Data provided: Tables of distribution functions

The University Of Sheffield.

SCHOOL OF MATHEMATICS AND STATISTICS **Spring Semester 2015–2016**

MAS6062 Bayesian Methods and Clinical Trials **3 hours**

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

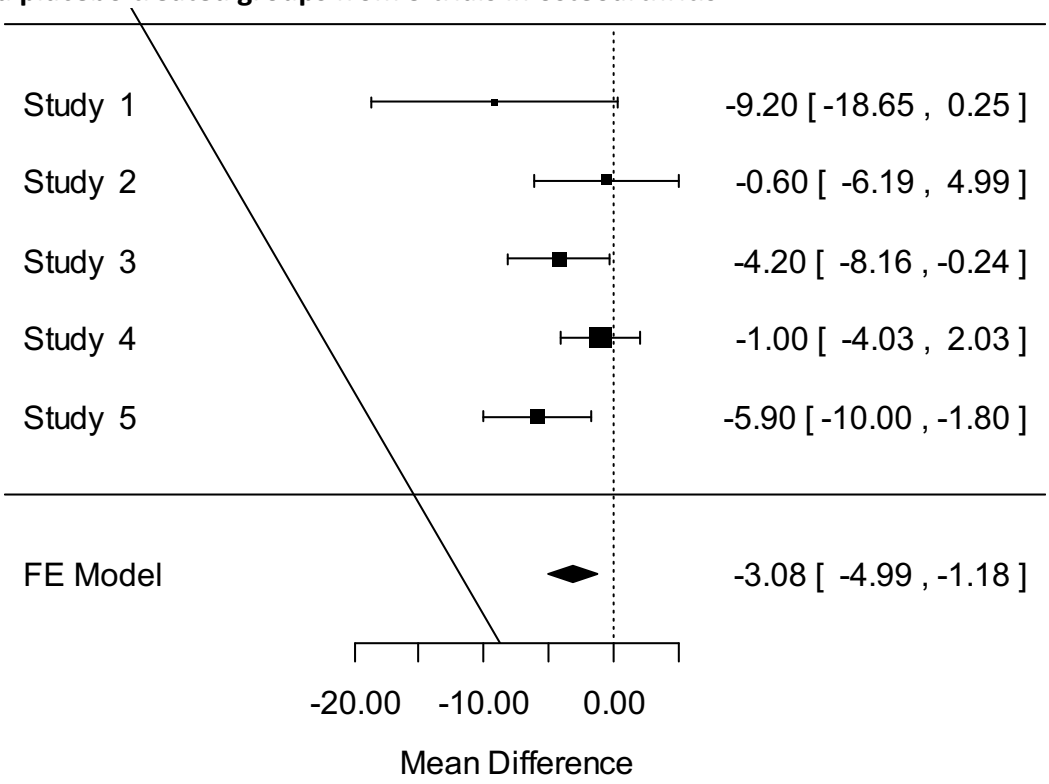
Please leave this exam paper on your desk
Do not remove it from the hall

Registration number from U-Card (9 digits)
to be completed by student

Blank

1 Machado et al (BMJ 2015) conducted a meta-analysis of paracetamol against placebo for a number of clinical conditions. For oesteoarthritis pain and disability scores were converted to a scale of 0 (no pain or disability) to 100 (worst possible pain or disability) on a continuous variable. The results from five trials used in the meta analysis are shown in Figure 1.

Figure 1 Forest Plot of the difference in mean pain/disability scores between paracetamol and placebo treated groups from 5 trials in osteoarthritis



Fixed-Effects Model (k = 5)

Test for Heterogeneity:
Q(df = 4) = 6.3073, P-value = 0.1773

Table 1. Results from Fixed Effects Analysis

				95% Confidence Interval	
Estimate	Standard Error	Z-value	P-value	Lower	Upper
-3.0843	0.9719	-3.1735	0.0015	-4.9892	-1.1794

i) Explain what a Fixed Effects model is in meta-analysis. What is 'fixed'? **(2 marks)**

Question 1 continued on next page

ii) Explain what the Q statistic means, how the degrees of freedom (df) are calculated and how the P-value was derived. **(3 marks)**

iii) Explain how the estimate of -3.0843 in Table 1 is computed. **(1 mark)**

iv) A simple average or mean of the five treatment differences reported in Figure 1 is 4.18. Contrast this simple mean difference with the one given in Table 1. Why are they different? **(2 marks)**

A random-effects model was run, with the following results

Random-Effects Model (k = 5; τ^2 estimator: REML)

τ^2 (estimated amount of total heterogeneity): 2.7624 (SE = 5.6029)
τ (square root of estimated τ^2 value): 1.6620
I^2 (total heterogeneity / total variability): ?

Test for Heterogeneity:
Q(df = 4) = 6.3073, P-value = 0.1773

Table 2. Results from Random Effects Analysis

				95% Confidence Interval	
Estimate	Standard Error	Z-value	P-value	Lower	Upper
-3.3457	1.2797	-2.6145	0.0089	-5.8538	-0.8375

v) Explain briefly what REML is. **(2 marks)**

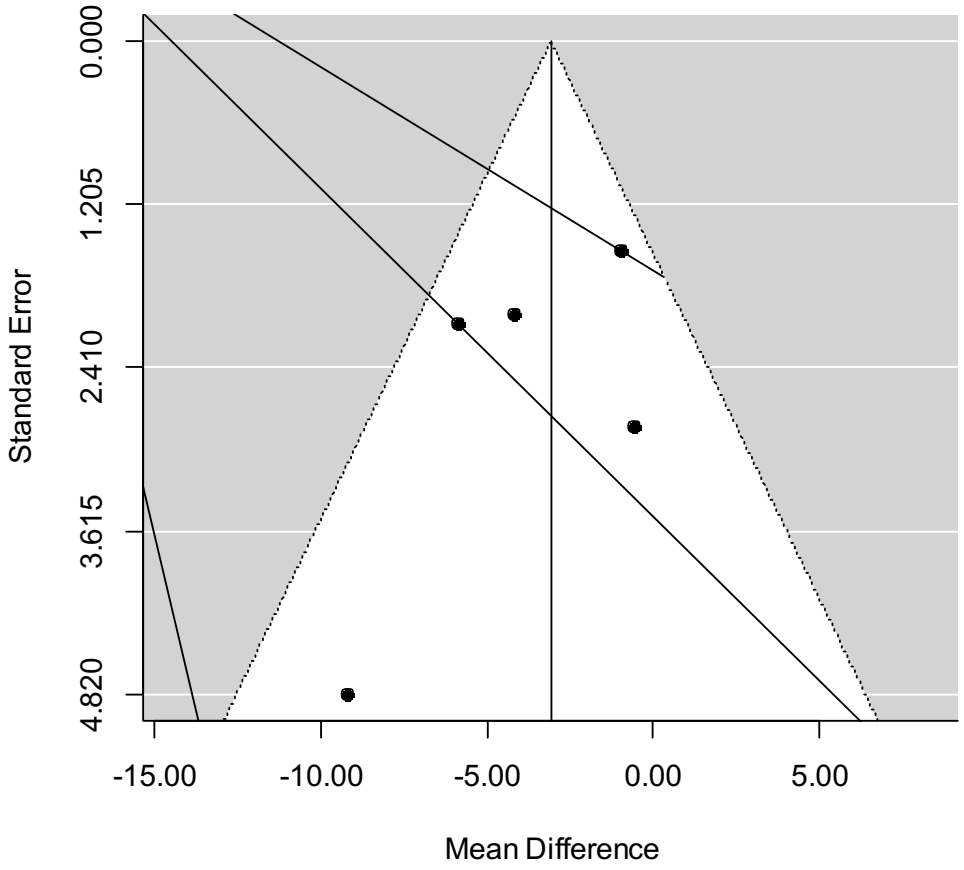
vi) Derive the value of the I^2 statistic from the output and comment on your results. **(2 marks)**

vii) Explain why the estimates, standard error and P-values differ between the fixed and random effects models shown in Table 1 and 2 respectively. **(3 marks)**

Question 1 continued on next page

A funnel plot of the data is shown in Figure 2

Figure 2 A funnel plot of the data from the five trials in osteoarthritis



viii) Describe Figure 2 and explain how Funnel plots may be used to detect publication bias. **(4 marks)**

ix) What conclusion do you draw from the meta-analysis of paracetamol against placebo for treatment of osteoarthritis? **(1 mark)**

Turn Over

2 A researcher is planning a randomised controlled clinical trial to compare a new treatment to standard treatment for colorectal cancer. The primary outcome/endpoint is (from randomisation) time to death. They have looked at four recently published randomised controlled trials on the standard treatment in this disease, two double-blind and two non-blinded studies, and estimate the median time death to be 12 months. The researcher thinks the new treatment will extend the median time to death by 33% and is planning a clinical trial to investigate this.

There is expected to be uniform accrual of patients over a 12 month recruitment period and a minimum 12 month follow-up period.

Using values from the Normal distribution to assist you with this question

X	Z_{1-X}
0.025	1.96
0.050	1.64
0.100	1.28
0.200	0.84

i) Give three factors the researcher should consider when using these published data to estimate the median time to death on standard. (**3 marks**)

ii) Define what the Type I error, the Type II error and the power are when designing a clinical trial. **(3 marks)**

iii) What are the accrual period and follow up periods in a clinical trial? **(1 mark)**

iv) Determine the number of events required to test this hypothesis with a 5% two-sided type I error α and 90% power (1-β). Using the formula below, determine the number of events. Note in formula $\theta=\log_e$(hazard ratio) (**2 marks**)

$$E = \frac{(e^{\theta}+1)^2 (Z_{1-\alpha/2} + Z_{1-\beta})^2}{(e^{\theta}-1)^2}$$

v) If 10% of patients are lost to follow up what would be the revised sample size be? **(1 mark)**

vi) For each of standard and new treatments, calculate the approximate probability of death (π) over the trial period (i.e. 12 months of uniform accrual (A) with entry rate λ plus 12 months minimum follow-up (F). Further, using an appropriate average, calculate the approximate probability of death across both treatment groups. (**3 marks**)

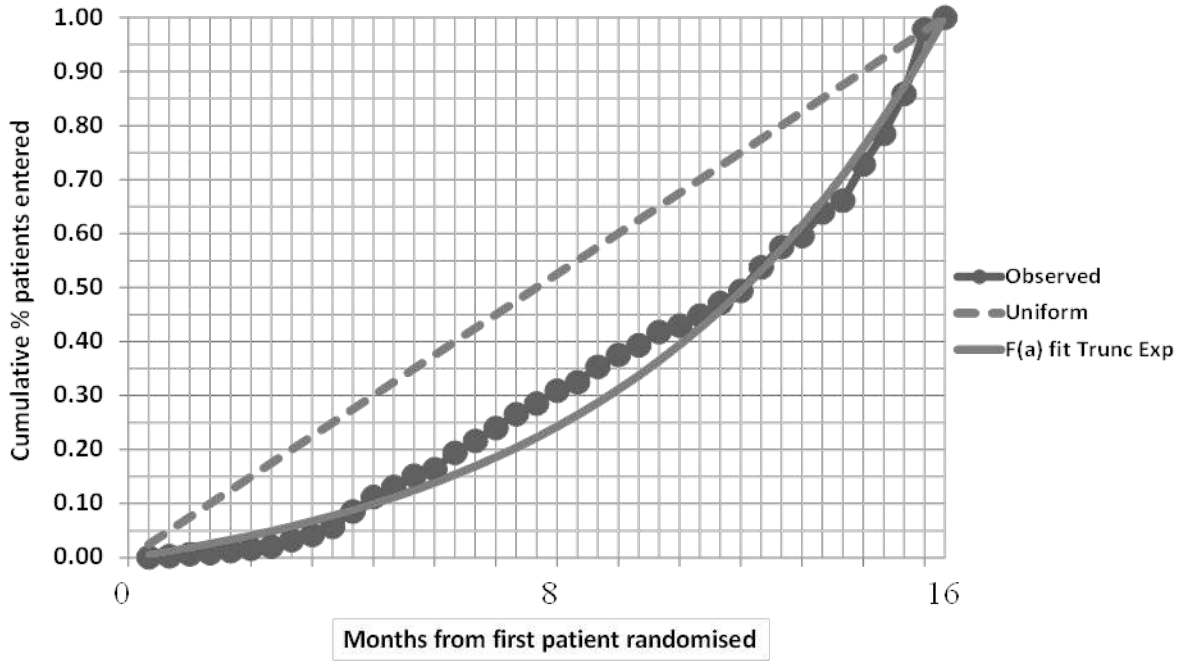
Note assuming exponential time to death, $\pi \approx 1 - e^{-\lambda\left(\frac{A}{2}+F\right)}$

Question 2 continued on next page

vii) Find the number of patients required to recruit into the trial using the results from (vi) and (v).

(**2 marks**)

You are shown the individual patient entry time data from a previous trial in the same population and the following is provided. The data do not seem to follow a uniform pattern of accrual but instead seems to be truncated exponential.



viii) What are uniform and non-uniform accrual in a trial. **(1 mark)**

ix) The expected entry time, $E(a)$, into the trial is 8 months. With a truncated exponential distribution the approximate probability of an event over the trial follow-up period π is $\approx 1 - e^{-\lambda(A+F-E(a))}$. Recalculate the number of patients required to recruit into the trial using the results from (vi) and (v) (**4 marks**)

Turn Over

3 A new drug is being developed for the treatment of migraine. The study is about to start clinical development with the first time into man study. The clinical program for this new drug will take 8 years and will involve a number of studies.

i) In the clinical research programme new migraine treatment will be developed a cross a number of different phases (Phases I, II, III and IV). With reference to this example

a) What is assessed in Phase I trials? **(2 marks)**

b) What is assessed in Phase II trials? **(2 marks)**

c) What is assessed in Phase III trials? **(1 marks)**

ii) For a new drug treatment for migraine why might a pilot or feasibility study be undertaken?

(2 marks)

iii) What are the advantages of a cross-over trial over a parallel group trial in a clinical development program for migraine? **(2 marks)**

A study is being designed to investigate new if practice nurses are as good as general practitioners in the planning of care for children with asthma in primary care. The objective of the trial is to assess if nurses can improve the care of children with asthma. The primary outcome will be the number of times children visit their general practice as they are ill with their asthma during the 12 months of the study.

iv) Given the above scenario would you recommend a cluster randomised controlled trial or an individually randomised controlled trial to investigate the research question. Please give reasons for your decision on the trial design **(5 marks)**

v) What are the different responsibilities of the researcher, the sponsor and the care organisation in running the above clinical trial? **(3 marks)**

vi) As the study involves children the ethics committee requested more information on safety monitoring. For safety monitoring in this trial what consideration would there need to be with respect to blinding and independence? Who is responsible for establishing relevant monitoring? Who would be independent to undetake trial monitoring? **(3 marks)**

Turn Over

4 In microscopic imaging it is common to model the number of photons arriving at the lens in each frame, X_i, as $\mathrm{Po}(x_i \mid \lambda)$, where λ is the rate of photon emission per frame. Given a random sample, $\boldsymbol{x} = \{x_1, \ldots, x_n\}$,

(i) (a) Show that $\pi(\lambda) = \mathrm{Ga}(\lambda \mid a, b)$ is a conjugate prior and give explicit expressions for the posterior parameters. ***(5 marks)***

(b) Find the Bayes estimator for λ under 0-1 loss; i.e. the posterior mode. ***(3 marks)***

(ii) (a) Calculate the predictive distribution of Y, the number of photons captured by the lens in the next random sample of m frames,

$$Y = \sum_{j=n+1}^{n+m} X_j$$

(7 marks)

(b) The scientist a priori believes that $\mathbb{E}[\lambda] = 10/3$ and $\mathbb{V}[\lambda] = 50/9$. Calculate the scientist's probability of observing not more than one photon in the next frame if 3 photons were detected in a sample of $n = 10$ frames. ***(5 marks)***

5 Consider the hierarchical model,

$$\begin{aligned} X_i &\sim \mathrm{Ber}(x_i \mid \theta_i) \text{ , ind. } i = 1, \ldots, n \\ \pi(\theta_i) &= \mathrm{Be}(\theta_i \mid a, a) \text{ , ind. } i = 1, \ldots, n \\ \pi(a) &= \mathrm{Ga}(a \mid c, d) \text{ , with } \mathbb{E}[a] = \frac{c}{d} \text{ .} \end{aligned}$$

(i) Write down the full conditional distributions for $\boldsymbol{\theta} = \{\theta_1, \ldots, \theta_n\}$ and a. ***(8 marks)***

(ii) Write pseudo-code for a Metropolis-within-Gibbs strategy to sample from $\pi(\boldsymbol{\theta}, a \mid \boldsymbol{x})$. ***(12 marks)***

6 Assume $\boldsymbol{X} = \{X_1, \ldots, X_n\}$ are independent random variables with

$$X_i \sim \mathrm{N}\left(x_i \,\middle|\, \mu, \frac{1}{a_i \lambda}\right),$$

for $i = 1, \ldots, n$; where $\boldsymbol{a} = \{a_1, \ldots, a_n\}$ are known constants with $0 < a_i < 1$ and $\sum_{i=1}^{n} a_i = 1$.

(i) Show that

$$\pi(\mu, \lambda) = \mathrm{N}\left(\mu \,\middle|\, m, \frac{1}{p\lambda}\right)\mathrm{Ga}(\lambda \,|\, a, b)$$

is a conjugate prior and provide explicit expressions for the posterior parameters. ***(15 marks)***

(ii) Show that

$$\mathbb{E}[\mu \mid \boldsymbol{x}] = w\widehat{\mu} + (1 - w)m \,,$$

where $0 < w < 1$ and $\widehat{\mu} = \sum_{i=1}^{n} a_i x_i$ is the MLE. ***(5 marks)***

End of Question Paper

Notation and distributions

Bayesian Statistics 2015–16

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 $\mathrm{N}\left(x \,\middle|\, \mu, \sigma^2\right)$. 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)	$\mathrm{U}(1, ..., k)$	$p(x) = 1/k$ $\mathcal{X} = \{1, ..., k\}, \quad \mathcal{K} = \mathbb{Z}_+$	$\frac{k+1}{2}$	$\frac{k^2-1}{12}$	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	$\mathrm{Bi}(x \mid n, \theta) \equiv \mathrm{Po}(x \mid n\theta)$ if n large, θ small

SOME CONTINUOUS DISTRIBUTIONS

Name	**Notation**	**p.d.f.** $f(x \mid \theta)$	$\mathbb{E}[X \mid \theta]$	$\mathbb{V}[X \mid \theta]$	**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.	Also parameterised in terms of $1/\beta$ $\mathrm{Ga}(x \mid 1, \lambda) \equiv \mathrm{Ex}(x \mid \lambda)$, $\mathrm{Ga}(x \mid \nu/2, 1/2) \equiv \chi^2_{(\nu)}(x)$
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. Also as a Bayesian conjugate prior.	$\mathrm{Be}(x \mid 1, 1) \equiv \mathrm{Un}(x \mid 0, 1)$ $\mathrm{Be}(x \mid \alpha, \beta)$ 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)	$\mathrm{N}\big(x \,\big\vert\, \mu, \sigma^2\big)$	$f(x) = \dfrac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\dfrac{x - \mu}{\sigma}\right)^2\right]$ $\mathcal{X} = \mathbb{R}$ $\Theta = \{(\mu, \sigma^2) \in \mathbb{R}^2 \,:\, \sigma^2 > 0\}$	μ	σ^2	Empirically and theoretically (via CLT) a useful model. Often parameterised in terms of the precision $\lambda = 1/\sigma^2$	$Y = aX + b \sim \mathrm{N}\big(y \,\big\vert\, a\mu + b, a^2\sigma^2\big)$ $Z = \frac{X - \mu}{\sigma} \sim \mathrm{N}(z \mid 0, 1)$ $\mathrm{P}[X \in (u, v)] = \mathrm{P}\Big[Z \in \Big(\dfrac{u - \mu}{\sigma}, \dfrac{v - \mu}{\sigma}\Big)\Big]$
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	$\mathrm{St}(x \mid \mu, \lambda, \nu)$	$f(x) = \frac{\Gamma[(\nu+1)/2]}{\Gamma[\nu/2]}\left(\frac{\lambda}{\nu\pi}\right)^{1/2} \times$ $\big(1 + \lambda(x - \mu)^2/\nu\big)^{-(\nu+1)/2}$ $\mathcal{X} = \mathbb{R}, \quad \mu \in \mathbb{R}, \lambda, \nu > 0$	μ (if $\nu > 1$)	$\lambda^{-1}\dfrac{\nu}{\nu - 2}$ (if $\nu > 2$)	Useful alternative to Gaussian for 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.** $f(x \mid \theta)$	$\mathbb{E}[X \mid \theta]$	$\mathbb{V}[X \mid \theta]$	**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	$\mathrm{N}_k(\boldsymbol{x} \mid \boldsymbol{\mu}, \Lambda)$	$f(\boldsymbol{x}) = \frac{\lvert\Lambda\rvert^{1/2}}{(2\pi)^{k/2}} \exp\big[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})'\Lambda(\boldsymbol{x}-\boldsymbol{\mu})\big]$ $\mathcal{X} = \boldsymbol{x} \in \mathbb{R}^k$ $\boldsymbol{\mu} \in \mathbb{R}^k$; Λ symmetric positive-definite	$\boldsymbol{\mu}$	Λ^{-1}	See univariate case	Usually parameterised in terms of the covariance matrix $\Sigma = \Lambda^{-1}$
Student	$\mathrm{St}_k(\boldsymbol{x} \mid \boldsymbol{\mu}, \Lambda, \nu)$	$f(\boldsymbol{x}) = \frac{\lvert\Lambda\rvert^{1/2}\Gamma((\nu+k)/2)}{(\nu\pi)^{k/2}\Gamma(\nu/2)} \times \left[1 + \frac{1}{\nu}(\boldsymbol{x}-\boldsymbol{\mu})'\Lambda(\boldsymbol{x}-\boldsymbol{\mu})\right]^{-(\nu+k)/2}$ $\mathcal{X} = \boldsymbol{x} \in \mathbb{R}^k$ $\boldsymbol{\mu} \in \mathbb{R}^k$; 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 $\Sigma = \Lambda^{-1}$