

MATH 598: TOPICS IN STATISTICS

MODEL SELECTION VIA REVERSIBLE JUMP MCMC

Consider two competing pharmacokinetic (PK) models for response data $y(t)$ measured at different time points t_1, \dots, t_n :

1. **Model M_1** : One compartment, elimination only;

$$\mathbb{E}[Y(t)] = A_1 \exp\{-\lambda_1 t\} \quad t \geq 0$$

2. **Model M_2** : One compartment, absorption and elimination;

$$\mathbb{E}[Y(t)] = A_2 (\exp\{-\lambda_{21} t\} - \exp\{-(\lambda_{21} + \lambda_{22}) t\}) \quad t \geq 0$$

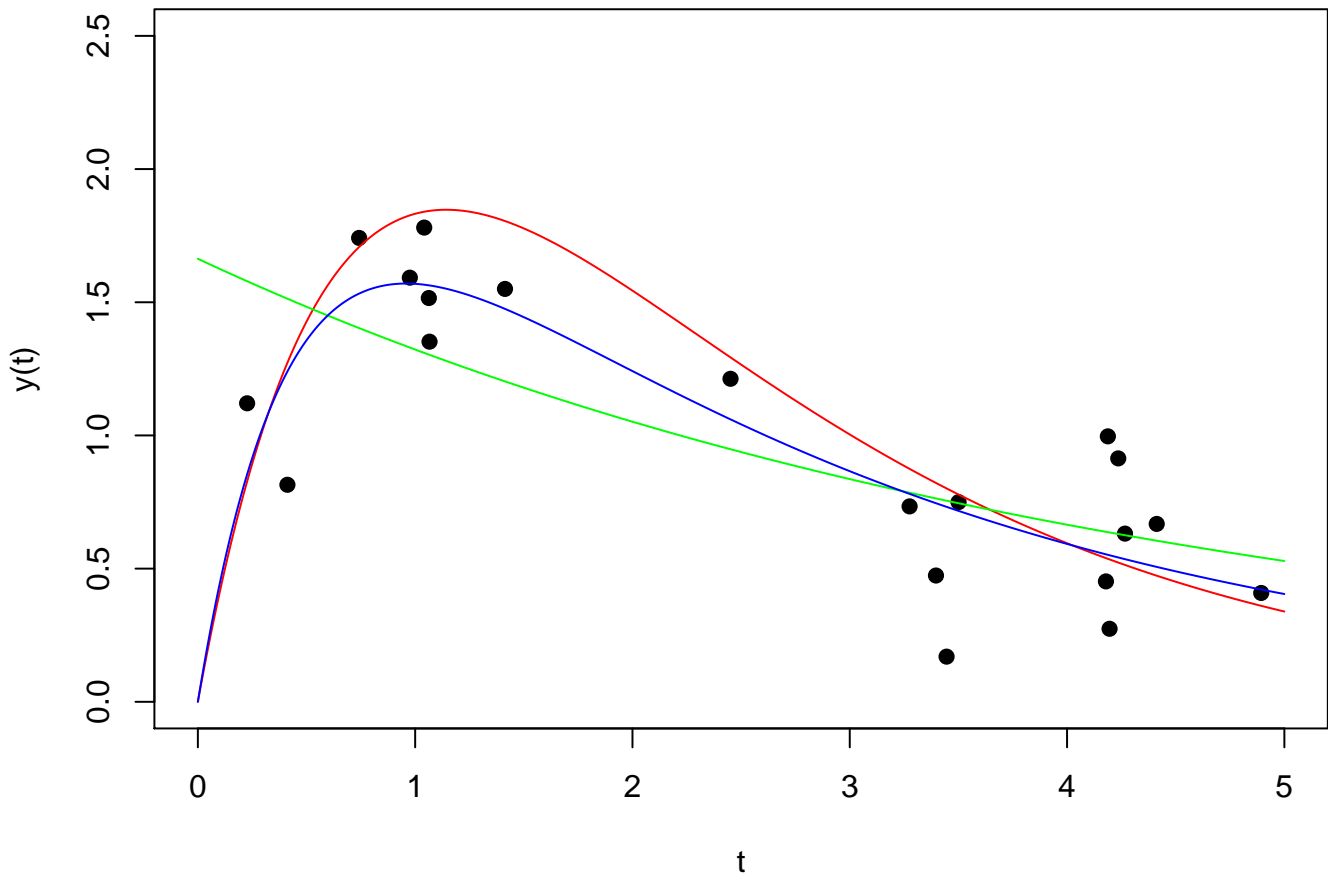
where (A_1, λ_1) and $(A_2, \lambda_{21}, \lambda_{22})$ are positive parameters. Under an assumption of additive, heteroscedastic Normal errors, we have two competing explanations for the observed data; both models can be fitted using ordinary least-squares.

In the following example, we simulate data from Model M_2 , with parameters

$$\theta_{21} = \log A_2 = 2 \quad \theta_{22} = \log \lambda_{21} = -0.5 \quad \theta_{23} = \log \lambda_{22} = -0.5$$

with noise variance $\sigma^2 = 0.3^2$.

```
library(MASS)
library(mvnfast)
#Simulate data
set.seed(4263374)
n<-20
x<-sort(runif(n,0,5))
th<-exp(c(2,-0.5,-0.5))
y<-funct.m2(x,th)+rnorm(n,0,.3)
par(mar=c(4,4,2,0))
plot(x,y,xlim=range(0,5),ylim=range(0,2.5),pch=19,xlab="t",ylab="y(t)")
xval<-seq(0,5,by=0.01)
yval<-funct.m2(xval,th)
lines(xval,yval,col="red")
fit.1<-optim(c(0,0,0),fn=loglike.optim,yv=y,xv=x,
            fv=funct.m1,np=3,control=list(maxit=1000),hessian=T)
yval1<-funct.m1(xval,exp(fit.1$par[1:2]))
lines(xval,yval1,col="green")
fit.2<-optim(c(0,0,0,0),fn=loglike.optim,yv=y,xv=x,
            fv=funct.m2,np=4,control=list(maxit=10000),hessian=T)
yval2<-funct.m2(xval,exp(fit.2$par[1:3]))
lines(xval,yval2,col="blue")
```



We can compare the two models by using the BIC: for $k = 1, 2$, we have

$$\text{BIC}_k = -2 \log \mathcal{L}_n^{M_k}(\hat{\theta}_k) + d_k \log n$$

where d_k is the total number of parameters for model M_k .

```
#Compute the BIC
BIC1<-2*fit.1$value+3*log(n)
BIC2<-2*fit.2$value+4*log(n)
c(BIC1,BIC2)

+ [1] 21.29536 13.59642

#Check using nls
PK.data<-data.frame(x,y)
AIC(nls(y~A*exp(-B*x),data=PK.data,start=list(A=1,B=0.5)),k=log(20))

+ [1] 21.29536

AIC(nls(y~A*exp(-B*x)*(1-exp(-C*x)),data=PK.data,start=list(A=1,B=1,C=1)),k=log(20))

+ [1] 13.59642
```

Under BIC, the evidence favours model M_2 , for which $\text{BIC}_2 = 13.596417$, over model M_1 , where $\text{BIC}_1 = 21.295358$.

We now consider a fully Bayesian solution using reversible jump MCMC. In the log-scale parameterization

$$\theta_1 = (\log A_1, \log \lambda_1) \quad \theta_2 = (\log A_2, \log \lambda_{21}, \log \lambda_{22}).$$

We place equal prior probabilities on M_1 and M_2 , and then place independent $N(0, \tau^2)$ priors on the components of θ_1 and θ_2 . The prior on the residual error variance σ^2 is Inverse Gamma with parameters 20 and 8.

The ML estimates $\hat{\theta}_1$ and $\hat{\theta}_2$ can be computed easily, as can the Hessian matrices $\hat{\mathbf{I}}_1$ and $\hat{\mathbf{I}}_2$; these likelihood-based results yield reasonable approximations to the posterior distributions to produce independence MH algorithms. Specifically, at the ML estimates for σ under the two models, we may approximate the conditional posterior for θ by the Normal density

$$\pi_n^{M_k}(\theta_k | \hat{\sigma}, \mathbf{y}) \simeq \text{Normal}(\hat{\theta}_k, n^{-1} \hat{\sigma}_k^2 \hat{\mathbf{I}}_k^{-1}) \quad (1)$$

We may also introduce the prior information and use the conjugate analysis to return an approximate Normal conditional posterior distribution. Finally, on fitting using ML, the estimates of σ under the two models are found to be quite similar ($M_1 : \hat{\sigma}_1 = 0.252$, $M_2 : \hat{\sigma}_2 = 0.329$).

Reversible Jump MCMC: A reversible jump MCMC algorithm can be constructed as follows: we again consider four move types:

- 1 $m = 1$: move **within** M_1 ; update θ_1 from $\pi_n^{M_1}(\theta_1 | M_1, \sigma)$
- 2 $m = 2$: move **within** M_1 ; update θ_2 from $\pi_n^{M_2}(\theta_2 | M_2, \sigma)$
- 3 $m = 3$: move **from** M_1 **to** M_2 ; propose a new θ_2 , and carry out an accept/reject step.
- 4 $m = 4$: move **from** Model M_2 **to** Model M_1 ; propose a new θ_1 , and carry out an accept/reject step.

with the remaining parameter σ^2 being updated in a Gibbs sampler algorithm at each iteration.

- For moves $m = 1, 2$, we use standard Metropolis-Hastings steps, either jointly on the whole parameter vector, or for each of the parameters individually. The asymptotic approximations in (1) can be used.
- Moves $m = 3, 4$ are a forward/reverse move pair. For move 3, several options are available; for example, we could adopt the earlier strategy, and generate a new variate u from the prior for the additional parameter, and then merely use the mapping

$$(\theta_{11}, \theta_{12}, u) \mapsto (\theta_{21} = \theta_{11}, \theta_{22} = \theta_{12}, \theta_{23} = u)$$

with reverse move setting $\theta_{23} = 0$.

This approach may be adequate, but more probably would not facilitate good mixing across the models. A better strategy is to consider a different augmentation, where we generate $\mathbf{u} = (u_1, u_2, u_3)$ from the model in (1) for $k = 2$, and map $(\theta_{11}, \theta_{12}, u_1, u_2, u_3)$ to

$$(\theta_{21} = u_1, \theta_{22} = u_2, \theta_{23} = u_3, v_1 = \theta_{11}, v_2 = \theta_{12})$$

with the paired reverse move being to generate $\mathbf{v} = (v_1, v_2)$ from the model in (1) for $k = 1$.

This guarantees that the proposed value θ_2 lies in a region with reasonably high posterior support under model M_2 , although it does not guarantee that the move will be accepted with high probability.

In the Hastings ratio, the Jacobian of the transformation is 1, and under equal probabilities of forward/reverse moves, we have that

$$\frac{\pi_n(M_2, \theta_2) p_V(v_1, v_2)}{\pi_n(M_1, \theta_1) p_U(u_1, u_2, u_3)}$$

can be written

$$\frac{\mathcal{L}_n^{M_2}(\theta_2) \left\{ \prod_{j=1}^3 \phi(\theta_{2j}/\tau)/\tau \right\} \phi_2(\theta_{11}, \theta_{12}; \hat{\theta}_1, \hat{\mathbf{I}}_1)}{\mathcal{L}_n^{M_1}(\theta_1) \left\{ \prod_{j=1}^2 \phi(\theta_{1j}/\tau)/\tau \right\} \phi_3(\theta_{21}, \theta_{22}, \theta_{23}; \hat{\theta}_2, \hat{\mathbf{I}}_2)}$$

where τ is the prior standard deviation for the θ parameters. The logic of this construction is that numerically

$$\mathcal{L}_n^{M_1}(\theta_1) \simeq \text{Normal}_2(\hat{\theta}_1, \hat{\mathbf{I}}_1^{-1}) \quad \mathcal{L}_n^{M_2}(\theta_2) \simeq \text{Normal}_3(\hat{\theta}_2, \hat{\mathbf{I}}_2^{-1}).$$

Note that if the prior information is also incorporated, the asymptotic approximation to the posterior may be improved further.

The algorithm was run for 100000 iterations to collect 10000 posterior samples.

```

#Prior
tau<-4
prior.al<-20
prior.be<-8

#Normal approximations to the conditional posterior distributions
var.1 <-solve(fit.1$hessian[1:2,1:2]+diag(1/tau^2,2))
ests.1<-var.1 %*% (fit.1$hessian[1:2,1:2] %*% fit.1$par[1:2])
sig.1<-exp(fit.1$par[3])
prec.1<-solve(var.1)

var.2<-solve(fit.2$hessian[1:3,1:3]+diag(1/tau^2,3))
ests.2<-var.2 %*% (fit.2$hessian[1:3,1:3] %*% fit.2$par[1:3])
sig.2<-exp(fit.2$par[4])
prec.2<-solve(var.2)

#Run MCMC
nsamp<-10000
nburn<-1000
nthin<-1
imod.tot<-c(0,0)
nr<-20
for(irep in 1:nr){
  ifix<-F
  imod<-1
  Res<-run.RJMCMC(nburn,nsamp,nthin)
  imod.tot<-imod.tot+table(Res$model)
  print(c(irep,as.numeric(table(Res$model)/nsamp)))
}

+ [1] 1.0000 0.6676 0.3324
+ [1] 2.00 0.65 0.35
+ [1] 3.0000 0.6581 0.3419
+ [1] 4.0000 0.6406 0.3594
+ [1] 5.0000 0.6609 0.3391
+ [1] 6.0000 0.6844 0.3156
+ [1] 7.0000 0.6858 0.3142
+ [1] 8.0000 0.6719 0.3281
+ [1] 9.0000 0.6405 0.3595
+ [1] 10.0000 0.6727 0.3273
+ [1] 11.0000 0.6706 0.3294
+ [1] 12.0000 0.6704 0.3296
+ [1] 13.0000 0.6678 0.3322
+ [1] 14.0000 0.6391 0.3609
+ [1] 15.0000 0.6733 0.3267
+ [1] 16.0000 0.6531 0.3469
+ [1] 17.0000 0.6539 0.3461
+ [1] 18.0000 0.6647 0.3353
+ [1] 19.0000 0.6688 0.3312
+ [1] 20.0000 0.6558 0.3442

#Overall estimate of model probabilities
imod.tot/(nr*nsamp)

+
+ 1 2
+ 0.6625 0.3375

```

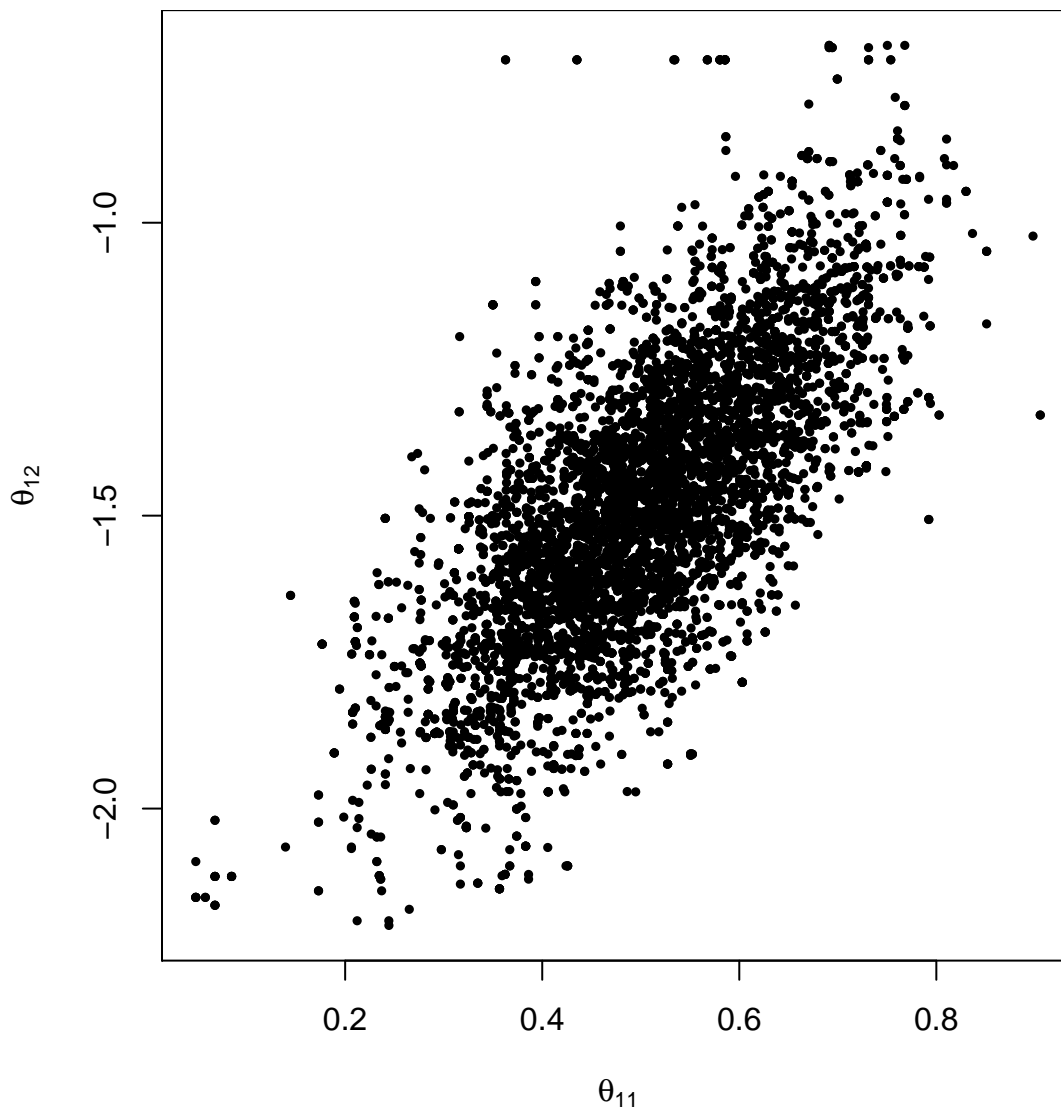
In this run with $\tau = 4$, the chain spent about 66 % of the time in model M_1 , indicating the posterior probabilities are

$$\pi_n(M_1) \simeq 0.66 \quad \pi_n(M_2) \simeq 0.34.$$

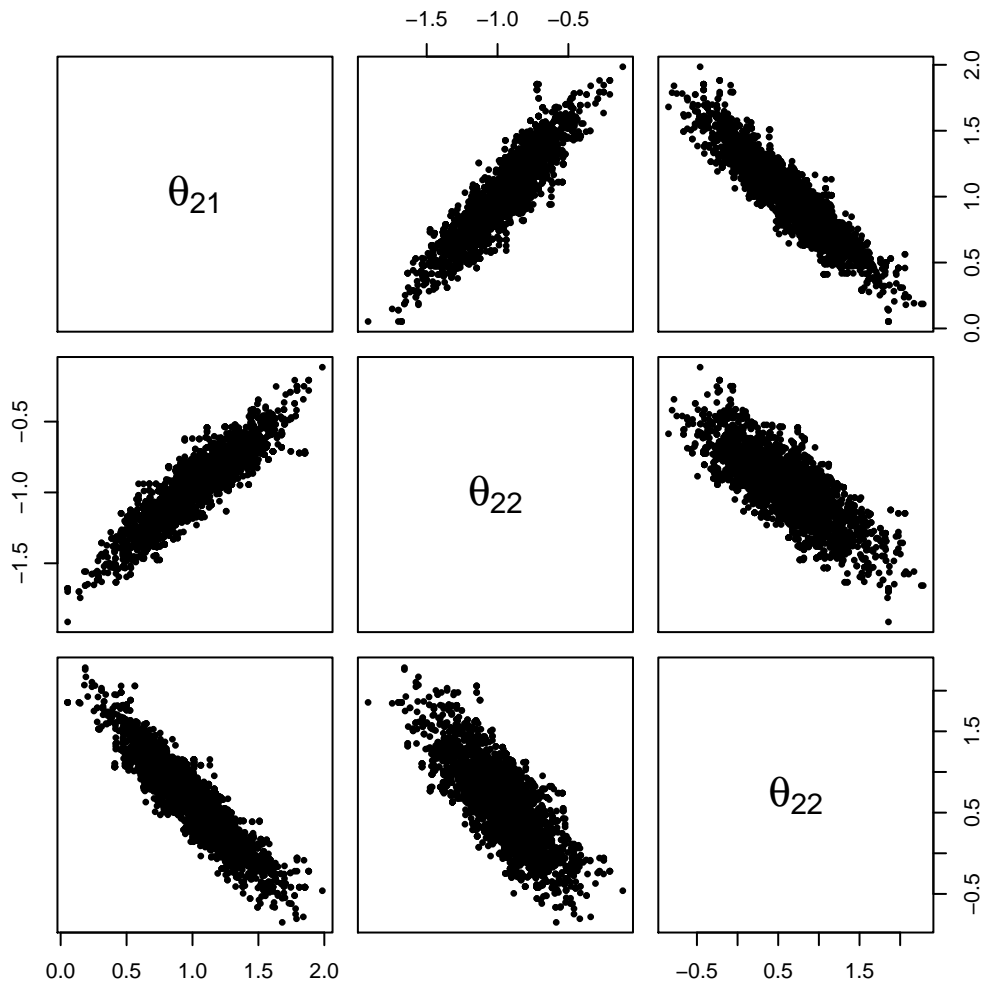
The model posterior probabilities vary with the choice of τ ; this is as expected, as the model probabilities are

closely related to the marginal likelihood, or prior predictive distribution, which is the expected value of the likelihood for the observed data with respect to the prior distribution. It is evident from the discussion that the prior specification acts as a penalty for complexity. For illustration, if $\tau = 1$, the model probabilities change to (0.43, 0.57); if $\tau = 10$, the model probabilities are approximately (0.80, 0.20).

```
th1<-Res$th[Res$model==1,]  
th2<-Res$th[Res$model==2,]  
  
par(mar=c(4,4,0,1))  
plot(th1,pch=19,cex=0.5,xlab=expression(theta[11]),ylab=expression(theta[12]))
```

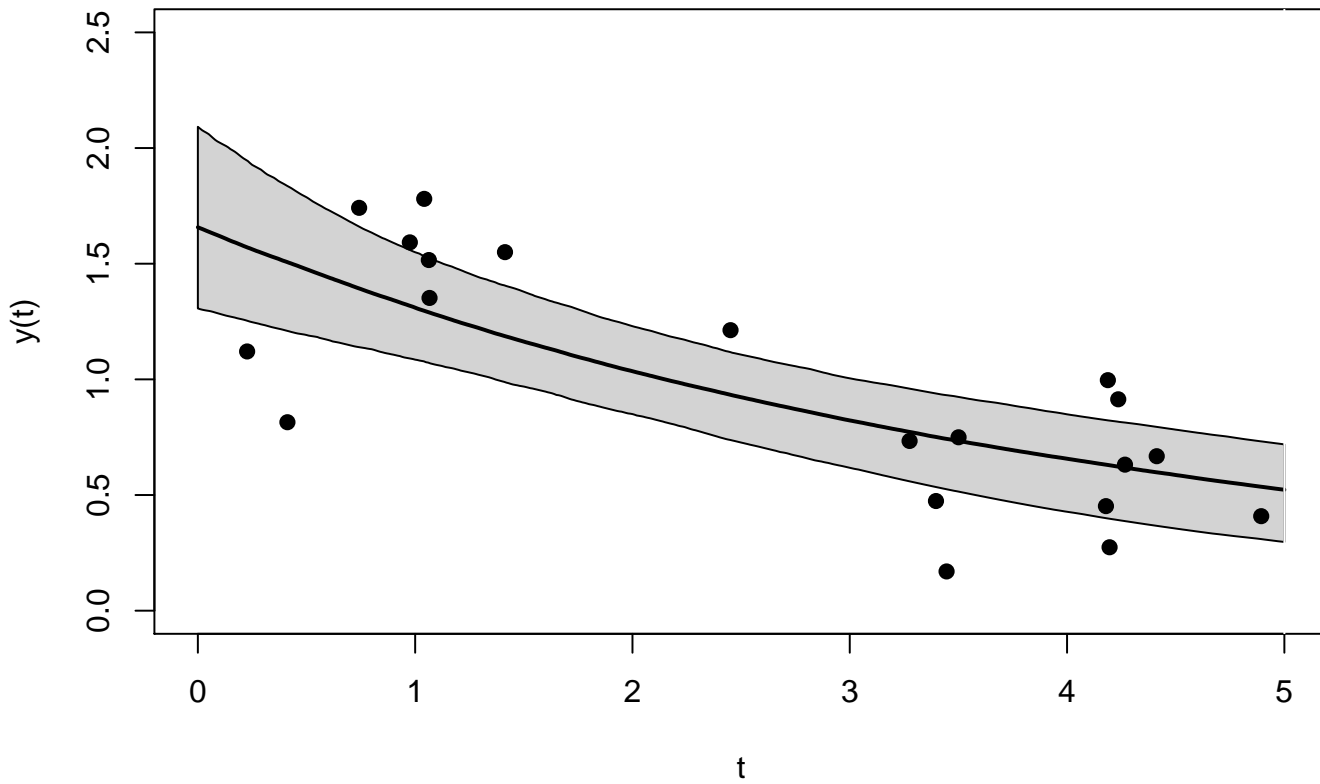


```
pairs(th2,pch=19,cex=0.5,labels=c(expression(theta[21]),expression(theta[22]),expression(theta[22])))
```



Conditional on M_1 or M_2 being true, we can perform inference about the parameters of the two models, and also reconstruct estimates and posterior credible intervals for $E[Y]$.

Model 1



Model 2

