STSB6816 Special Test of 2022

Mathematical Statistics and Actuarial Science; University of the Free State

2022/06/23

Time: 150 minutes; Marks: 40

MEMORANDUM

Instructions

- Answer all questions in a single R Markdown document. Please knit to Word or PDF at the end and submit both the PDF/Word document and the .Rmd file for assessment, in that order.
- Label questions clearly, as it is done on this question paper.
- All results accurate to at least 1 decimal place, ensure that simulation error is small enough (by doing enough simulations).
- Show all derivations, formulas, code, sources and reasoning.
- Intervals should cover 95% probability unless stated otherwise.
- No communication software, no devices, and no communication capable websites may be accessed prior to submission. You may not (nor even appear to) attempt to communicate or pass information to another student.

Question 1

Consider the Laplace density: $f(y|\mu,b) = \frac{1}{2b}\exp\left(-\frac{|y-\mu|}{b}\right)$.

1.1) Write down the log density. **[1]**

$$g = \log f(y|\mu,b) = -\log 2 - \log b - \frac{|y-\mu|}{b}$$

1.2) Derive the MDI prior. **[2]**

$$E(g) = -\log 2 - \log b - \frac{E|y-\mu|}{b} = -\log 2 - \log b - 1$$

$$\pi_{MDI}(\mu,b) \propto b^{-1}$$

1.3) Derive the Jeffreys prior for b assuming $\mu = 0$. **[4]**

$$\frac{\partial g}{\partial b} = -\frac{1}{b} + \frac{|y|}{b^2}$$

$$\frac{\partial^2 g}{\partial b^2} = \frac{1}{b^2} - 2\frac{|y|}{b^3}$$

$$\pi_{Jeffreys}(b) \propto \sqrt{\frac{1}{b^2}} = b^{-1}$$

1.4) Now consider that you have a sample of values $y_1, \dots, y_n$ that are explained by a single continuous explanatory variable $x_1, \dots, x_n$ but with Laplace residuals. Write down the log likelihood and then show that the log posterior is given by the expression below. **[3]**

$$\pi(\boldsymbol{\theta}|\mathbf{y}, \mathbf{x}) = -(n+1)\log b - \frac{1}{b}\sum_{i=1}^{n}|y_i - (\beta_0 + \beta_1 x_i)|$$

1.5) What is the theoretical impact of using a Laplace distribution for the residuals instead of a normal distribution? **[3]**

The Laplace distribution places less emphasis on large deviations since it does not square them, thus it is a more robust regression. [3]

1.6) Use the code below to capture a set of data, and show that the average y value is 5.2 **[2]**

```
d <- data.frame(x = 1:20, y = c(8.6, 5.6, 7.2, 6.9, 6.8, 5.4, 6.8, 5.9, 4.4, 5.9,
4.8, 5.2, 3.7, 5.3, 3.4, 4, 4.2, 3.1, 3.6, 3.2))
mean(d$y)
```

```
| [1] 5.2
```

Typing in code and numbers correctly [2].

1.7) Fit an OLS regression through the points as given. Give a summary of the parameters. **[2]**

```
OLS <- lm(y ~ x, data = d)
OLS |> summary() |> broom::tidy() |> kable(digits = 2)
```

term	estimate	std.error	statistic	p.value
(Intercept)	7.59	0.35	21.77	0
x	-0.23	0.03	-7.81	0

Fitting a regression model [1], giving summary [1].

1.8) Fit a Bayesian linear model with Laplace errors through the points as given. Give a trace plot and summary of the parameters. **[10]**

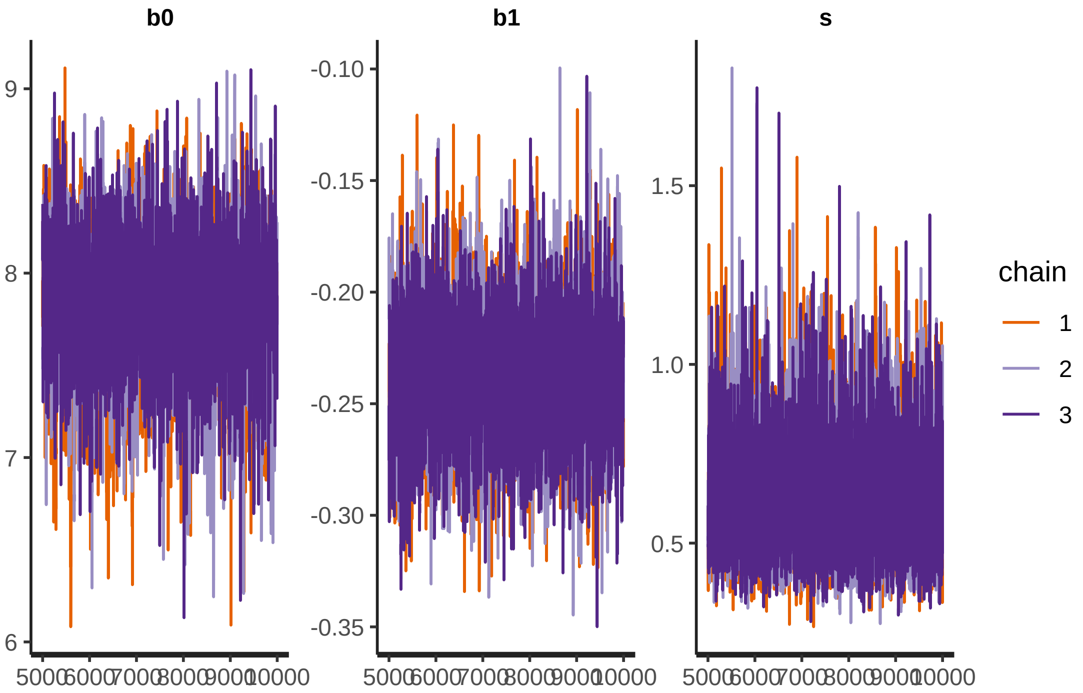
```
library(rstan)
mycores <- 3
options(mc.cores = mycores)

// This Stan block defines a simple Laplace regression model, by Sean van der Merwe,
UFS
data {
  int<lower=1> n;                    // number of observations
  real y[n];                 // observations
  real x[n];                 // explanatory variables
}
// The parameters of the model
parameters {
  real b0;
  real b1;
  real<lower=0> s;
```

```
}
model {
  for (i in 1:n) {
    y[i] ~ double_exponential(b0 + b1*x[i], s);
  }
  target += -log(s);    // joint objective prior
}
generated quantities {
  vector[n] preds;
  for (i in 1:n) {
    preds[i] = double_exponential_rng(b0 + b1*x[i], s);
  }
}

ModelFit <- sampling(LaplaceReg, list(n = nrow(d), y = d$y, x = d$x), iter = 10000,
chains = mycores)

ModelFit |> traceplot(pars = c('b0', 'b1', 's'))
```




```
(ModelFit |> summary(pars = c('b0', 'b1', 's')))$summary |> kable(digits = 2)
```

	mean	se_mean	sd	2.5%	25%	50%	75%	97.5%	n_eff	Rhat
b0	7.83	0	0.32	7.08	7.67	7.86	8.02	8.42	4160.84	1
b1	-0.24	0	0.03	-0.29	-0.25	-0.24	-0.22	-0.18	4434.85	1
s	0.62	0	0.15	0.39	0.51	0.60	0.70	0.98	5402.89	1

Coding the model [4], using the prior as given [2], running the model and giving trace plot plus parameter summary [4].

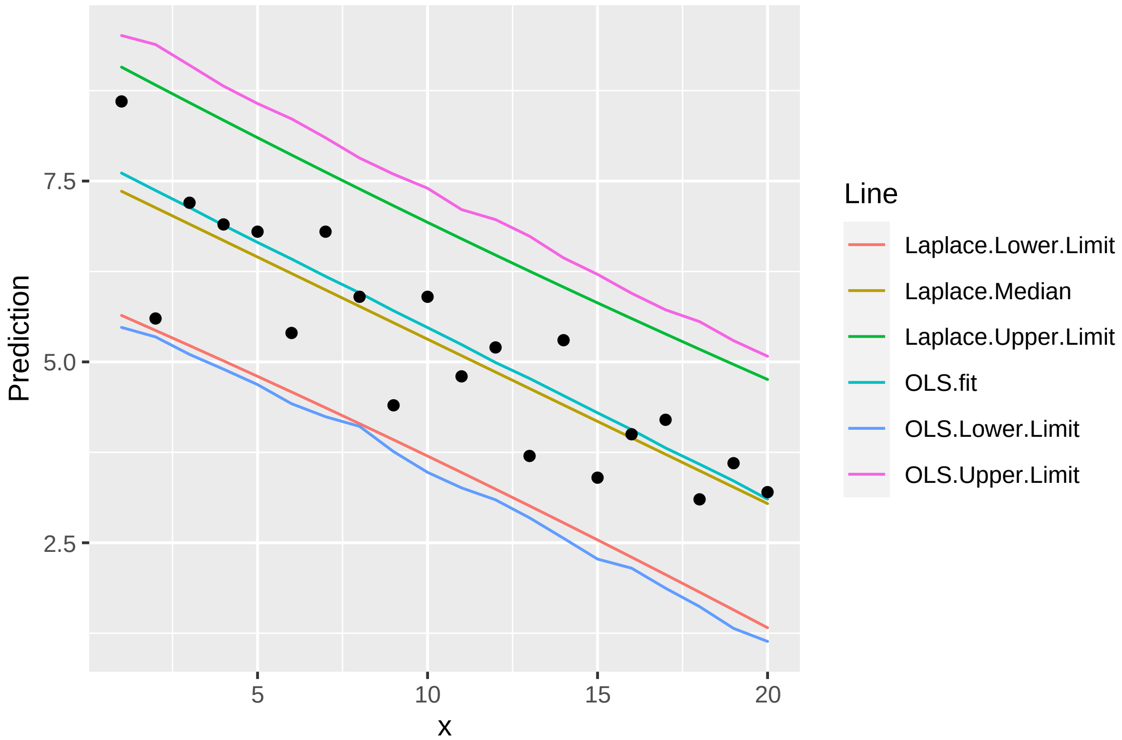
1.9) Draw a single plot showing the data points and both model fits, as well as prediction intervals for both models. **[10]**

HINT You can use posterior mean or median estimates to produce Laplace predicted values without intervals. Leave the Laplace intervals for last, after you have already done the plot and some interpretation, as they require you to simulate new Laplace values.

```
ModelFit |> extract() -> draws

OLSpreds <- predict(OLS, newdata = d, interval = 'prediction')
draws$preds |> apply(2, quantile, c(0.5, 0.025, 0.975)) |> t() |>
data.frame(OLSpreds) -> allpreds
names(allpreds) <- c('OLS fit', 'OLS Lower Limit', 'OLS Upper Limit', 'Laplace
Median', 'Laplace Lower Limit', 'Laplace Upper Limit')

suppressPackageStartupMessages(library(tidyverse))
allpreds |> data.frame() |>
  pivot_longer(everything(), names_to = 'Line', values_to = 'Prediction') |>
  data.frame(x = rep(d$x, each = ncol(allpreds))) -> plot_data
plot_data |>
  ggplot(aes(x = x, y = Prediction)) + geom_line(aes(colour = Line)) +
geom_point(data = d, mapping = aes(x = x, y = y))
```



Data points [2], OLS predictions [1], OLS Intervals [1], Laplace predictions [2], Laplace intervals [4].

1.10) Explain, in detail, which model appears to be a better fit based on your plot. **[3]**

Capturing more/less of the relevant information, width and coverage and key factors to talk about. In particular we note that the Laplace has smaller intervals as it is less affected by the outlying observations. [3].

Points total

The points on the test add up to **40**
