# Bayes example for guidance paper

*Paul Gustafson*

*September 11, 2018*

## Data wrangling

```
require(xlsx)
```

```
## Loading required package: xlsx
```

```
dtf <- read.xlsx("open_data_20180607.xlsx", 1)

require(tidyverse)
```

```
## Loading required package: tidyverse
```

```
## -- Attaching packages --------------------------------------------------------- tidyverse 1.2.1 --
```

```
## v ggplot2 3.0.0     v purrr   0.2.5
## v tibble  1.4.2     v dplyr   0.7.6
## v tidyr   0.8.1     v stringr 1.3.1
## v readr   1.1.1     v forcats 0.3.0
```

```
## -- Conflicts ------------------------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
dtf <- dplyr::mutate(dtf, xstr=log(POTASS_FFQ1/ENERGY_FFQ1),
                          r1=log(POTASS_UR1/TEE1),
                          r2=log(POTASS_UR2/TEE2))

### look at naive regression
summary(lm(BMI ~ xstr + AGE + GENDER, data=dtf))
```

```
##
## Call:
## lm(formula = BMI ~ xstr + AGE + GENDER, data = dtf)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -10.2326  -3.6116  -0.9421   2.6134  23.6675
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 27.27970    1.76370  15.467   <2e-16 ***
## xstr        -1.68862    0.93369  -1.809   0.0711 .
## AGE          0.03885    0.02941   1.321   0.1871
## GENDER      -0.37743    0.49134  -0.768   0.4428
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.263 on 479 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.01201,    Adjusted R-squared:  0.005827
## F-statistic: 1.942 on 3 and 479 DF,  p-value: 0.122
```

### Bayesian model specification

```
genmod.string <- "model{

gm.0 ~ dnorm(0,.001)
gm.age ~ dnorm(0,.001)
gm.gnd ~ dnorm(0,.001)
prc.x ~ dgamma(.01,.01)

bt.0 ~ dnorm(0,.001)
bt.x ~ dnorm(0,.001)
bt.age ~ dnorm(0,.001)
bt.gnd ~ dnorm(0,.001)
prc.y ~ dgamma(.01,.01)

al.0 ~ dnorm(0,.001)
al.x ~ dnorm(0,.001)
al.age ~ dnorm(0,.001)
al.gnd ~ dnorm(0,.001)
prc.xstr ~ dgamma(.01,.01)

prc.r1 ~ dgamma(.01,.01)
prc.r2 ~ dgamma(.01,.01)

for (i in 1:n) {
  x[i] ~ dnorm(gm.0 + gm.age*age[i] + gm.gnd*gnd[i], prc.x)
  y[i] ~ dnorm(bt.0 + bt.x*x[i] + bt.age*age[i] + bt.gnd*gnd[i], prc.y)
  xstr[i] ~ dnorm(al.0 + al.x*x[i] + al.age*age[i] + al.gnd*gnd[i], prc.xstr)
}

for (i in 1:n.sub) {
  r1[i] ~ dnorm(x[i], prc.r1)
  r2[i] ~ dnorm(x[i], prc.r2)
}

}"

require(rjags)
```

```
## Loading required package: rjags

## Loading required package: coda

## Linked to JAGS 4.3.0

## Loaded modules: basemod,bugs
```

```
mod <- jags.model(
  textConnection(genmod.string),
  data=list(y=dtf$BMI, xstr=dtf$xstr, r1=dtf$r1, r2=dtf$r2,
            age=dtf$AGE, gnd=dtf$GENDER, n=dim(dtf)[1], n.sub=250),
  n.chains=3)
```

```
## Compiling model graph
##    Resolving undeclared variables
##    Allocating nodes
## Graph information:
##    Observed stochastic nodes: 1210
##    Unobserved stochastic nodes: 758
##    Total graph size: 5255
##
## Initializing model
```

## Produce MCMC output

```
set.seed(13)  ### want beta.hat insensitive to seed, to plus/minus .01 say

update(mod, 5000) #burn-in

opt.jags <- coda.samples(mod, n.iter=500000, thin=50,
                         variable.names=c("bt.x", "al.0", "al.x", "al.age", "al.gnd","prc.y", "prc.x",
```

## Check traceplots

```
require(rjags)
```

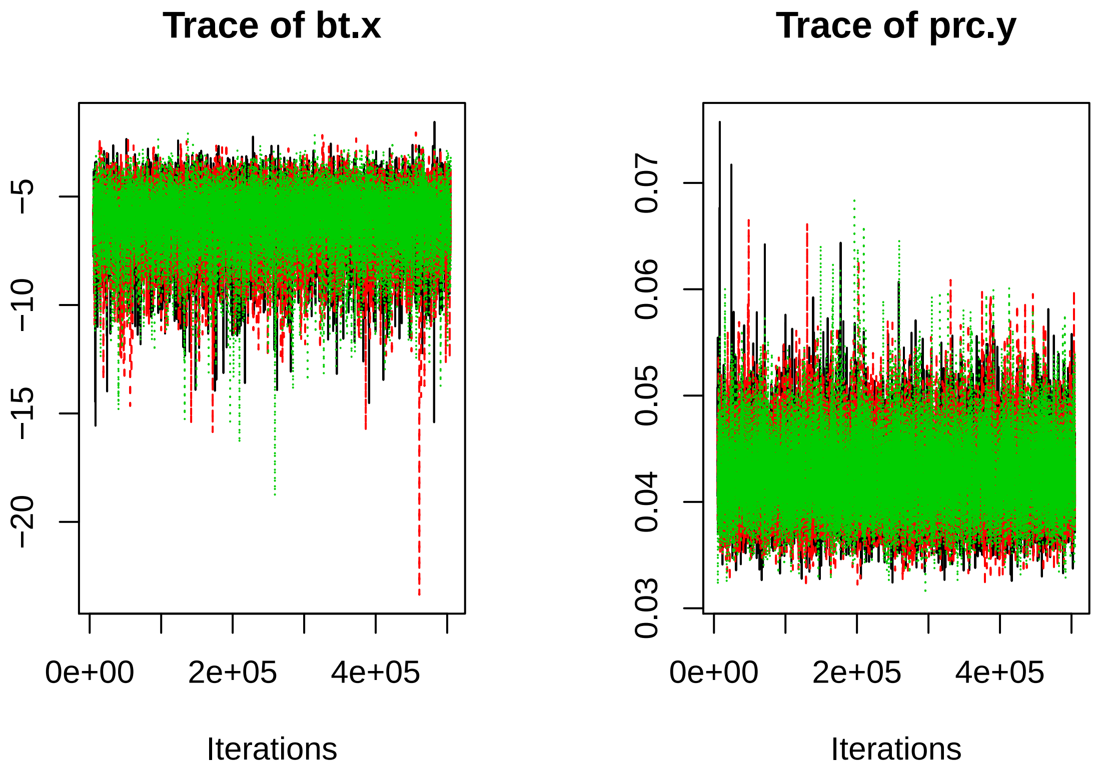
```
## Loading required package: rjags
```

```
## Loading required package: coda
```

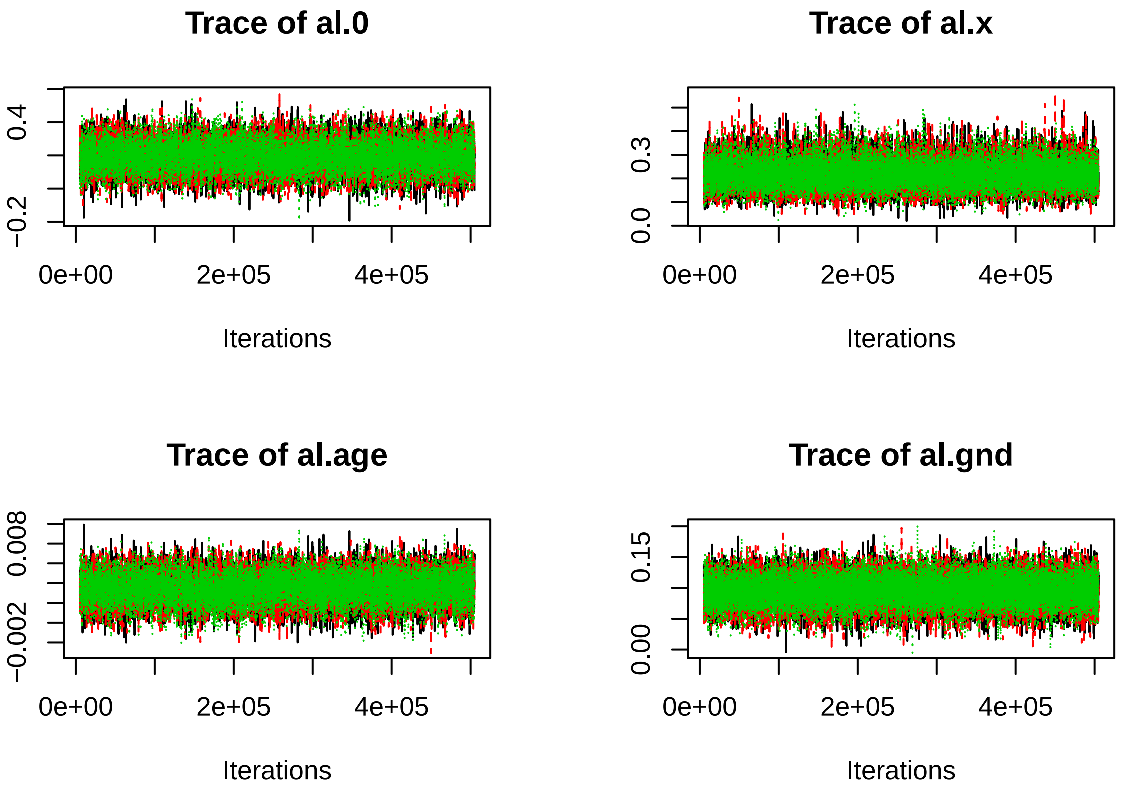
```
## Linked to JAGS 4.3.0
```

```
## Loaded modules: basemod,bugs
```

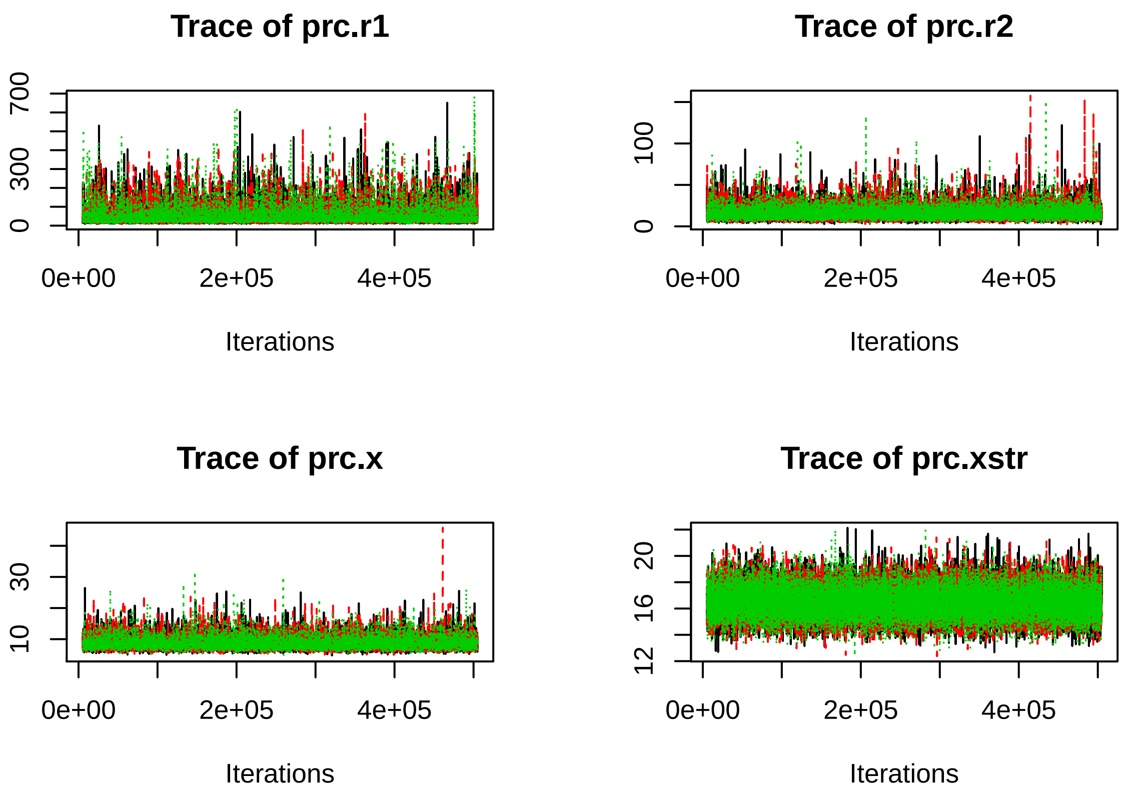
```
par(mfrow=c(1,2))
traceplot(opt.jags[,c("bt.x","prc.y")])
```



```r
par(mfrow=c(2,2))
traceplot(opt.jags[,c("al.0","al.x","al.age","al.gnd")])
```



```r
par(mfrow=c(2,2))
traceplot(opt.jags[,c("prc.r1","prc.r2","prc.x","prc.xstr")])
```



### Inference summaries in general

```
summary(opt.jags)
```

```
##
## Iterations = 5050:505000
## Thinning interval = 50
## Number of chains = 3
## Sample size per chain = 10000
##
## 1. Empirical mean and standard deviation for each variable,
##    plus standard error of the mean:
##
##               Mean        SD  Naive SE Time-series SE
## al.0      0.196429  0.090135 5.204e-04      8.220e-04
## al.age    0.003489  0.001526 8.810e-06      1.383e-05
## al.gnd    0.095132  0.024193 1.397e-04      1.521e-04
## al.x      0.217684  0.058538 3.380e-04      4.089e-04
## bt.x     -6.082040  1.425443 8.230e-03      1.201e-02
## prc.r1   50.217420 48.428993 2.796e-01      4.698e-01
## prc.r2   15.783393  8.361564 4.828e-02      5.118e-02
## prc.x     8.729091  2.117072 1.222e-02      1.896e-02
## prc.xstr 16.484532  1.202749 6.944e-03      6.944e-03
## prc.y     0.042673  0.003659 2.113e-05      2.280e-05
##
```

```
## 2. Quantiles for each variable:
##
##                2.5%       25%      50%       75%     97.5%
## al.0      0.0207289  0.135548  0.19586  0.255830   0.37645
## al.age    0.0004142  0.002471  0.00350  0.004521   0.00645
## al.gnd    0.0476083  0.078951  0.09503  0.111375   0.14246
## al.x      0.1126051  0.178429  0.21374  0.252761   0.34558
## bt.x     -9.3956942 -6.820615 -5.91849 -5.135711  -3.78521
## prc.r1   12.9872045 22.315294 33.97098 58.294730 188.84791
## prc.r2    5.9740090 10.444226 13.95472 18.883758  36.35952
## prc.x     6.1105319  7.319191  8.26175  9.605445  14.07561
## prc.xstr 14.2854394 15.654088 16.43320 17.260691  18.98890
## prc.y     0.0363033  0.040160  0.04244  0.044876   0.05057
```

## Inference summaries for $beta_x$

```
### focus on beta.x
tmp <- as.matrix(opt.jags[,"bt.x"])

### point estimate - posterior mean
round(mean(tmp),2)
```

```
## [1] -6.08
```

```
### analog to SE, the posterior SD
round(sqrt(var(tmp)),2)
```

```
##      var1
## var1 1.43
```

```
### 95% equal-tailed posterior credible interval
round(quantile(tmp, c(.025, .975)),2)
```

```
##  2.5% 97.5%
## -9.40 -3.79
```

```
### how much posterior probability above zero?
mean(tmp>0)
```

```
## [1] 0
```