UNIVERSITEIT VAN DIE VRYSTAAT
UNIVERSITY OF THE FREE STATE

STSB 6816

WISKUNDIGE STATISTIEK & AKTUARIËLE WETENSKAP/ MATHEMATICAL STATISTICS & ACTUARIAL SCIENCE

Test 1 — 3 May 2019

MEMORANDUM

TYD/TIME: 160 Minutes PUNTE/MARKS: 45

INSTRUCTIONS:

- Answer all questions in a single Word document. Please convert to .pdf at the end.
- Label questions clearly, as it is done on this question paper.
- All results accurate to 2 decimal places.
- Show all derivations, formulas, code, sources and reasoning.
- Intervals should cover 95% probability unless stated otherwise.
- No communication software, devices or websites may be accessed prior to submission.

Question 1

Groups of participants in an on-line study are presented with a neutral statement that is nevertheless politically charged and divisive. Each group's consensus response to the statement in presented on a scale from 0 to 1.

You are asked to model the responses using the following density:

$$f(x) = 1 - \frac{\alpha}{4} + \alpha|x - 0.5| \ , \qquad 0 \le x \le 1 \ , \qquad -4 \le \alpha \le 4$$

The CDF of this density is

$$F(x) = \begin{cases} x\left(\frac{\alpha}{2}(0.5 - x) + 1\right) \ , & 0 \le x \le 0.5 \\ x\left(\frac{\alpha}{2}(x - 1.5) + 1\right) + \frac{\alpha}{4} \ , & 0.5 < x \le 1 \end{cases}$$

A real sample of responses ($\mathbf{x}_r$) was observed as 0.84, 0.24, 0.59, 0.83, 0.20, 0.91, 0.02, 0.07, 0.86, 0.59, 0.10, 0.80, 0.13, 0.05.

(a) Show in outline that the inverse CDF is given by

$$\Rightarrow F^{-1}(y) = \begin{cases} \left(\frac{1}{\alpha} + \frac{1}{4}\right) \pm \sqrt{\left(\frac{1}{\alpha} + \frac{1}{4}\right)^2 - \frac{2y}{\alpha}} \ni 0 \le F^{-1}(y) \le 0.5 \ , & 0 \le y \le 0.5 \\ \left(\frac{3}{4} - \frac{1}{\alpha}\right) \pm \sqrt{\left(\frac{1}{\alpha} - \frac{3}{4}\right)^2 + \frac{2y}{\alpha} - \frac{1}{2}} \ni 0.5 < F^{-1}(y) \le 1 \ , & 0.5 < y \le 1 \end{cases}$$

if $\alpha \neq 0$, and $F^{-1}(y) = y$ if $\alpha = 0$. [3]

Let $y = F(x)$ then $F^{-1}(y) = x$, so we must solve for x in ✔

$$y = \begin{cases} x\left(\frac{\alpha}{2}(0.5 - x) + 1\right), & 0 \leq x \leq 0.5 \\ x\left(\frac{\alpha}{2}(x - 1.5) + 1\right) + \frac{\alpha}{4}, & 0.5 < x \leq 1 \end{cases}$$

Since $F(0.5) = 0.5$ the split in y is at 0.5. ✔ Thus, if $\alpha \neq 0$,

$$x = \begin{cases} \frac{-\left(1+\frac{\alpha}{4}\right) \pm \sqrt{\left(1+\frac{\alpha}{4}\right)^2 - 4*\left(-\frac{\alpha}{2}\right)*(-y)}}{2*\left(-\frac{\alpha}{2}\right)} \ ✔ \ni 0 \leq x \leq 0.5, & 0 \leq y \leq 0.5 \\ \frac{-\left(1-\frac{3\alpha}{4}\right) \pm \sqrt{\left(1-\frac{3\alpha}{4}\right)^2 - 4*\left(\frac{\alpha}{2}\right)*\left(\frac{\alpha}{4}-y\right)}}{2*\left(\frac{\alpha}{2}\right)} \ni 0.5 < x \leq 1, & 0.5 < y \leq 1 \end{cases}$$

$$\Rightarrow F^{-1}(y) = \begin{cases} \left(\frac{1}{\alpha} + \frac{1}{4}\right) \pm \sqrt{\left(\frac{1}{\alpha} + \frac{1}{4}\right)^2 - \frac{2y}{\alpha}} \ni 0 \leq F^{-1}(y) \leq 0.5, & 0 \leq y \leq 0.5 \\ \left(\frac{3}{4} - \frac{1}{\alpha}\right) \pm \sqrt{\left(\frac{1}{\alpha} - \frac{3}{4}\right)^2 + \frac{2y}{\alpha} - \frac{1}{2}} \ni 0.5 < F^{-1}(y) \leq 1, & 0.5 < y \leq 1 \end{cases}$$

And $F^{-1}(y) = y$ if $\alpha = 0$ (Uniform distribution).

(b) Create a function that simulates a sample from this density given value for the sample size (n) and α by directly using the inverse CDF. Give a histogram of a random sample of size 10000 with $\alpha = 3$. [4]

```r
rhouse <- function(n,a) {
  sims <- runif(n)
  if (length(a)<n) { a <- c(matrix(a,n,1)) }
  for (i in 1:n) {
  if (a[i] != 0) {
      y <- runif(1)
      if (y <= 0.5) {
        x3 <- sqrt((1/a[i] + 0.25)^2 - 2*y/a[i])
        x2 <- (1/a[i] + 0.25) - x3
        x1 <- (1/a[i] + 0.25) + x3
        if (x2 < 0) {sims[i] <- x1} else { sims[i] <- x2 }
      } else {
        x3 <- sqrt((1/a[i] - 0.75)^2 + 2*y/a[i]-0.5)
        x2 <- (0.75 - 1/a[i]) - x3
        x1 <- (0.75 - 1/a[i]) + x3
        if (x2 < 0.5) {sims[i] <- x1} else { sims[i] <- x2 }
      }
    }
  }
  return(sims)
}
hist(rhouse(10000,3))
```

✔✔✔

✔

(c) Assume a $Uniform(-4, 4)$ prior for α and derive the posterior density of α. Create a function that calculates the posterior density at a vector of alpha values given a sample, but unscaled due to the unknown constant. [4]

$$p(\alpha|\mathbf{x}) \propto p(\alpha) * Like(\mathbf{x}|\alpha)$$

$$= 1/8 * \checkmark \prod_{i=1}^{n} \checkmark \left[1 - \frac{\alpha}{4} + \alpha|x_i - 0.5|\right], \qquad -4 \leq \alpha \leq 4$$

```r
postalpha <- function(alphas,xs) { sapply(alphas,function(a) { prod(1-(a/4)+a*abs(xs
    -0.5)) }) }
```

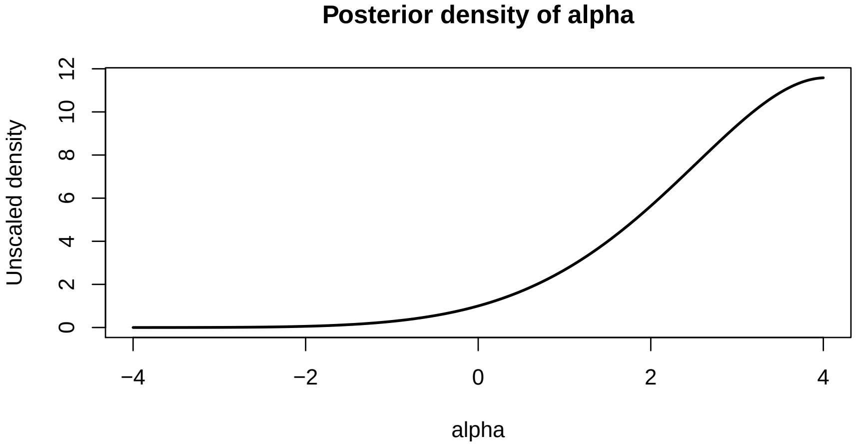
✔✔

(d) Create a discrete vector of possible α values over the domain of α, with length

1001. Draw a plot of the posterior distribution of the real sample at these values ($p(\alpha|\mathbf{x}_r)$). Estimate the posterior mean of α using the values calculated here (via discretisation). [4]

```
alphas <- seq(-4,4,l=1001)
xr <- c(0.84, 0.24, 0.59, 0.83, 0.20, 0.91, 0.02, 0.07, 0.86, 0.59, 0.10, 0.80, 0.13,
    0.05)
postalphas <- postalpha(alphas,xr)
plot(alphas,postalphas,type='l',main='Posterior density of alpha',xlab='alpha',ylab='
    Unscaled density',lwd=2)
(postmean1 <- sum(alphas*postalphas/sum(postalphas)))
```

✔✔



✔

Posterior mean estimate should be 2.573225. ✔

(e) Simulate a sample of 10000 random values from this posterior distribution and use these simulations to estimate the posterior mean and posterior median estimates of α. [Note that another 3 (bonus) marks will be awarded for correctly implementing acceptance rejection sampling to do this simulation.] [3]

```
postsims <- sample(alphas,10000,T,postalphas)
(postmean2 <- mean(postsims))
(postmedian2 <- median(postsims))
```

✔✔

Mean close to 2.6, median close to 2.8. ✔

(f) Simulate a sample of size 10000 from the posterior predictive distribution of a single new observation ($p(x_{new}|\mathbf{x}_r)$). Use these simulations to estimate the $P[|x-0.5|>0.2] = P[(x<0.3)\cup(x>0.7)]$. [3]

```
postpred <- rhouse(10000,postalphas)
(est1 <- mean(abs(postpred-0.5)>0.2))
```

✔✔

About 0.753 ✔

(g) Simulate 1000 random samples from the original density using $\alpha = 2$ and $n = 14$. For each sample simulate 10000 values from its posterior distribution and calculate a 95% HPD credibility interval for α. Give the empirical coverage of your intervals. [6]

```
hpd.interval <- function(postsims,alpha=0.05) { # Coded by Sean van der Merwe, UFS
sorted.postsims <- sort(postsims)
nsims <- length(postsims)
numints <- floor(nsims*alpha)
gap <- round(nsims*(1-alpha))
widths <- sorted.postsims[(1+gap):(numints+gap)] - sorted.postsims[1:numints]
HPD <- sorted.postsims[c(which.min(widths),(which.min(widths) + gap))]
return(HPD) }

truealpha <- 2
nsamples <- 1000
intervals <- matrix(NA,nsamples,2)
for (i in 1:nsamples) {
  x <- rhouse(14,truealpha)
  postalphas <- postalpha(alphas,x)
  postsims <- sample(alphas,10000,T,postalphas)
  intervals[i,] <- hpd.interval(postsims)
}
(coverage <- mean((intervals[,1]<=truealpha)&(intervals[,2]>=truealpha)))
```

✔✔✔✔

Coverage should be close to 95%. ✔✔

(h) Show in outline that the MDI prior is proportional to

$$\exp\left\{\frac{1}{\alpha}\left[\left(1+\frac{\alpha}{4}\right)^2\ln\left(1+\frac{\alpha}{4}\right)-\left(1-\frac{\alpha}{4}\right)^2\ln\left(1-\frac{\alpha}{4}\right)\right]\right\}$$

[3]

$$MDIP \propto \exp\left\{E_X\left[\ln f(x|\boldsymbol{\theta})\right]\right\} ✔$$

$$= \exp\left\{\int_0^1\left[1-\frac{\alpha}{4}+\alpha|x-0.5|\right]\ln\left[1-\frac{\alpha}{4}+\alpha|x-0.5|\right]dx\right\} ✔$$

Splitting the integral at 0.5 ✔ and using the result that $\int u\ln u du = 0.5u^2(\ln u - 0.5) + c$ ✔ combined with $\frac{du}{dx} = \alpha * sign(x-0.5)$ ✔ we get the given result after simplifying.

(i) Explain how you would determine whether the MDI prior works better for the given sample ($\mathbf{x}_r$) (don't implement it, just explain the steps you would take). [4]

1. I would adapt the simulation study in part (g) to include more discrepancy measures, like mean square error of best estimate. ✔
2. Then repeat it for different values of α and n. ✔
3. Store all summary measures for comparison. ✔
4. Then do every step above for the MDI prior and compare. ✔

Total for Question 1: 34

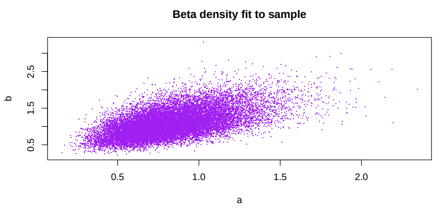
Question 2

Consider the sample given in the previous question ($\mathbf{x}_r$). We now wish to model the sample using a $Beta(a, b)$ density instead. Assume $Exp(0.00001)$ priors for each parameter and implement the following:

(a) Simulate a sample from the joint posterior distribution (at least 10000 draws). Draw a scatterplot of the simulations of b vs a. [8]

```
library(R2OpenBUGS)
betamodel <- function() {
  for (i in 1:n) {
    y[i] ~ dbeta(a,b)
  }
  a ~ dexp(0.00001)
  b ~ dexp(0.00001)
}
write.model(betamodel,'betamodel.txt')
BUGSdata <- list(n=length(xr), y=xr)
ab1 <- mean(xr)*mean(1-xr)/var(xr)
inits <- function() {list(a=(mean(xr)*(ab1-1)), b=(mean(1-xr)*(ab1-1))) }
bugsoutput <- bugs(BUGSdata,inits,c('a','b'),20000, 'betamodel.txt', 2, 10000, 2,
    debug=TRUE)
plot(bugsoutput$sims.list$a, bugsoutput$sims.list$b, pch='.', col='purple', main='
    Beta density fit to sample', xlab='a', ylab='b')
```

✓✓✓✓✓✓



✓✓

(b) Simulate a sample of size 10000 from the posterior predictive distribution of a single new observation ($p(x_{new}|\mathbf{x}_r)$). Use these simulations to estimate the $P[|x - 0.5| > 0.2] = P[(x < 0.3) \cup (x > 0.7)]$. [3]

```
postpred2 <- rbeta(10000,bugsoutput$sims.list$a[1:10000],bugsoutput$sims.list$b
    [1:10000])
(est2 <- mean(abs(postpred2-0.5)>0.2))
```

✓✓

About 0.63 ✓

Total for Question 2: 11

Total half marks on memo = 90 vs. 90 = Double total margin points (=45).