

Exercise 10

107.330 - Statistical Simulation and Computerintensive Methods, WS24

11912007 - Yahya Jabary

10.10.2024

```
set.seed(11912007)
```

Assignment

Simulate $n = 500$ observations from from a mixture of two normal linear regression models, with the mixture weights depending on the predictor, with different error variances and with a non-linear mean function for the second component, $y_i|x_i \sim^{ind.} e^{-2x_i} \cdot N(y_i|x_i, 0.01) + (1 - e^{-2x_i})N(y_i|x_i^4, 0.04), i = 1, \dots, n$, where the predictor values x_i are simulated from a uniform distribution, $x_i \sim^{iid} U(0, 1)$.

Apply the DPcdensity function from package DPpackage, was fitted using the following hyper-parameters: $a0 = 10, b0 = 1, \nu1 = \nu2 = 4, m2 = (\hat{y}, \hat{x})', \tau1 = 6.01, \tau2 = 3.01, S2 = \Psi - 12$, where S is the sample covariance matrix for the response and predictor. A total number of 25,000 scans of the Markov chain cycle implemented in the DPcdensity function were completed. A burn-in period of 5,000 samples was considered and the chain was subsampled every 4 iterates to get a final sample size of 5,000. The following commands were used to fit the model, where the conditional density estimates were evaluated on a grid of 100 points on the range of the response.

Provide the traceplot and density plots for

- the true conditional densities of $y | x$ along with
- the posterior mean estimates and
- point-wise 95% HPD intervals for several values of x, such as: (a) $x = 0.1$, (b) $x = 0.25$, (c) $x = 0.5$, (d) $x = 0.75$.

```
# simulate data
n <- 500
x <- runif(n)
y1 <- x + rnorm(n, 0, sqrt(0.01))
y2 <- x^4 + rnorm(n, 0, sqrt(0.04))
u <- runif(n)
prob <- exp(-2*x)
y <- ifelse(u < prob, y1, y2)

# set up prior and MCMC parameters / hyperparameters
w <- cbind(y,x)
wbar <- apply(w, 2, mean)
wcov <- var(w)

prior <- list(a0=10,
              b0=1,
              nu1=4,
              nu2=4,
              s2=0.5*wcov,
              m2=wbar,
              psiinv2=2*solve(wcov),
              tau1=6.01,
              tau2=2.01)

mcmc <- list(nburn=5000,
             nsave=5000,
             nskip=3,
             ndisplay=100)

# fit the model
```

```
xpred <- c(0.1, 0.25, 0.5, 0.75)
fit <- DPcdensity(y=y, x=x, xpred=xpred, ngrid=100,
  prior=prior, mcmc=mcmc,
  state=NULL, status=TRUE,
  compute.band=TRUE, type.band="HPD")
```

```
##
## MCMC scan 100 of 5000 (CPU time: 11.539 s)
## MCMC scan 200 of 5000 (CPU time: 15.501 s)
## MCMC scan 300 of 5000 (CPU time: 19.286 s)
## MCMC scan 400 of 5000 (CPU time: 22.867 s)
## MCMC scan 500 of 5000 (CPU time: 26.789 s)
## MCMC scan 600 of 5000 (CPU time: 30.547 s)
## MCMC scan 700 of 5000 (CPU time: 34.075 s)
## MCMC scan 800 of 5000 (CPU time: 37.769 s)
## MCMC scan 900 of 5000 (CPU time: 41.512 s)
## MCMC scan 1000 of 5000 (CPU time: 45.238 s)
## MCMC scan 1100 of 5000 (CPU time: 49.067 s)
## MCMC scan 1200 of 5000 (CPU time: 52.855 s)
## MCMC scan 1300 of 5000 (CPU time: 56.526 s)
## MCMC scan 1400 of 5000 (CPU time: 60.176 s)
## MCMC scan 1500 of 5000 (CPU time: 64.018 s)
## MCMC scan 1600 of 5000 (CPU time: 67.684 s)
## MCMC scan 1700 of 5000 (CPU time: 71.506 s)
## MCMC scan 1800 of 5000 (CPU time: 75.296 s)
## MCMC scan 1900 of 5000 (CPU time: 78.887 s)
## MCMC scan 2000 of 5000 (CPU time: 82.607 s)
## MCMC scan 2100 of 5000 (CPU time: 86.318 s)
## MCMC scan 2200 of 5000 (CPU time: 89.950 s)
## MCMC scan 2300 of 5000 (CPU time: 93.668 s)
## MCMC scan 2400 of 5000 (CPU time: 97.334 s)
## MCMC scan 2500 of 5000 (CPU time: 101.049 s)
## MCMC scan 2600 of 5000 (CPU time: 104.745 s)
## MCMC scan 2700 of 5000 (CPU time: 108.346 s)
## MCMC scan 2800 of 5000 (CPU time: 112.009 s)
## MCMC scan 2900 of 5000 (CPU time: 115.633 s)
## MCMC scan 3000 of 5000 (CPU time: 119.260 s)
## MCMC scan 3100 of 5000 (CPU time: 123.028 s)
## MCMC scan 3200 of 5000 (CPU time: 126.670 s)
## MCMC scan 3300 of 5000 (CPU time: 130.324 s)
## MCMC scan 3400 of 5000 (CPU time: 134.070 s)
## MCMC scan 3500 of 5000 (CPU time: 137.678 s)
## MCMC scan 3600 of 5000 (CPU time: 141.288 s)
## MCMC scan 3700 of 5000 (CPU time: 145.053 s)
## MCMC scan 3800 of 5000 (CPU time: 148.650 s)
## MCMC scan 3900 of 5000 (CPU time: 152.109 s)
## MCMC scan 4000 of 5000 (CPU time: 155.898 s)
## MCMC scan 4100 of 5000 (CPU time: 159.842 s)
## MCMC scan 4200 of 5000 (CPU time: 163.520 s)
## MCMC scan 4300 of 5000 (CPU time: 167.101 s)
## MCMC scan 4400 of 5000 (CPU time: 170.628 s)
## MCMC scan 4500 of 5000 (CPU time: 174.397 s)
## MCMC scan 4600 of 5000 (CPU time: 177.974 s)
## MCMC scan 4700 of 5000 (CPU time: 181.585 s)
## MCMC scan 4800 of 5000 (CPU time: 185.404 s)
## MCMC scan 4900 of 5000 (CPU time: 189.099 s)
## MCMC scan 5000 of 5000 (CPU time: 192.781 s)
```

```
#
# plot
#

# true conditional density function
dtrue <- function(grid, x) {
```

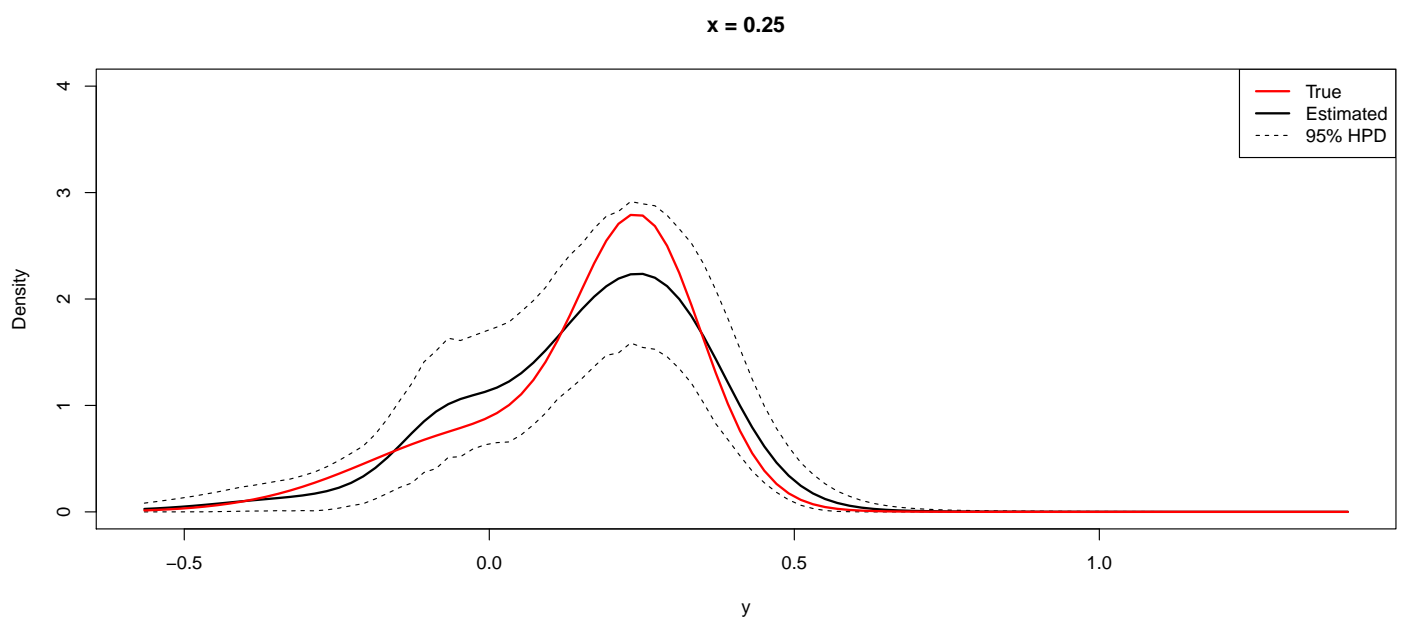
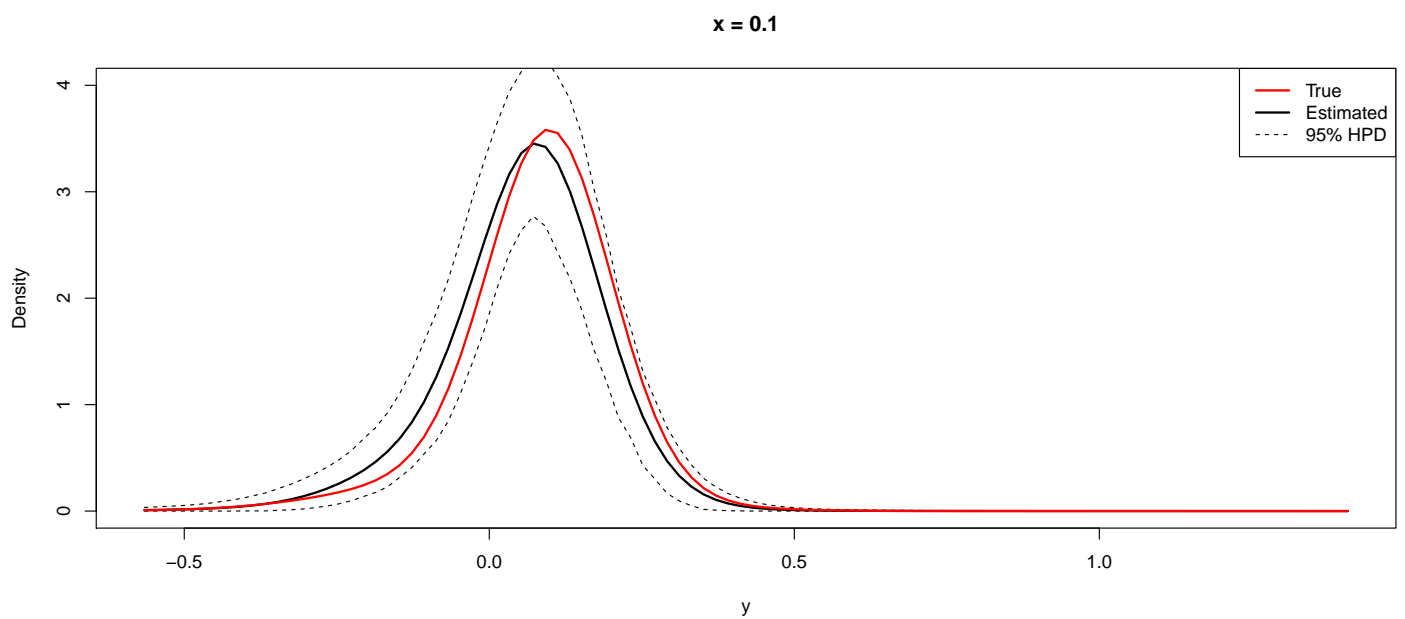
```

exp(-2*x)*dnorm(grid, mean=x, sd=sqrt(0.01)) +
  (1-exp(-2*x))*dnorm(grid, mean=x^4, sd=sqrt(0.04))
}

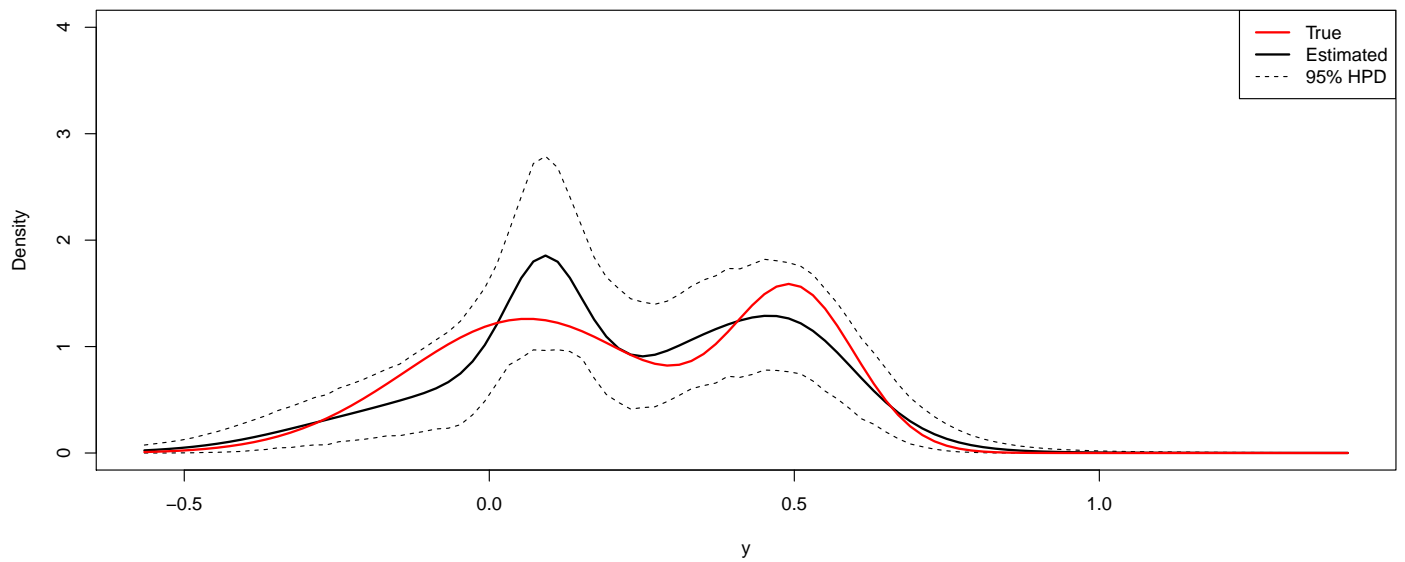
plot_conditional_density <- function(fit, x_val, x_index) {
  par(mfrow=c(1,1))
  plot(fit$grid, fit$densp.h[x_index,], lwd=1, type="l", lty=2,
       main=paste("x =", x_val), xlab="y", ylab="Density", ylim=c(0,4))
  lines(fit$grid, fit$densp.l[x_index,], lwd=1, type="l", lty=2)
  lines(fit$grid, fit$densp.m[x_index,], lwd=2, type="l", lty=1)
  lines(fit$grid, dtrue(fit$grid, x_val), lwd=2, type="l", lty=1, col="red")
  legend("topright", legend=c("True", "Estimated", "95% HPD"),
        col=c("red", "black", "black"), lty=c(1,1,2), lwd=c(2,2,1))
}

for (i in 1:length(xpred)) {
  plot_conditional_density(fit, xpred[i], i)
}

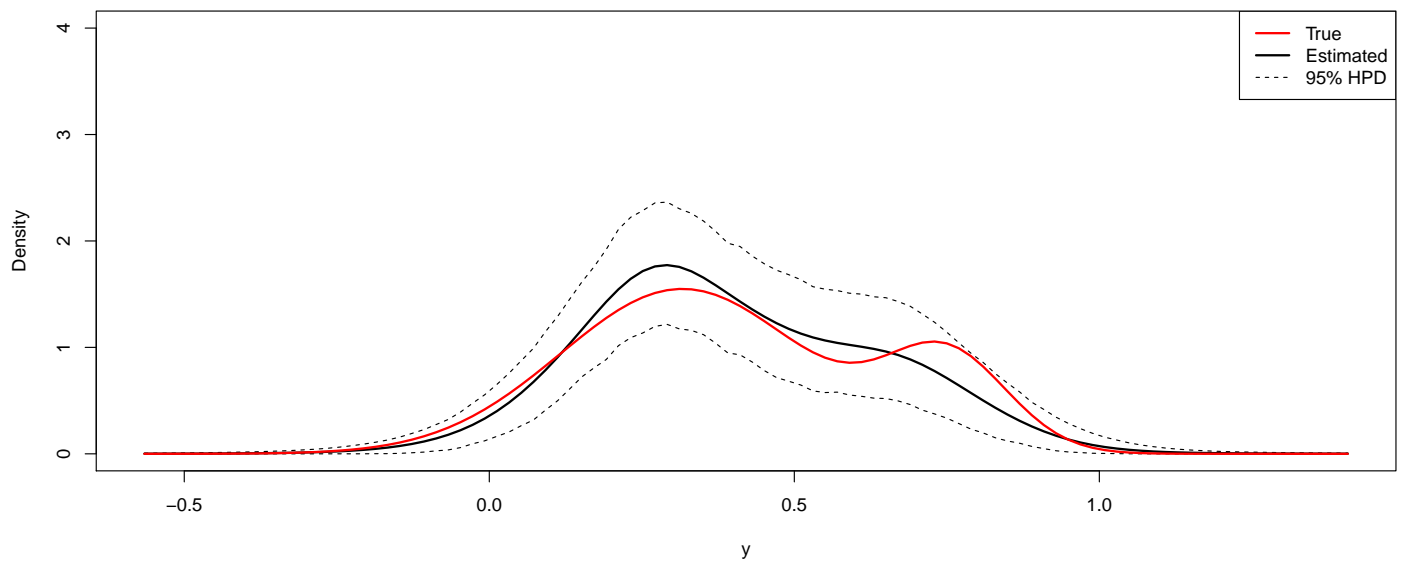
```



x = 0.5



x = 0.75



```
# plot traceplots and density plots
mcmc_samples <- as.mcmc(fit$save.state$thetasave)
# plot(mcmc_samples)
mcmc_subset <- as.mcmc(fit$save.state$thetasave[, 1:4]) # only plot the first 4 parameters
par(mfrow=c(2,2), mar=c(4,4,2,1))
plot(mcmc_subset)
```

