UNIVERSITEIT VAN DIE VRYSTAAT
UNIVERSITY OF THE FREE STATE

STSB 6815 and 7910

WISKUNDIGE STATISTIEK & AKTUARIËLE WETENSKAP/ MATHEMATICAL STATISTICS & ACTUARIAL SCIENCE

Test 2 — 4 May 2016

MEMORANDUM

TYD/TIME: 200 Minutes PUNTE/MARKS: 35

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

Consider the BUGS representation of the Logistic distribution:

$$f(x|\mu,\tau) = \frac{\tau e^{\tau(x-\mu)}}{\left(1+e^{\tau(x-\mu)}\right)^2}\ , \quad -\infty < x, \mu < \infty\ ,\ 0 < \tau < \infty$$

The Logistic distribution is a popular distribution because of its similarities to the Normal.

(a) Show that the MDI prior is proportional to τ. [5]

$$
\begin{aligned}
\ln \pi(\mu,\tau) &= E(\ln f(x|\mu,\tau)) + c_1 \checkmark \\
&= E\left[\ln \tau + \tau(x-\mu) - 2\ln\left(1+e^{\tau(x-\mu)}\right)\right] + c_1 \checkmark \\
&= \ln \tau + 0 - 2E\left[\ln\left(1+e^{\tau(x-\mu)}\right)\right] + c_1
\end{aligned}
$$

Let $w = 1 + e^{\tau(x-\mu)}$, then $\frac{dw}{dx} = \tau e^{\tau(x-\mu)}$, so that

$$
\begin{aligned}
E\left[\ln\left(1+e^{\tau(x-\mu)}\right)\right] &= \int_{-\infty}^{\infty} \ln\left(1+e^{\tau(x-\mu)}\right) \frac{\tau e^{\tau(x-\mu)}}{\left(1+e^{\tau(x-\mu)}\right)^2} dx \checkmark \\
&= \int_{1}^{\infty} \ln w \; w^{-2} dw \\
&= \left[\ln w(-w^{-1})\right]_1^{\infty} - \int_1^{\infty} w^{-1}(-w^{-1}) dw \\
&= \left[-w^{-1}\right]_1^{\infty} = 1 \checkmark \checkmark
\end{aligned}
$$

Finally we see that $\pi(\mu,\tau) \propto e^{\ln \tau} = \tau$.

The following data represent changes in the earth's rotation (i.e. day length). Units 0.00001 second. The data is for a random sample of 23 years.
-12,110,78,126,-35,104,111,22,-31,92,51,36,231,-13,65,119,21,104,112,-15,137,139,101

Reference: Acta Astron. Sinica Vol. 15, p79-85. The observed mean is 71.87 (use as a check). A possible question is whether the mean could actually be greater than 100 (1ms)?

(b) Fit a Logistic distribution to the data by simulating the full posterior distribution of the parameters. You are advised to approximate the MDI prior by assigning a flat prior to μ and a $Gamma(2, 0.0001)$ prior to τ. [6]

```
x <- c
    (-12,110,78,126,-35,104,111,22,-31,92,51,36,231,-13,65,119,21,104,112,-15,137,139,101)

n <- length(x)
library(R2OpenBUGS)
daysmodel <- function() {
for (i in 1:n) {
  y[i] ~ dlogis(mu,taou)
}
mu ~ dflat()
taou ~ dgamma(2,0.0001)
}
write.model(daysmodel,'daysmodel.txt')
bugsdata <- list(y=x,n=n)
inits <- function() { list(mu = mean(x), taou = 1/sd(x)) }
daysoutput <- bugs(bugsdata,inits,c('mu','taou'),40000,'daysmodel.txt',debug=TRUE)
```

Likelihood and priors ✔ ✔ ✔. Decent initial values and enough burnin and simulations to achieve clean posterior. ✔ ✔Successful simulation ✔.

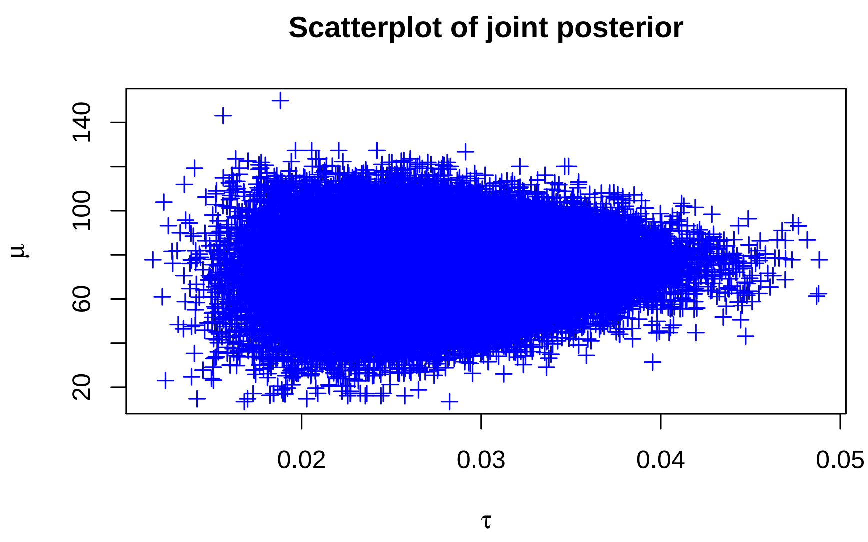
(c) Use the output to calculate the probability that the mean (μ) is more than 100. [1]

```
attach.bugs(daysoutput)
mean(mu>100)
```

≈ 0.03 ✔

(d) Also give a neat scatter plot of μ versus τ. [3]

```
windows(6,4)
plot(taou,mu,main='Scatterplot of joint posterior',xlab=expression(tau),ylab=
    expression(mu),col=3)
```



Scatterplot ✔ ✔labels ✔

(e) Fit a Normal distribution to the data by simulating the full posterior distribution of the parameters. You are advised to use a flat prior for μ and a $Gamma(0.001, 0.001)$ prior for $\tau = \sigma^{-2}$. Determine which distribution fits better by comparing the DIC values. [6]

```
daysmodelN <- function() {
for (i in 1:n) {
  y[i] ~ dnorm(mu,taou)
}
mu ~ dflat()
taou ~ dgamma(0.001,0.001)
}
write.model(daysmodelN,'daysmodelN.txt')
bugsdata <- list(y=x,n=n)
inits <- function() { list(mu = mean(x), taou = 1/sd(x)) }
daysoutputN <- bugs(bugsdata,inits,c('mu','taou'),40000,'daysmodelN.txt',debug=TRUE)
daysoutputN$DIC
daysoutput$DIC
```

Likelihood and priors ✔ ✔. Decent initial values and enough burnin and simulations to achieve clean posterior. ✔Successful simulation ✔. DIC values ✔and saying which model is better according to lowest DIC ✔. Normal: 261.6 < Logistic: 262.4

Question 2

The odds ratios measuring the effect of smoking on lung cancer from seven studies are given in 'ratios.csv'. Also given are 95% lower and upper confidence limits. Your task is to combine the results of the seven studies and estimate the underlying odds ratio.

It is known that the log odds ratio is asymptotically Normal. We thus consider the following model:

$$\ln \widehat{OR}_i \sim N(\theta_i,\ \hat{\sigma}_i^2)$$
$$\theta_i \sim N(\theta,\ \sigma_\theta^2)$$
$$\theta \sim N(0,\ 1000)$$
$$\tau_\theta = \sigma_\theta^{-2} \sim Gamma(0.001\ , 0.001)$$

(a) Before implementing the model you need to transform your data to the log scale. You also need to estimate $\hat{\tau}_i = \hat{\sigma}_i^{-2}$ for each i. Do this, and give your 14 answers in a table. You may use the natural approximation $\hat{\sigma}_i = \frac{\ln(U_i/L_i)}{2\times 1.959964}$. [3]

```
(ratiosOriginal <- read.csv(file.choose()))
attach(ratiosOriginal)

(LOR <- log(OddsRatio))
(LL95 <- log(L95))
(LU95 <- log(U95))
(sighat <- (LU95 - LL95)/qnorm(0.975)/2)
(tauhat <- sighat^(-2))

Estimates <- cbind(LOR,tauhat)
```

✓✓

Estimates	LOR	tauhat
1	1.3584	1.8596
2	1.3788	11.0345
3	1.3558	18.9371
4	2.8605	91.9747
5	1.6771	6.2259
6	2.2083	15.9578
7	1.2267	173.2233

✓

(b) Obtain 100,000 simulations of θ and give the standard deviation. [6]

```
library(R2OpenBUGS)
ORmodel <- function() {
  for (i in 1:n) {
    y[i] ~ dnorm(thetas[i],taus[i])
    thetas[i] ~ dnorm(theta,taou)
  }
  theta ~ dnorm(0,0.001)
  taou ~ dgamma(0.001,0.001)
}
write.model(ORmodel,'ORmodel.txt')

BUGSdata <- list(y=LOR,taus=tauhat,n=length(LOR))
inits <- function(){list(thetas=LOR,theta=mean(LOR),taou=mean(tauhat))}

ORoutput <- bugs(BUGSdata,inits,c('theta','taou','thetas'),60000,'ORmodel.txt'
    ,2,10000,debug=TRUE)

attach.bugs(ORoutput)
sd(theta)
```

✔ ✔ ✔ ✔ ✔

0.3110023 ✔.

(c) Based on your simulations, estimate the posterior median of θ and give a 90% symmetric interval for θ itself. [2]

```
median(theta)
quantile(theta,c(0.05,0.95),6)
```

1.77 ✔

5%	95%	
1.264	2.260	✔

(d) Now estimate the posterior mean of the overall odds ratio (e^{θ}) and give a 95% HPD interval for the odds ratio. [3]

```
postOR <- exp(theta)
mean(postOR)
hpd.interval(postOR,0.05)
```

6.1477 ✔ ✔

2.7374 10.0543 ✔

Total for Question 2: 14

Total half marks on memo = 70 vs. 70 = Double total margin points (=35).