UNIVERSITEIT VAN DIE VRYSTAAT
UNIVERSITY OF THE FREE STATE

STSB 6816

WISKUNDIGE STATISTIEK & AKTUARIËLE WETENSKAP/
MATHEMATICAL STATISTICS & ACTUARIAL SCIENCE

Test 1 — 3 May 2019

TYD/TIME: 160 Minutes PUNTE/MARKS: 45

INSTRUCTIONS:

- Answer all questions in a single Word document.
- 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:

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

The CDF of this density is

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

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

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

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

(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]

(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]

(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]

(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]

(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 $P[|x - 0.5| > 0.2] = P[(x < 0.3) \cap (x > 0.7)]$. [3]

(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]

(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]

(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]

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]

(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 $P[|x - 0.5| > 0.2] = P[(x < 0.3) \cap (x > 0.7)]$. [3]

Total for Question 2: 11