(ii) What types of trials are undertaken in the Phase I, Phase II and Phase III of clinical development for a new pharmaceutical treatment? ***(4 marks)***

(iii) What are the objectives of each phase of development? ***(3 marks)***

(iv) What is the difference between a pilot and feasibility study? ***(2 marks)***

(v) What is being investigated when the pharmacokinetics of a drug is being assessed? ***(2 marks)***

(vi) What is being investigated when the pharmacodynamics of a drug is being assessed? ***(2 marks)***

(vii) If a pharmaceutical alternative test formulation is shown to be bioequivalent when compared to a reference formulation what has been demonstrated? ***(2 marks)***

(viii) What can be concluded with respect to safety and efficacy if a new pharmaceutical alternative is shown to be bioequivalent ***(2 marks)***

(ix) In bioequivalence trials 90% confidence intervals are quoted. What is the Type I error for these studies? ***(1 mark)***

6 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 $y_i \sim \mathrm{Un}(y_i \mid 0, \theta)$, 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. ***(3 marks)***

(ii) The engineer decides to use

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

as a prior distribution.

(a) Sketch the engineer's prior distribution. ***(3 marks)***

(b) Show that her posterior distribution is $\mathrm{Pa}(\theta \mid a^\star, b^\star)$, with $a^\star = n + a$ and $b^\star = \max\left\{b, \widehat{\theta}\right\}$. ***(7 marks)***

(c) Discuss the implications on the Bayesian learning process if $b > \widehat{\theta}$. ***(3 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 $f(\boldsymbol{x} \mid \boldsymbol{\phi})$ 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 Normal distribution with mean μ and variance σ^2, its density is denoted by $\mathrm{N}\left(x \mid \mu, \sigma^2\right)$. Tables below present some density functions used throughout the course.

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

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

respectively stand for the 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)	$\mathrm{U}(1, ..., k)$	$p(x) = 1/k$ $\mathcal{X} = \{1, ..., k\}$, $\mathcal{K} = \mathbb{Z}_+$	$\frac{k+1}{2}$	$\frac{k^2-1}{12}$	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	$\mathrm{Ge}(x \mid \theta)$	$p(x) = \theta(1-\theta)^x$ $\mathcal{X} = 0, 1, 2, ...$ $\Theta = (0, 1)$	$\frac{1-\theta}{\theta}$	$\frac{1-\theta}{\theta^2}$	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	$\mathrm{NB}(x \mid m, \theta)$	$p(x) = \binom{m+x-1}{x}\theta^m(1-\theta)^x$ $\mathcal{X} = 0, 1, 2, ...$ $\Theta = (0, 1)$	$\frac{m(1-\theta)}{\theta}$	$\frac{m(1-\theta)}{\theta^2}$	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	$\mathrm{Po}(x \mid \lambda)$	$p(x) = \frac{e^{-\lambda}\lambda^x}{x!}$ $\mathcal{X} = 0, 1, 2, ...$ $\Lambda = \mathbb{R}^+$	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)	$\mathrm{N}\big(x \,\big|\, \mu, \sigma^2\big)$	$f(x) = \dfrac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\dfrac{x-\mu}{\sigma}\right)^2\right]$ $\mathcal{X} = \mathbb{R}$ $\Theta = \{(\mu, \sigma^2) \in \mathbb{R}^2 \,:\, \sigma^2 > 0\}$	μ	σ^2	Empirically and theoretically (via CLT) a useful model. Often parameterised in terms of the precision $\lambda = 1/\sigma^2$	$Y = aX + b \sim \mathrm{N}\big(y \,\big|\, a\mu + b, a^2\sigma^2\big)$ $Z = \frac{X-\mu}{\sigma} \sim \mathrm{N}(z \mid 0, 1)$ $\mathrm{P}[X \in (u, v)] = \mathrm{P}\Big[Z \in \Big(\dfrac{u-\mu}{\sigma}, \dfrac{v-\mu}{\sigma}\Big)\Big]$
Chi-square	$\chi^2_{(\nu)}(x)$	$f(x) = \dfrac{2^{-\nu/2}}{\Gamma(\nu/2)} x^{\nu/2-1} e^{-x/2}$ $\mathcal{X} = \mathbb{R}_+$; $\Theta = \mathbb{R}_+$	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	$\mathrm{St}(x \mid \mu, \lambda, \nu)$	$f(x) = \frac{\Gamma[(\nu+1)/2]}{\Gamma[\nu/2]}\left(\frac{\lambda}{\nu\pi}\right)^{1/2} \times$ $\big(1 + \lambda(x-\mu)^2/\nu\big)^{-(\nu+1)/2}$ $\mathcal{X} = \mathbb{R}, \;\; \mu \in \mathbb{R}, \lambda, \nu > 0$	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	$\mathrm{NG}(x, y \mid \mu, \lambda, \alpha, \beta)$	$f(x, y) = \mathrm{N}\big(x \,\big|\, \mu, (y\lambda)^{-1}\big)\mathrm{Ga}(y \mid \alpha, \beta)$ $\mathcal{X} = \{(x, y) \,:\, x \in \mathbb{R}, y > 0\}$ $\mu \in \mathbb{R};\ \lambda, \alpha, \beta > 0$	$\mathbb{E}[x] = \mu$ $\mathbb{E}[y] = \alpha\beta^{-1}$	$\mathbb{V}[x] = \frac{\beta}{\lambda(\alpha-1)}$ $\mathbb{V}[y] = \alpha\beta^{-2}$	Conjugate prior for Gaussian data	$f(x) = \mathrm{St}\big(x \,\big|\, \mu, \lambda\alpha\beta^{-1}, 2\alpha\big)$
Gaussian	$\mathrm{N}_k(\boldsymbol{x} \mid \boldsymbol{\mu}, \Lambda)$	$f(\boldsymbol{x}) = \frac{\lvert\Lambda\rvert^{1/2}}{(2\pi)^{k/2}} \exp\big[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})'\Lambda(\boldsymbol{x}-\boldsymbol{\mu})\big]$ $\mathcal{X} = \boldsymbol{x} \in \mathbb{R}^k$ $\boldsymbol{\mu} \in \mathbb{R}^k;\ \Lambda$ 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 $\Sigma = \Lambda^{-1}$