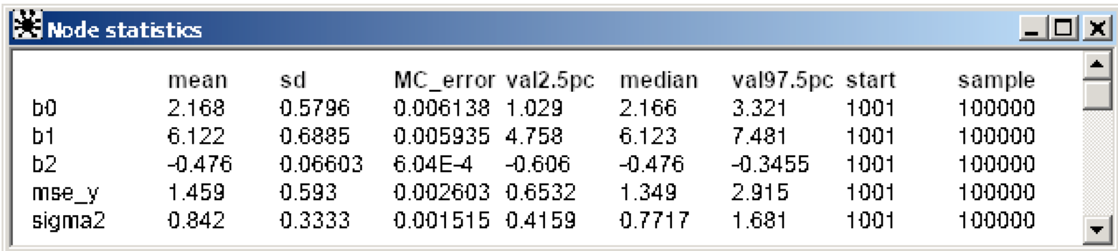
ISyE 6420
Solution **Homework 5** November 17, 2019

1 Cross-validating a Bayesian Regression.

By running OpenBUGS code, we have the result shown in Figure 1. The OpenBUGS code is attached in Appendix A. The Bayesian estimator of $\beta_0, \beta_1, \beta_2$ and σ are close to the true


Node statistics

	mean	sd	MC_error	val2.5pc	median	val97.5pc	start	sample
b0	2.168	0.5796	0.006138	1.029	2.166	3.321	1001	100000
b1	6.122	0.6885	0.005935	4.758	6.123	7.481	1001	100000
b2	-0.476	0.06603	6.04E-4	-0.606	-0.476	-0.3455	1001	100000
mse_y	1.459	0.593	0.002603	0.6532	1.349	2.915	1001	100000
sigma2	0.842	0.3333	0.001515	0.4159	0.7717	1.681	1001	100000

Figure 1: OpenBUGS result for problem 1

values.

2 Body Fat from Linear Regression.

(a) We consider five different models with single predictor and compare the Bayesian R^2 of each model. The result is shown in Table 1. We choose BAI as the single best predictor.

Predictor	Bayesian R^2	b_0	b_1	τ
Age	0.08494	19.31	0.2206	0.01638
BAI	0.5474	-5.97	1.183	0.03311
BMI	0.2965	3.611	0.9619	0.0213
BB	0.4763	11.95	0.02161	0.02862
Gender	0.232	23.7	7.883	0.01951

Table 1: The result of Bayesian R^2 for five models with single predictor

The OpenBUGS code of both models are attached in Appendix B.

(b) We obtain the result for the prediction based on two model. The result of the predicted BF is shown in 2 and 3, respectively. The predicted BF is 15.12 and 24.81 based on the first and second model, respectively.

Node statistics

	mean	sd	MC_error	val2.5pc	median	val97.5pc	start	sample
BFpredict	15.12	4.057	0.0159	7.157	15.13	23.02	1001	100000
BR2	0.7521	0.006201	2.101E-5	0.7397	0.7522	0.764	1001	100000
b0	-33.01	1.808	0.09751	-36.5	-33.11	-29.25	1001	100000
b1	0.07333	0.007253	1.544E-4	0.05913	0.07331	0.08759	1001	100000
b2	0.7737	0.06177	0.003224	0.6439	0.7744	0.8928	1001	100000
b3	1.896	0.07539	0.004006	1.739	1.901	2.036	1001	100000
b4	-0.02403	0.002109	1.138E-4	-0.0281	-0.02413	-0.01967	1001	100000
b5	10.61	0.2212	0.007157	10.18	10.6	11.04	1001	100000
sigma2	16.55	0.4141	0.001403	15.76	16.54	17.38	1001	100000

Figure 2: Predicted BF based on the first model

Node statistics

	mean	sd	MC_error	val2.5pc	median	val97.5pc	start	sample
BFpredict	24.81	5.509	0.01785	14.06	24.82	35.61	1001	100000
BR2	0.5474	0.01132	3.655E-5	0.5248	0.5476	0.569	1001	100000
b0	-5.971	0.55	0.01347	-7.043	-5.97	-4.893	1001	100000
b1	1.183	0.01888	4.624E-4	1.146	1.183	1.22	1001	100000
sigma2	30.22	0.7556	0.00244	28.78	30.21	31.73	1001	100000

Figure 3: Predicted BF based on the second model

3 Shocks.

By the description, we model the responses via a logistic regression as:

$$p(x) \sim \text{logit}(\beta_0 + \beta_1 \cdot x),$$

where x denotes the shocks time. The OpenBUGS code is provided in Appendix C. Eventually, we obtain the following result:

Node statistics

	mean	sd	MC_error	val2.5pc	median	val97.5pc	start	sample
pstart	0.4532	0.03408	4.271E-5	0.3868	0.4532	0.5202	1	1000000

Figure 4: 95% Credible Set for the Population Proportion under the New Setting.

A Code for Problem 1

```
model {
# train the linear regression model
for (i in 1:20) {
y[i] ~ dnorm(mu[i], tau)
mu[i] <- b0 + b1 * x1[i] + b2 * x2[i]
}
b0 ~ dnorm(0, 0.001)
b1 ~ dnorm(0, 0.001)
b2 ~ dnorm(0, 0.001)
tau ~ dgamma(0.001, 0.001)
sigma2 <- 1/tau

# evaluate the model
for (i in 21:40) {
mu_hat[i] <- b0 + b1 * x1[i] + b2 * x2[i]
y_hat[i-20] ~ dnorm(mu_hat[i], tau)
se[i-20] <- pow((y_hat[i] - y[i]), 2)
}
mse_y <- (1/20)*sum(se[])
}

DATA
x1[] x2[] y[]
0.0356 3 0.1020
0.8862 9 3.1455
0.2469 5 0.3497
0.0089 9 -2.3172
0.8149 1 7.9718
0.1405 9 -1.0208
0.8799 1 7.6091
0.0954 9 -0.0395
0.3526 9 0.5557
0.5934 6 2.8787
0.5852 2 6.0721
0.6677 2 4.2824
0.6480 8 1.3424
0.4334 9 1.7148
0.1398 1 2.7020
0.7519 8 2.4779
0.2418 10  -1.1827
```

```
0.6505 4 3.6914
0.8574 7 3.4107
0.0844 4 0.1968
0.9721 3 5.8893
0.0315 8 -2.5662
0.8354 8 2.9645
0.8357 3 5.0545
0.0499 6 0.4526
0.5459 5 2.9637
0.9432 1 7.9761
0.3215 1 3.3170
0.8065 5 4.2736
0.6014 3 5.1150
0.7896 4 5.9125
0.7992 3 5.3215
0.0496 2 2.6514
0.2832 4 2.0798
0.6535 7 2.7156
0.4897 3 5.1915
0.9729 7 5.5657
0.7485 5 4.7167
0.5678 5 3.2289
0.2990 2 2.3631
END

INITS
list(b0=1, b1=0, b2=0, tau=1)
```

B Code for Problem 2

```
model{

for(i in 1:N){
BF[i] ~ dnorm(mu[i], tau)
BB[i] <- BAI[i] * BMI[i]

# model 1
mu[i] <- b0 + b1 * Age[i] + b2*BAI[i] + b3*BMI[i] + b4*BB[i] + b5* Gender[i]


# model 2
```

```
#mu[i] <- b0 + b1 * BAI[i]
}

# find a prediction
Age.new <- 35
BAI.new <- 26
BMI.new <- 20
Gender.new <- 0
BB.new <- 520

# model 1
BFmean <- b0+b1*Age.new+b2*BAI.new+b3*BMI.new+b4*BB.new+b5*Gender.new

# model 2
#BFmean <- b0 + b1 * BAI.new

BFpredict ~ dnorm(BFmean, tau)

b0 ~ dnorm(0, 0.001)
b1 ~ dnorm(0, 0.001)
b2 ~ dnorm(0, 0.001)
b3 ~ dnorm(0, 0.001)
b4 ~ dnorm(0, 0.001)
b5 ~ dnorm(0, 0.001)
tau ~ dgamma(0.001, 0.001)
sigma2 <- 1/tau

p <- 6
# Bayesian R^2
nminusp <- N-p
sse <- nminusp * sigma2
for (i in 1:N) {
CBF[i] <- BF[i] - mean(BF[])
}
sst <- inprod(CBF[], CBF[])
BR2 <- 1 - sse/sst

}
```

C Code for Problem 3

```
model{
for (i in 1:N){
Y[i] ~ dbin(p[i], 70)
logit(p[i]) <- beta0 + beta1 * X[i]
}
beta0 ~ dnorm(0.0, 1.0E-6)
beta1 ~ dnorm(0.0, 1.0E-6)

# New subject
X_new <- 2.5
logit(pstart) <- beta0 + beta1 *  X_new
}

DATA
list(N=6, X=c(0,1,2,3,4,5), Y=c(0,9,21,47,60,63))

INITS
list(beta0=0,beta1=0)
```