Bayesian Model Averaging

PUBH 8442: Bayes Decision Theory and Data Analysis

Eric F. Lock
UMN Division of Biostatistics, SPH
elock@umn.edu

04/14/2024

Example: Body Fat (cont.)

- Recall: % body fat ($BF\%$) measured for 100 adult males.
- Also measured 9 predictor variables
  - *Age*, *Weight*, *Height*; circumference of *neck*, *chest*, *abdomen*, *ankle*, *bicep*, and *wrist*.
- Consider the model

$$\mathbf{y} = \boldsymbol{\beta} X + \boldsymbol{\epsilon}$$

  where

  - $\mathbf{y}$ is population-centered $BF\%$
  - X is the standardized matrix of predictor variables
  - $\epsilon \sim \text{Normal}(\mathbf{0}, \sigma^2 I)$

Example: Body Fat (cont.)

- Previously used iid normal prior for $\beta_i's$:

$$\beta \sim \text{Normal}(0, 0.62\sigma^2 I)$$

with $\hat{\tau}^2 = 0.62$ estimated empirically.

- $IG(3, 20)$ prior for σ^2
- Now, allow some $\beta_i's$ to be identically 0, with probability 1/2:

$$\beta_i \sim \begin{cases} 0 \text{ with probability } 1/2 \\ N(0, 0.62\sigma^2) \text{ with probability } 1/2 \end{cases}$$

- Equivalently, incorporate model inclusion indicators ζ:

$$y_i = \zeta_1\beta_1 x_{i1} + \zeta_2\beta_1 x_{i2} + \ldots + \zeta_9\beta_9 x_{i9} + \epsilon_i$$

where the $\zeta_i's$ have iid Bernoulli(1/2) priors.

Example: Body Fat (cont.)

- rjags code:
  http://www.ericfrazerlock.com/bma_rjags.R

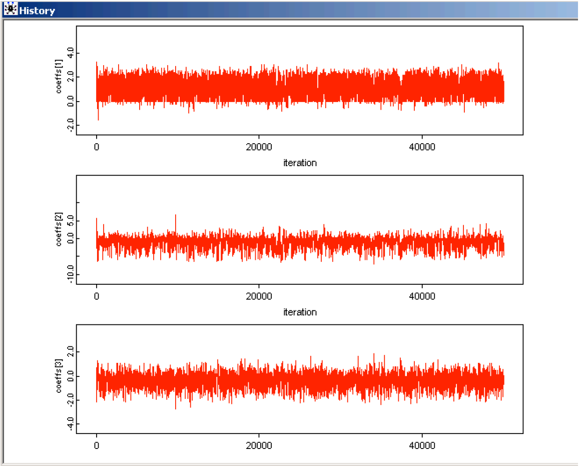
Example: Body Fat (cont.)

- Data, in BUGS list format, are available at http://www.ericfrazerlock.com/BMI_for_WinBugs.txt
- BUGS model specification:

```
model{
  #sampling model
   for(i in 1:100){
     Y[i] ~ dnorm(mu[i], Prec)
     mu[i] <- coeffs[1]*X1[i]+coeffs[2]*X2[i]+coeffs[3]*X3[i]+
              coeffs[4]*X4[i]+ coeffs[5]*X5[i]+coeffs[6]*X6[i]+
              coeffs[7]*X7[i]+coeffs[8]*X8[i]+coeffs[9]*X9[i]}
   #Priors
   PrecBeta <- (1/0.62)*Prec
   for(j in 1:9){
     beta[j] ~ dnorm(0,PrecBeta)
     zeta[j] ~ dbern(0.5)
     coeffs[j]<-beta[j]*zeta[j]}
   Prec~dgamma(3,20)
}
```

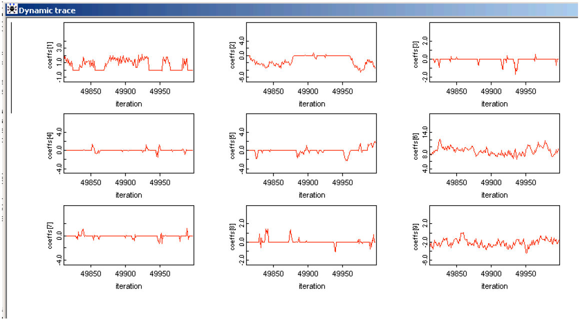
Example: Body Fat (cont.)

- MCMC history, coefficients 1-3



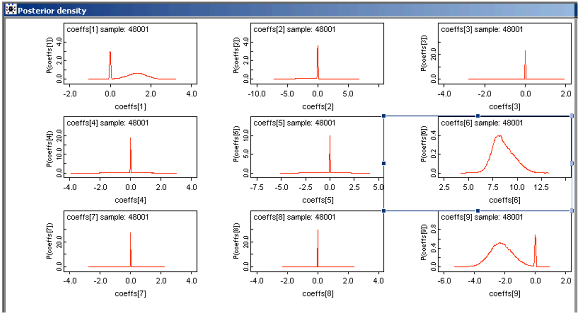
Example: Body Fat (cont.)

- Final 200 draws, all coefficients:



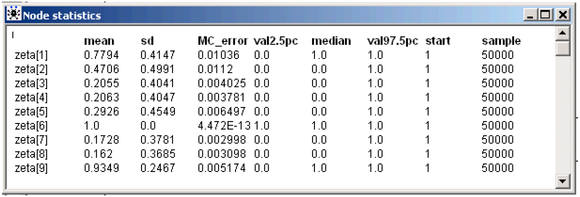
Example: Body Fat (cont.)

- Coefficient density estimates (2000 burn-in):



Example: Body Fat (cont.)

- Zeta draw statistics:



	mean	sd	MC_error	val2.5pc	median	val97.5pc	start	sample
zeta[1]	0.7794	0.4147	0.01036	0.0	1.0	1.0	1	50000
zeta[2]	0.4706	0.4991	0.0112	0.0	0.0	1.0	1	50000
zeta[3]	0.2055	0.4041	0.004025	0.0	0.0	1.0	1	50000
zeta[4]	0.2063	0.4047	0.003781	0.0	0.0	1.0	1	50000
zeta[5]	0.2926	0.4549	0.006497	0.0	0.0	1.0	1	50000
zeta[6]	1.0	0.0	4.472E-13	1.0	1.0	1.0	1	50000
zeta[7]	0.1728	0.3781	0.002998	0.0	0.0	1.0	1	50000
zeta[8]	0.162	0.3685	0.003098	0.0	0.0	1.0	1	50000
zeta[9]	0.9349	0.2467	0.005174	0.0	1.0	1.0	1	50000

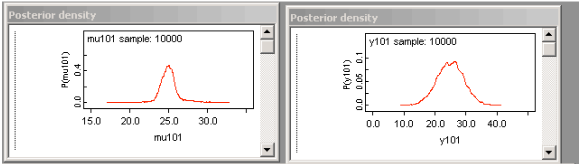
- Abdominal circumference (6) is non-zero for all draws
- Wrist circumference (9) is included in 93%, Age (1) in 78%
- Remaining variables are most often excluded from model

Example: Body Fat (cont.)

- Assume an individual has the following standardized measurements:
  - *Age*: 1.20, *Weight*:0.5, *Height*:-0.4; circumference of *neck*:2.2, *chest*:0.3, *abdomen*:0.8, *ankle*:0.2, *bicep*:0.3, and *wrist:0.6*.
- Let μ_{101} be mean %BF for above measurements, and y_{101} be the individual's BMI
- Compute for each MCMC iteration
  - For each draw, some subset of coefficients are 0
- WinBUGS code in model:

```
mu101 <- 18.5+coeffs[1]*1.2+coeffs[2]*0.5-coeffs[3]*0.4
         -coeffs[4]* 2.2+coeffs[5]*0.3+coeffs[6]*0.8
         +coeffs[7]*0.2+coeffs[8]*0.3+coeffs[9]*0.6
y101 ~ dnorm(mu101,Prec)
```

- Posterior for μ_{101} and y_{101}:



Node statistics

node	mean	sd	MC_error	val2.5pc	median	val97.5pc	start	sample
mu101	25.03	1.266	0.03202	22.89	24.95	28.31	50001	10000

Node statistics

node	mean	sd	MC_error	val2.5pc	median	val97.5pc	start	sample
y101	25.03	4.321	0.05307	16.51	25.08	33.49	50001	10000

Bayesian model averaging (BMA)

- For many applications, consider multiple models $M_1, \ldots, M_m$
- Rather than selecting the "best" model, combine them
- Let γ be a quantity that is defined in every model
- Prior probabilities $p(M_i)$ define the combined prior distribution

$$p(\gamma) = \sum_{i=1}^{m} p(\gamma \mid M_i) p(M_i)$$

- The combined posterior distribution for γ is

$$p(\gamma \mid \mathbf{y}) = \sum_{i=1}^{m} p(\gamma \mid M_i, \mathbf{y}) p(M_i \mid \mathbf{y})$$

where

$$p(M_i \mid \mathbf{y}) = \frac{p(\mathbf{y} \mid M_i) p(M_i)}{\sum_{j=1}^{m} p(\mathbf{y} \mid M_i) p(M_i)}.$$

Bayesian model averaging (BMA)

- In previous example, considered $m = 2^9 = 512$ models
  - Each model M_i includes some subset of 9 predictor variables
  - Prior: $p(M_i) = 2^{-9}$ for $i = 1, \ldots, m$
- If $p(M_1) = \cdots = p(M_m)$, $p(M_i \mid \mathbf{y}) \propto p(\mathbf{y} \mid M_i)$ and

$$\frac{P(M_i \mid \mathbf{y})}{P(M_j \mid \mathbf{y})}$$

  is the Bayes factor of model i over model j.
- Under posterior sampling, $p(M_i \mid \mathbf{y})$ is approximated by the proportion of times model i is chosen
  - Useful for computing Bayes factors.

BMA comments

- Even if no parameters are shared, can use BMA for posterior predictive: $\gamma = y_{n+1}$
- BMA appropriately accounts for uncertainty in model choice
  - Selecting the single "best" model can underestimate uncertainty, overfit.
- In practice, may consider a large set of models, but omit less plausible models in BMA
  - For example, only include those models i such that

$$\frac{\max_j p(M_j \mid \mathbf{y})}{p(M_i \mid \mathbf{y})} < C$$

For some C (say, $C = 20$).