AMS 207
Model comparison examples

Department of Applied Mathematics and Statistics
University of California, Santa Cruz

April 19, 2016

Deviance

- deviance:$D(y,\theta) = -2\log p(y \mid \theta)$
- posterior estimate: $D_{\hat{\theta}}(y) = D(y, \hat{\theta}(y))$

BIC

- penalty: $p_{BIC} = k \log n$
- k is the number of estimated parameters, and n is the sample size
- $BIC = D_{\hat{\theta}}(y) + p_{BIC}$

DIC

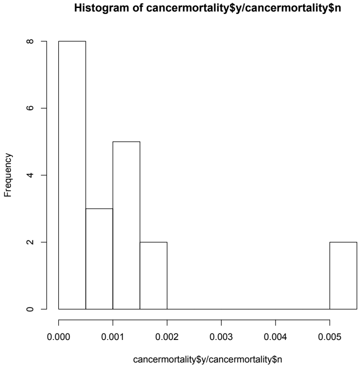
- posterior expected deviance: $D_{avg} = \mathbb{E}_{post}[\log p(y \mid \theta)]$
- which is estimated, using posterior samples, as

$$\hat{D}_{avg} = \frac{1}{S} \sum_{t=1}^{S} D(y, \theta^{(t)})$$

- penalty: $p_{DIC} = 2(D_{\hat{\theta}}(y) - D_{avg})$
- $DIC = D_{\hat{\theta}}(y) + p_{DIC}$

Cancer mortality data

- Number of stomach cancer deaths and individuals at risk in a city in Missouri
- data available in R package LearnBayes
- see Albert (2009), p. 91
- library(LearnBayes)
- data(cancermortality)



Model A

- Binomial likelihood (known trials), common success probability
$$y_i \sim \mathcal{B}in(\theta; n_i)$$
- conjugate prior
$$\theta \sim \mathcal{B}eta(\alpha, \beta)$$
- α and β fixed

Binomial likelihood

```
lbinlike=function(theta,data){
        y = data[, 1]
        n = data[, 2]
        N = length(y)
        val=sum(lchoose(n,y))+sum(y)*log(theta)+sum(n-y)*log(1-theta)
        return(val)
}
```

(direct) Posterior sampling

```
## prior parameters
alpha=1
beta=1

## posterior parameters
alpha_post=alpha+sum(cancermortality$y)
beta_post=beta+sum(cancermortality$n)-sum(cancermortality$y)

##
S=10000
post_sample<-rbeta(S,alpha_post,beta_post)
```

BIC & DIC

```
hlp<-NULL
for(t in 1:S){
        hlp[t]<-lbinlike(post_sample[t],cancermortality)
}

lph<-lbinlike(mean(post_sample),cancermortality)

pdic=2*(lph-mean(hlp))
DIC_A=-2*lph+2*pdic

N<-length(cancermortality$y)
BIC_A=-2*lph+log(N)
```

Model B

- Beta-Binomial likelihood

$$y_i \sim \mathcal{B}e\text{-}\mathcal{B}in(\eta, K)$$

- unknown parameters; vague prior

$$\pi(\eta, K) \propto \frac{1}{\eta(1-\eta)} \frac{1}{(1+K)^2}$$

Likelihood

```
##original parametrization
 lbetabinlike=function (params, data){
        eta = params[1]
        K = params[2]
        y = data[, 1]
        n = data[, 2]
        val = sum(lbeta(K * eta + y, K * (1 - eta) + n - y)
                       - lbeta(K * eta, K * (1 - eta))+ lchoose(n,y))
        return(val)
}

##transformed parameters (logit(eta),log(K))
 lbetabinliketrans=function (theta, data){
        eta = exp(theta[1])/(1 + exp(theta[1]))
        K = exp(theta[2])
        y = data[, 1]
        n = data[, 2]
        val = lbetabinlike(c(eta,K),data)
        val = val + theta[1]+theta[2] - 2 * log(1 + exp(theta[1]))
        return(val)
}
```

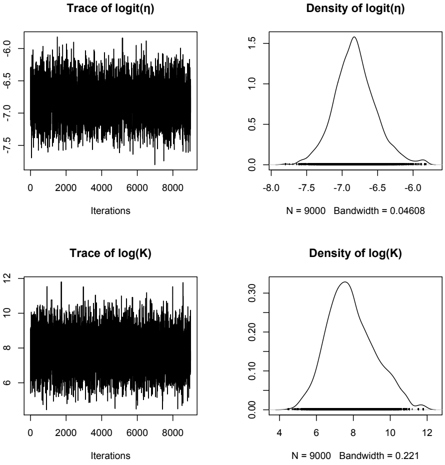
Posterior

```
## transformed parameters
 lbetabinposttrans=function (theta, data){
         eta = exp(theta[1])/(1 + exp(theta[1]))
         K = exp(theta[2])
         val = lbetabinliketrans(theta,data)
         val = val -log(eta)-log(1-eta) - 2 * log(1 + K)
         return(val)
}
```

Posterior sampling

```
##  get Laplace approximation; use as proposal
fit=laplace(lbetabinposttrans,c(-7,6),cancermortality)


## sampling from posterior (M-H)
S=10000
X<-matrix(,S,2)

X[1,]<-c(-7,6)
for(t in 2:S){
        Y<-rmnorm(1,fit$mode,fit$var)
        lrho<-lbetabinposttrans(Y,cancermortality)+dmnorm(X[t-1,],fit$mode,fit$var,log=TRUE)
                        -lbetabinposttrans(X[t-1,],cancermortality)-dmnorm(Y,fit$mode,fit$var,log=TRUE)
        X[t,]<-X[t-1,]+(Y-X[t-1,])*(log(runif(1))<lrho)
}

## burn-in
BURN=1000
X<-X[-(1:BURN),]

## sanity check
library(MCMCpack)
plot(as.mcmc(X))
```

Convergence diagnosis



BIC & DIC

```
##transform samples original parameters
post_sample<-cbind(exp(X[,1])/(1+exp(X[,1])),exp(X[,2]))

hlp<-NULL
pred<-NULL
for(t in 1:(S-BURN)){
        hlp[t]<-lbetabinlike(post_sample[t,],cancermortality)
        pred[t]<-rbeta(1,post_sample[t,1],post_sample[t,2])
}

lph<-lbetabinlike(colMeans(post_sample),cancermortality)

pdic=2*(lph-mean(hlp))
DIC_B=-2*lph+2*pdic

N<-length(cancermortality$y)
BIC_B=-2*lph+log(N)
```

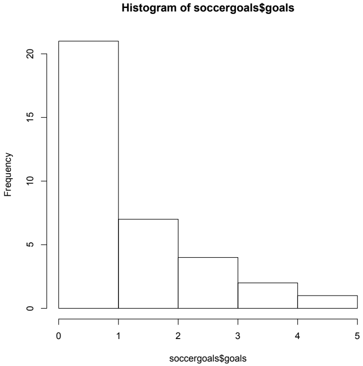
Results

	BIC	DIC
Model A	56.88	55.84
Model B	59.23	61.67

- Both BIC and DIC slightly favor the simpler model A.

Soccer goals data

- Also from Albert (2009), see p. 187
- data available in R package LearnBayes
- number of goals scored by a team in the Major League Soccer for the 2006 season in 35 games.
- data(soccergoals)



Model A

- Poissson likelihood, common rate

$$y_i \sim \mathcal{P}oiss(\lambda)$$

- conjugate prior

$$\lambda \sim \mathcal{G}amma(\alpha, \beta)$$

- α and β fixed

(direct) Posterior sampling

```
## prior parameters
alpha=5
beta=1

## posterior parameters
N<-length(soccergoals$goals)
goal_sum<-sum(soccergoals$goals)
alpha_post=a+goal_sum
beta_post=b+N

## sample from posterior
S=10000
post_sample<-rgamma(S,alpha_post,beta_post)
```

Posterior predictive

```
pred_A<-NULL
for(t in 1:S){
        pred_A[t]<-rpois(1,post_sample[t])
}
```

BIC & DIC

```
hlp<-dpois(goal_sum,N*post_sample,log=TRUE)

lph<-dpois(goal_sum,N*mean(post_sample),log=TRUE)

pdic=2*(lph-mean(hlp))
DIC_A=-2*lph+2*pdic

BIC_A=-2*lph+log(N)
```

Model B

- Binomial likelihood (known trials), common success probability

$$y_i \sim \mathcal{B}in(\theta; n_i)$$

- conjugate prior

$$\theta \sim \mathcal{B}eta(a, b)$$

- a and b fixed

(direct) Posterior sampling

```
n=10

## prior parameters
a=1
b=1

## posterior parameters
N<-length(soccergoals$goals)
goal_sum<-sum(soccergoals$goals)
a_post=alpha+goal_sum
b_post=beta+n*N-goal_sum

## sample from posterior
S=10000
post_sample<-rbeta(S,a_post,b_post)
```

Posterior predictive

```
pred_B<-NULL
for(t in 1:S){
        pred_B[t]<-rbinom(1,10,post_sample[t])
}
```

BIC & DIC

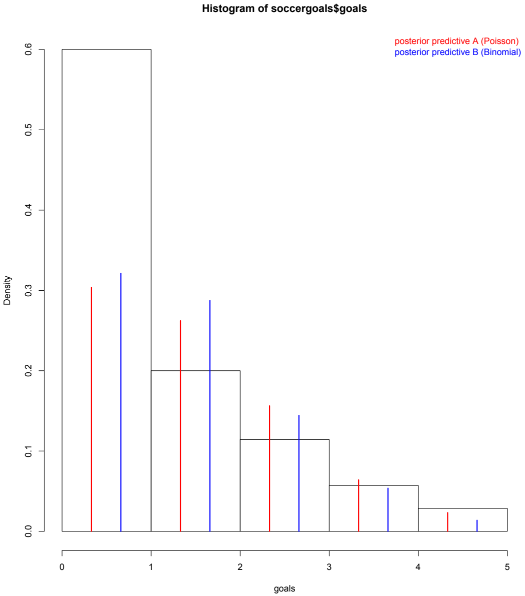
```
hlp<-dbinom(goal_sum,n*N,post_sample,log=TRUE)

lph<-dbinom(goal_sum,n*N,mean(post_sample),log=TRUE)

pdic=2*(lph-mean(hlp))
DIC_B=-2*lph+2*pdic

BIC_B=-2*lph+log(N)
```

Results

	BIC	DIC
Model A	9.07	7.97
Model B	8.71	7.67

- Models are practically indistinguishable based on BIC & DIC

Posterior predictives



References

- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian Data Analysis (3th ed.).* CRC Press.
- Albert, J. (2009). *Bayesian computation with R.* Springer Science & Business Media.