MAST30027: Modern Applied Statistics

Assignment 4

Due: 1pm Friday September 25 (week 9)

This assignment is worth 3 1/3% of your total mark.

1. The Dirichlet distribution is a multivariate generalisation of the beta distribution. It takes values in $\{\mathbf{x} = (x_1, \ldots, x_d) : x_i \in [0,1], \sum_i x_i = 1\}$, and, for $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_d) \in \mathbb{R}_+^d$, has density

$$f(\mathbf{x}) = \frac{1}{B(\boldsymbol{\alpha})} \prod_{i=1}^{d} x_i^{\alpha_i - 1}, \text{ where } B(\boldsymbol{\alpha}) = \frac{\prod_i \Gamma(\alpha_i)}{\Gamma\left(\sum_i \alpha_i\right)}.$$

   (a) Prove that if $\mathbf{X} = (X_1, \ldots, X_d)$ has a Dirichlet distribution with parameter $\boldsymbol{\alpha}$ (we write $\mathbf{X} \sim \mathrm{Dir}(\boldsymbol{\alpha})$), then $\mathbb{E}X_i = \alpha_i / \sum_i \alpha_i$. Hint: $\int \cdots \int f(\mathbf{x}) dx_1 \cdots dx_d = 1$.

   (b) Show that the Dirichlet distribution is the conjugate prior for the multinomial. That is, if $\mathbf{p} \sim \mathrm{Dirichlet}(\boldsymbol{\alpha})$ and $\mathbf{X}|\mathbf{p} \sim \mathrm{multinomial}(n, \mathbf{p})$, then $\mathbf{p}|\{\mathbf{X} = \mathbf{x}\} \sim \mathrm{Dirichlet}(\boldsymbol{\beta})$, where $\boldsymbol{\beta}$ depends on $\boldsymbol{\alpha}$ and $\mathbf{x}$.

   (c) In 2003 Briggs, Ades and Price reported on a trial for the treatment of asthma. Patients received one of two treatments (seretide or fluticasone), and their status was monitored from week to week. Possible states were

   STW Successfully treated week

   UTW Unsuccessfully treated week

   HEX Hospital managed exacerbation

   PEX Primary-care managed exacerbation

   TF Treatment failure (treatment ceased and patient removed from the trial)

   For patients on seretide, the number of transitions from one state to another were

	To					
	STW	UTW	HEX	PEX	TF	
From						Total
STW	210	60	0	1	1	272
UTW	88	641	0	4	13	746
HEX	0	0	0	0	0	0
PEX	1	0	0	0	1	2
TF	0	0	0	0	81	81

   The rows of this table can be considered as observations from independent multinomial random variables. Using Dir(1,. . . ,1) priors, give Bayesian estimates (posterior means) for

$$p_{ij} = \mathbb{P}(\text{ state changes from } i \text{ to } j \text{ })$$

   for $i = \mathrm{STW}, \ldots, \mathrm{PEX}$ and $j = \mathrm{STW}, \ldots, \mathrm{TF}$.

   (d) What prior would be appropriate for the transitions from state TF?

2. The $t(3)$ distribution has pdf $p(x) = \frac{2}{\sqrt{3}\pi}\left(1 + \frac{x^2}{3}\right)^{-2}$, $-\infty < x < \infty$, and the Cauchy$(\sqrt{3}, 0)$ distribution has pdf $g(y) = \frac{1}{\sqrt{3}\pi}\left(1 + \frac{y^2}{3}\right)^{-1}$, $-\infty < y < \infty$.

   (a) Suppose $U \overset{d}{=} U(0,1)$, and $Y = \sqrt{3}\tan(\pi(U - \frac{1}{2}))$. Show that $Y \overset{d}{=}$ Cauchy$(\sqrt{3}, 0)$.

   (b) Construct an A-R sampling algorithm (or a mixture of A-R and transformation algorithm) for generating random numbers from $t(3)$ by using the result of (a).

   (c) Write an R function to implement (b). Then use it to generate a sample of 1000 numbers from $t(3)$, and compare the sample pdf curve with the actual $t(3)$ curve.

   Note: Your `R` outputs should be concise and properly organised before they are printed and submitted with the assignment.