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: $\int \cdots \int f(\mathbf{x}) dx_1 \cdots dx_d = 1$.

Solution:

$$\begin{aligned}
\mathbb{E}X_1 &= \int \cdots \int x_1 \frac{1}{B(\alpha_1, \ldots, \alpha_d)} \prod_{i=1}^d x_i^{\alpha_i - 1} dx_1 \cdots dx_d \\
&= \frac{B(\alpha_1 + 1, \ldots, \alpha_d)}{B(\alpha_1, \ldots, \alpha_d)} \int \cdots \int \frac{1}{B(\alpha_1 + 1, \ldots, \alpha_d)} x_1^{\alpha_1 + 1 - 1} \prod_{i=2}^d x_i^{\alpha_i - 1} dx_1 \cdots dx_d \\
&= \frac{B(\alpha_1 + 1, \ldots, \alpha_d)}{B(\alpha_1, \ldots, \alpha_d)} \\
&= \frac{\Gamma\left(\sum_i \alpha_i\right) \Gamma(\alpha_1 + 1) \prod_{i=2}^d \Gamma(\alpha_i)}{\prod_i \Gamma(\alpha_i) \Gamma\left(\alpha_1 + 1 + \sum_{i=2}^d \alpha_i\right)} \\
&= \frac{\alpha_1}{\sum_i \alpha_i} \text{ since } \Gamma(n) = (n-1)\Gamma(n-1).
\end{aligned}$$

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

Solution:

$$\begin{aligned}
f(\mathbf{p}|\mathbf{x}) &\propto f(\mathbf{x}|\mathbf{p}) f(\mathbf{p}) \\
&\propto \prod_i p_i^{x_i} \prod_i p_i^{\alpha_i - 1} \\
&= \prod_i p_i^{x_i + \alpha_i - 1}
\end{aligned}$$

This is the kernel of a $\mathrm{Dir}(x_1 + \alpha_1, \ldots, x_d + \alpha_d)$ random variable, as required.

(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

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

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

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

Solution: For transitions from STW we have prior $\text{Dir}(1, 1, 1, 1, 1)$ and posterior $\text{Dir}(211, 61, 1, 2, 2)$. The posterior means are thus $211/277, 61/277, 1/277, 2/277, 2/277$. We can proceed in the same way for the next three rows, however for the last row we know that TF is an absorbing state. Thus we can use the deterministic prior which gives probability 1 to the response TF, and probability 0 to all the others (if you like, you can think of this as a Dir(0,0,0,0,∞) prior). Thus we get estimated transition probabilities

	To STW	UTW	HEX	PEX	TF
From					
STW	0.762	0.220	0.004	0.007	0.007
UTW	0.119	0.855	0.001	0.007	0.019
HEX	0.200	0.200	0.200	0.200	0.200
PEX	0.286	0.143	0.143	0.143	0.286
TF	0	0	0	0	1

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 $Y \overset{d}{=} \text{Cauchy}(\sqrt{3}, 0)$.

Solution: The cdf of Y is

$$\begin{aligned} F_Y(y) &= P[Y \le y] = P[\sqrt{3}\tan(\pi(U - \frac{1}{2})) \le y] \\ &= P[U \le \frac{1}{2} + \pi^{-1}\arctan\frac{y}{\sqrt{3}}] = \frac{1}{2} + \pi^{-1}\arctan\frac{y}{\sqrt{3}}, \quad -\infty < y < \infty \end{aligned}$$

The pdf of Y is then

$$f_Y(y) = F_Y'(y) = \frac{1}{\sqrt{3}\pi}\left(1 + \frac{y^2}{3}\right)^{-1}$$

which is the same as 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).

Solution: Since $p(x) = \frac{2}{\sqrt{3}\pi}\left(1+\frac{x^2}{3}\right)^{-2} = 2\left(1+\frac{x^2}{3}\right)^{-1} g(x) \le 2g(x)$, the following algorithm will generate random numbers from a $t(3)$ distribution:

1° Generate $U \sim U(0,1)$, and calculate $X = \sqrt{3}\tan(\pi(U-\frac{1}{2}))$.
2° Generate $Y \sim U(0, 2g(X))$ independently.
3° If $Y \le p(X)$ deliver X; otherwise go to 1°.

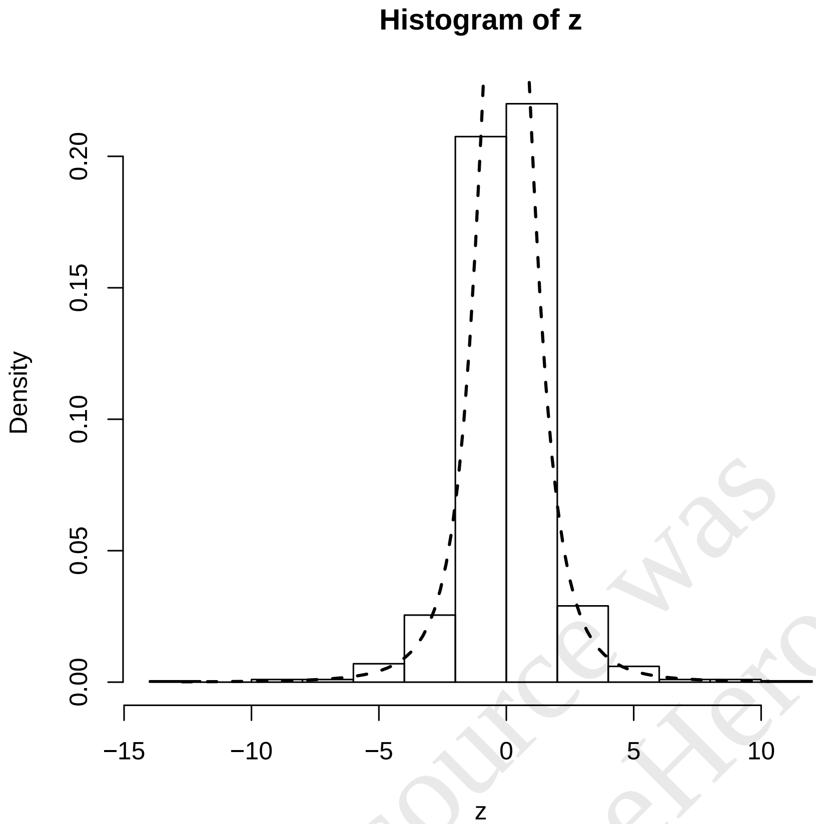
We can in fact make this a little more efficient. Note that if $V \sim U(0,1)$ then $2g(x)V \sim U(0, 2g(x))$ and $2g(x)V \le p(x)$ iff $V \le p(x)/(2g(x)) = \left(1+\frac{x^2}{3}\right)^{-1}$. Thus the following algorithm does the same as before, but instead of calculating $2g(x)$ and $p(x)$, it only has to calculate $\left(1+\frac{x^2}{3}\right)^{-1}$.

1° Generate $U \sim U(0,1)$, and calculate $X = \sqrt{3}\tan(\pi(U-\frac{1}{2}))$.
2° Generate $V \sim U(0,1)$ independently.
3° If $V \le 1/\left(1+X^2/3\right)$ deliver X; otherwise go to 1°.

(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.

Solution:

```
> t3sim <- function(){
+     # generate a t(3) random number.
+     u <- runif(1)
+     x <- sqrt(3.0)*tan(pi*(u-0.5))
+     v <- runif(1)
+     while(v > 1/(1+x^2/3)){
+         u <- runif(1)
+         x <- sqrt(3.0)*tan(pi*(u-0.5))
+         v <- runif(1)
+     }
+     return(x)
+ }
> set.seed(3456)
> n <- 1000
> z <- rep(0, n)
> for (i in 1:n) z[i] <- t3sim()
> hist(z, freq=F)
> curve(dt(x,3), from=min(z), to=max(z), add=T, lwd=2, lty=2)
```


Histogram of z
Density
0.00
0.05
0.10
0.15
0.20
−15
−10
−5
0
5
10
z