Tutorial/Lab Problems # 10
MATH3871/MATH5970

Suppose we have the following Bayesian model for the data $\boldsymbol{y} = (y_1, \ldots, y_n)^\top$:

$$\begin{aligned} \text{prior: } (\boldsymbol{\beta} \,|\, \sigma^2) &\sim \mathsf{N}(\mathbf{0}, \sigma^2 \mathbf{D}) \\ \text{likelihood (normal linear model): } (\boldsymbol{Y} \,|\, \boldsymbol{\beta}, \sigma^2) &\sim \mathsf{N}(\underbrace{\mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2}_{\mathbf{X}\boldsymbol{\beta}}, \sigma^2 \mathbf{I}) \ , \end{aligned}$$

where $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are unknown vectors of dimension k and $p-k$, respectively; and $\mathbf{X}_1$ and $\mathbf{X}_2$ are full-rank model matrices of dimensions $n \times k$ and $n \times (p-k)$, respectively. Above we implicitly defined $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2)$ and $\boldsymbol{\beta}^\top = (\boldsymbol{\beta}_1^\top, \boldsymbol{\beta}_2^\top)$. Suppose we wish to test the hypothesis $H_0 : \boldsymbol{\beta}_2 = \mathbf{0}$ against $H_1 : \boldsymbol{\beta}_2 \neq \mathbf{0}$ within a Bayesian paradigm. In other words, we wish to compare two nested models — the unrestricted model $m = 2$ having the full set of parameters $\boldsymbol{\beta}$ and the restricted model $m = 1$ with parameters $\boldsymbol{\beta}_1$. Find a formula for the Bayes factor needed to compare the two models under the assumptions that:

- σ^2 is known and fixed;
- σ^2 is unknown with an improper prior $g(\sigma^2) \propto 1/\sigma^2$.

You may use the following background facts (which you may derive if you have the background knowledge):

$$\begin{aligned} (\boldsymbol{\beta} \,|\, \sigma^2, \boldsymbol{y}) &\sim \mathsf{N}(\hat{\boldsymbol{\beta}}, \sigma^2 \Sigma), \\ \text{where } \Sigma &:= (\mathbf{D}^{-1} + \mathbf{X}^\top \mathbf{X})^{-1} \\ \hat{\boldsymbol{\beta}} &:= \Sigma \mathbf{X}^\top \boldsymbol{y} \ . \end{aligned}$$

Here $\hat{\boldsymbol{\beta}}$ is the maximum aposteriori estimate of $\boldsymbol{\beta}$ (the mode of $g(\boldsymbol{\beta}, \sigma^2 \,|\, \boldsymbol{y})$). As an aside we note that since $(\sigma^{-2} \,|\, \boldsymbol{y}) \sim \mathsf{Gamma}(n/2, \boldsymbol{y}^\top(\mathbf{I} - \mathbf{X}\Sigma\mathbf{X}^\top)\boldsymbol{y}/2)$, the posterior mean estimate of σ^{-2} is $\mathbb{E}[\sigma^{-2} \,|\, \boldsymbol{y}] = n/\boldsymbol{y}^\top(\mathbf{I} - \mathbf{X}\Sigma\mathbf{X}^\top)\boldsymbol{y} =: 1/\hat{\sigma}^2$, and we can show that

$$g(\boldsymbol{\beta} \,|\, \boldsymbol{y}) = \frac{\Gamma((n+p)/2)}{\Gamma(n/2)(n\pi)^{p/2}|\hat{\sigma}^2\Sigma|^{1/2}} \left[1 + \frac{(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^\top (\hat{\sigma}^2\Sigma)^{-1} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})}{n} \right]^{-(n+p)/2} .$$

In other words, $\boldsymbol{\beta} \,|\, \boldsymbol{y} \sim \mathsf{t}_n(\hat{\boldsymbol{\beta}}, \hat{\sigma}^2\Sigma)$ follows a multivariate student-t distribution with n degrees of freedom, location parameter $\hat{\boldsymbol{\beta}}$, and scale parameter (matrix) $\hat{\sigma}^2\Sigma$.

Answer:

Let

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}^\top & \Sigma_{22} \end{bmatrix}, \quad \mathbf{D} = \begin{bmatrix} \mathbf{D}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_{22} \end{bmatrix}, \quad \hat{\boldsymbol{\beta}} = \begin{bmatrix} \hat{\boldsymbol{\beta}}_1 \\ \hat{\boldsymbol{\beta}}_2 \end{bmatrix}$$

be the partitioning corresponding to $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$.

case of known σ^2. Using the fact that the marginal densities of a multivariate normal are also normal, we derive the Savage-Dickey density ratio:

$$\frac{g(\boldsymbol{\beta}_2 \,|\, \sigma^2, \boldsymbol{y})}{g(\boldsymbol{\beta}_2 \,|\, \sigma^2)} = \frac{|\mathbf{D}_{22}|^{1/2}}{|\Sigma_{22}|^{1/2}} \exp\left(-\frac{(\boldsymbol{\beta}_2 - \hat{\boldsymbol{\beta}}_2)^\top \Sigma_{22}^{-1} (\boldsymbol{\beta}_2 - \hat{\boldsymbol{\beta}}_2) - \boldsymbol{\beta}_2^\top \mathbf{D}_{22}^{-1} \boldsymbol{\beta}_2}{2\sigma^2} \right) .$$

Therefore,

$$B_{1\,|\,2} = \frac{|\mathbf{D}_{22}|^{1/2}}{|\Sigma_{22}|^{1/2}} \exp\left(-\frac{\hat{\boldsymbol{\beta}}_2^\top \Sigma_{22}^{-1} \hat{\boldsymbol{\beta}}_2}{2\sigma^2} \right) .$$

We can use this formula to compare different linear models with the aid of a computer. Note that one can show (using matrix inversion) that

$$\Sigma_{22}^{-1} = \mathbf{D}_{22} + \mathbf{X}_2^\top \mathbf{X}_2 - \mathbf{X}_2^\top \mathbf{X}_1 (\mathbf{X}_1^\top \mathbf{X}_1 + \mathbf{D}_{11}^{-1})^{-1} \mathbf{X}_1^\top \mathbf{X}_2 ,$$

but this formula is again only computable with a computer.

We can verify this formula with a direct computation of the Bayes factor, which requires that we compute the ratio of the marginal likelihoods of model one and model two. For model two, we have

$$g(\boldsymbol{y} \,|\, \sigma^2) = \int \frac{\mathrm{e}^{-\frac{(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})^\top \Sigma^{-1} (\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})}{2\sigma^2} - \frac{\boldsymbol{y}^\top (\mathbf{I}-\mathbf{X}\Sigma\mathbf{X}^\top)\boldsymbol{y}}{2\sigma^2}}}{(2\pi\sigma^2)^{(n+p)/2} |\mathbf{D}|^{1/2}} \mathrm{d}\boldsymbol{\beta} = \frac{|\Sigma|^{1/2} \mathrm{e}^{-\frac{\boldsymbol{y}^\top (\mathbf{I}-\mathbf{X}\Sigma\mathbf{X}^\top)\boldsymbol{y}}{2\sigma^2}}}{(2\pi\sigma^2)^{n/2} |\mathbf{D}|^{1/2}}$$

and we can obtain the marginal likelihood of model 1 by computing the limit:

$$g(\boldsymbol{y} \,|\, \sigma^2, m = 1) = \lim_{\substack{\mathbf{D}_{22} = \epsilon \mathbf{I} \\ \epsilon \to 0}} g(\boldsymbol{y} \,|\, \sigma^2) .$$

We can check (numerically in R/Matlab) that for small ϵ

$$B_{1\,|\,2} \approx \frac{g(\boldsymbol{y}, \mathbf{D}_{22} = \epsilon \mathbf{I} \,|\, \sigma^2)}{g(\boldsymbol{y} \,|\, \sigma^2)} .$$

case of unknown σ^2. Note that the prior $g(\boldsymbol{\beta}) = \int_0^\infty g(\boldsymbol{\beta} \,|\, \sigma^2) g(\sigma^2) \mathrm{d}\sigma^2$ is the pdf of the $\mathsf{t}_0(\mathbf{0}, \mathbf{D})$ distribution, which is not a proper density and so we cannot apply the Savage-Dickey ratio formula. Instead, we must take the ratio of the marginal likelihoods for model 1 and model 2:

$$B_{1\,|\,2} = \frac{g(\boldsymbol{y}\,|\,m=1)}{g(\boldsymbol{y})}$$

For model 2 we have

$$g(\boldsymbol{y}) = \int_0^\infty \frac{|\Sigma|^{1/2}\mathrm{e}^{-\frac{\boldsymbol{y}^\top(\mathbf{I}-\mathbf{X}\Sigma\mathbf{X}^\top)\boldsymbol{y}}{2\sigma^2}}}{(2\pi)^{n/2}(\sigma^2)^{n/2+1}|\mathbf{D}|^{1/2}}\mathrm{d}\sigma^2 = \frac{|\Sigma|^{1/2}\Gamma(n/2)}{|\mathbf{D}|^{1/2}(\pi n\hat{\sigma}^2)^{n/2}}$$

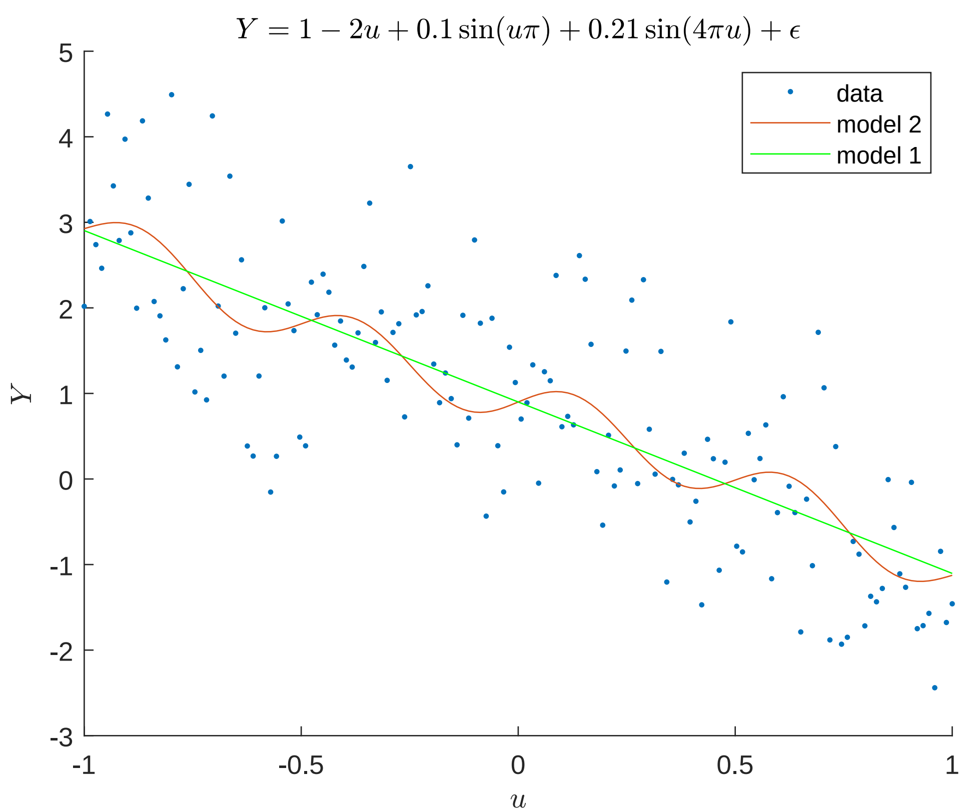
and for model 1:

$$g(\boldsymbol{y}\,|\,m=1) = \frac{|\Sigma_1|^{1/2}\Gamma(n/2)}{|\mathbf{D}_{11}|^{1/2}(\pi n\hat{\sigma}_1^2)^{n/2}},$$

where

$$\hat{\sigma}_1^2 := \frac{\boldsymbol{y}^\top(\mathbf{I}-\mathbf{X}_1\Sigma_1\mathbf{X}_1^\top)\boldsymbol{y}}{n}, \qquad \Sigma_1 := (\mathbf{D}_{11}^{-1} + \mathbf{X}_1^\top\mathbf{X}_1)^{-1}\,.$$

The following figure gives an example of a case where the Bayes factors are close to unity ($B_{1\,|\,2} = 0.9759$ if we know σ and $B_{1\,|\,2} = 0.6145$ if unknown) and it is difficult to decide which model is preferable.



To produce the picture we used the following code.

```
clear all,clc,randn('seed',10),rand('seed',10),clf
n=150;a=-1;b=1;
u=[a:(b-a)/(n-1):b]';
X=[ones(n,1),u,sin(u*pi),sin(u*4*pi)];
beta=[1,-2,.1,.21]';
p=length(beta);
sig=1;
y=X*beta+sig*randn(n,1); % generate data

D=eye(p);Sig=inv(inv(D)+X'*X);
bhat=Sig*X'*y;
% case of known sigma^2
Sig22=Sig(3:4,3:4);b2=bhat(3:4);
% Savage-Dickey ratio
B12=exp(-.5*b2'*inv(Sig22)*b2/sig^2)/sqrt(det(Sig22))
% direct computation as ratio of marginal likelihoods
exp(marlik(D(1:2,1:2),y,X(:,1:2),sig^2)-marlik(D,y,X,sig^2))
% unknown sigma^2, only direct computation possible:
exp(marlik(D(1:2,1:2),y,X(:,1:2))-marlik(D,y,X))

plot(u,y,'.'), box off, hold on
%model 2
plot(u,X*bhat)
% model 1
X1=X(:,1:2);
Sig1=inv(inv(D(1:2,1:2))+X1'*X1);
bhat1=Sig1*X1'*y;
plot(u,X1*bhat1,'g ')
```

The code uses a function that computes the logarithm of the marginal likelihood.

```
function lnml=marlik(D,y,X,sig2)
% outputs log of marginal likelihood
[n,p]=size(X);
Sig=inv(inv(D)+X'*X);
sig2hat=y'*(eye(n)-X*Sig*X')*y/n;

if nargin>3 % if sigma is known
 lnml=-.5*sig2hat*n/sig2+.5*log(det(Sig))-.5*log(det(D))-...
      n/2*log(2*pi*sig2);
else % if sigma is unknown
    lnml=gammaln(n/2)+.5*log(det(Sig))-.5*log(det(D))-...
      n/2*log(2*pi*sig2hat*n/2);
end
```