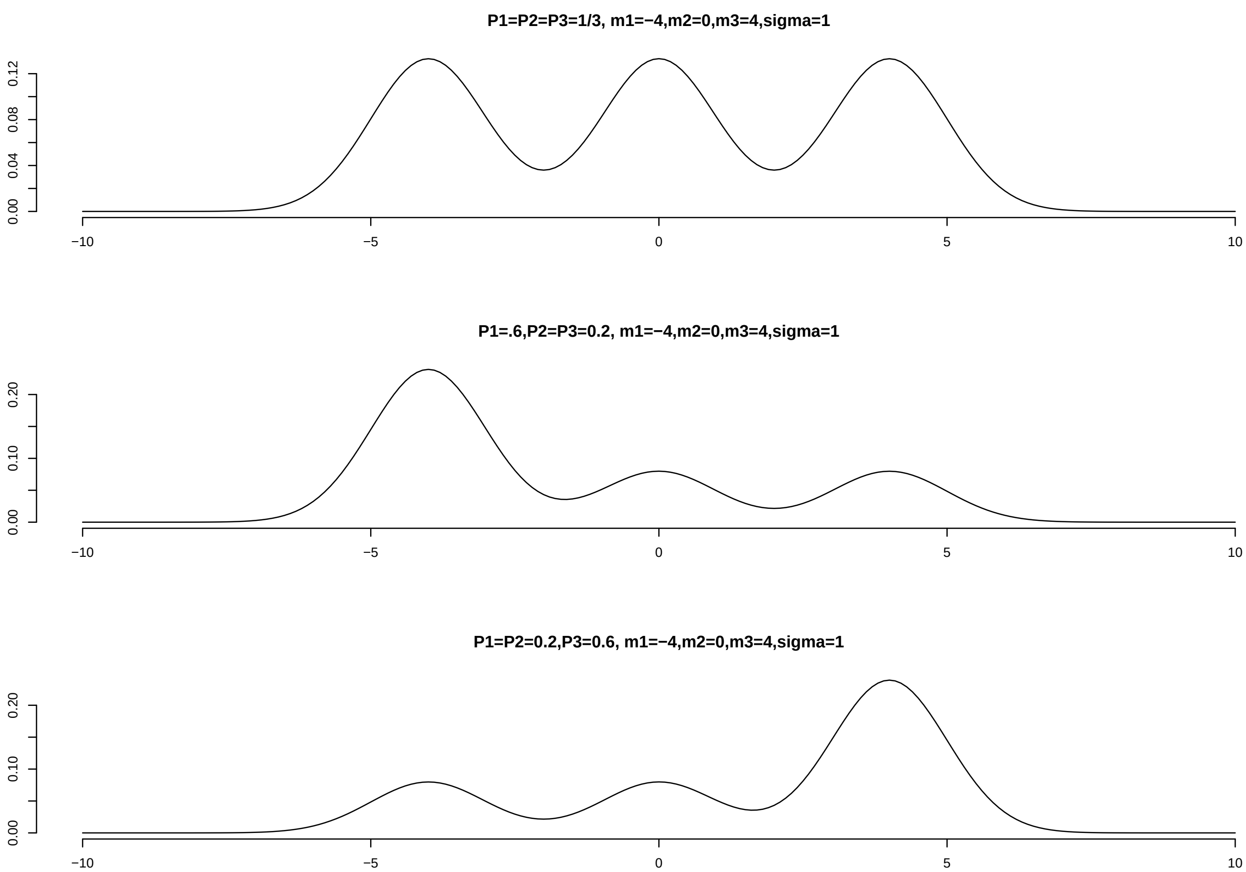
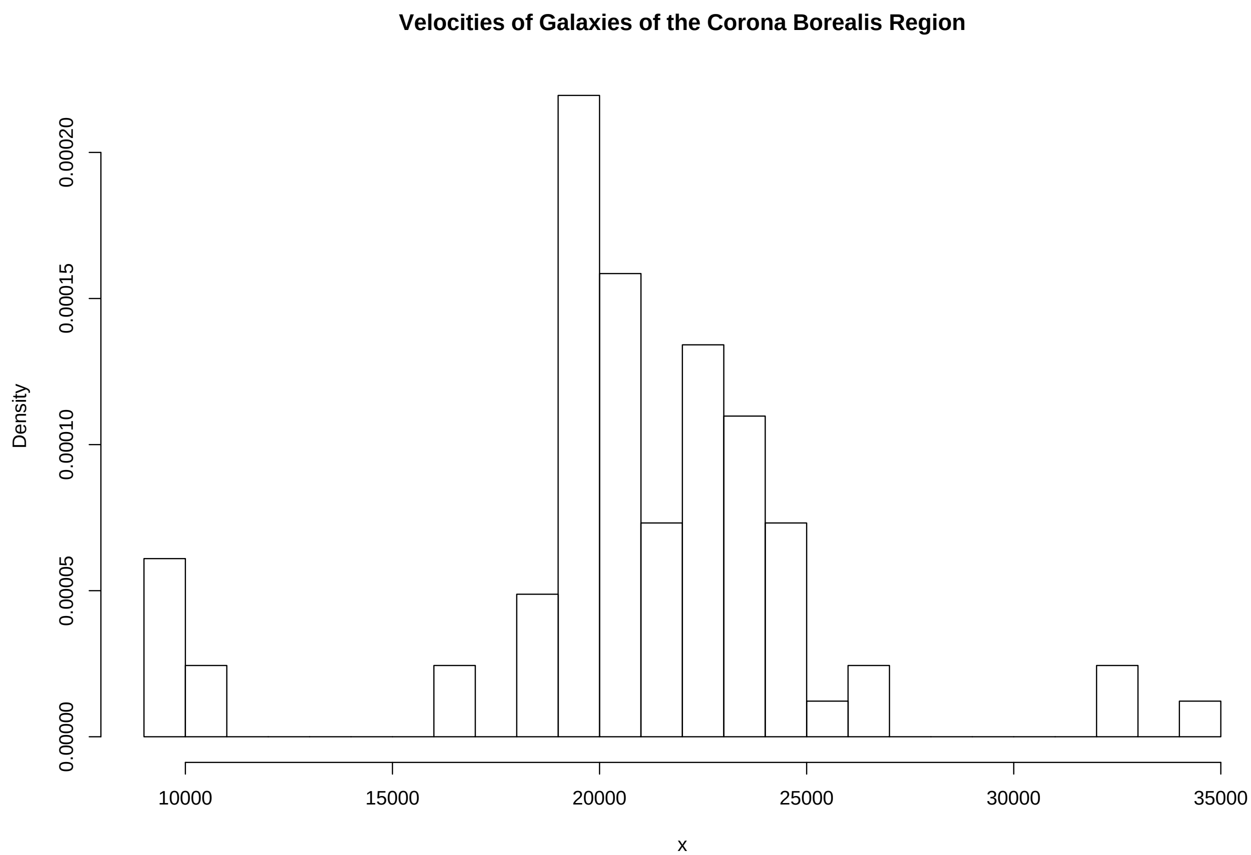
Examples for a 3 component mixture of normals

The picture shows the pdfs for a mixture of normals with 3 components.

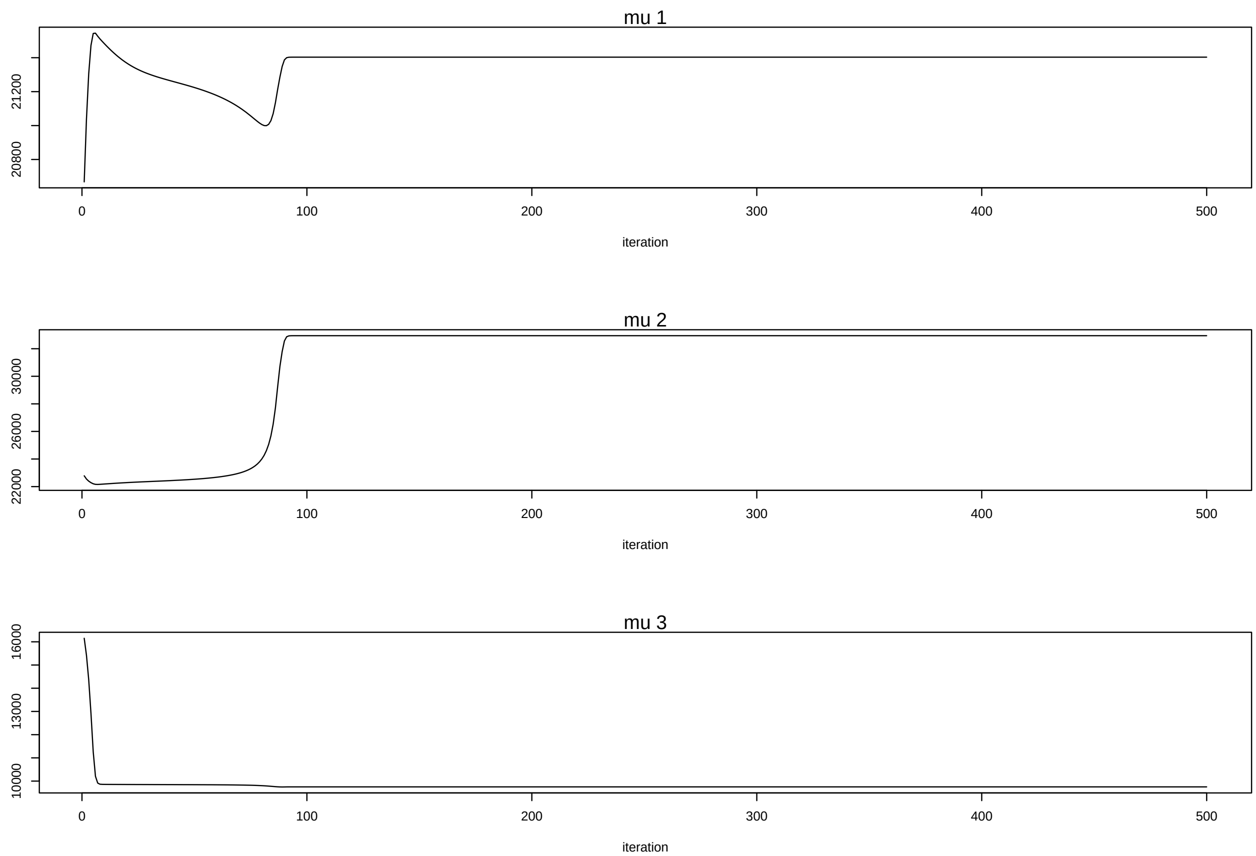


EM-algorithm example. Fitting a mixture of normals to Galaxy Data

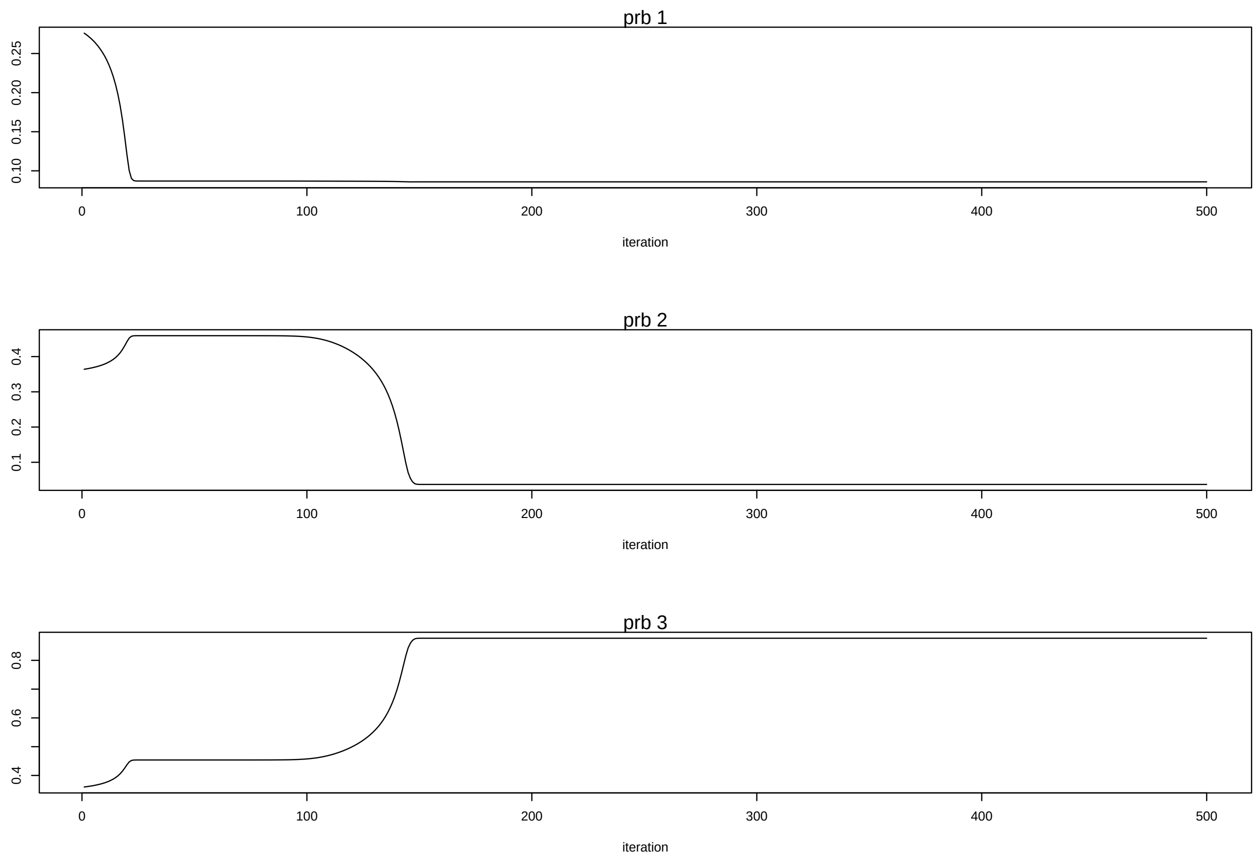
- Picture shows a histogram with 30 bins of 82 Velocities of Galaxies for an Unfilled Survey of the Corona Borealis region. Velocity (km per second) (Tanner's book page 88).



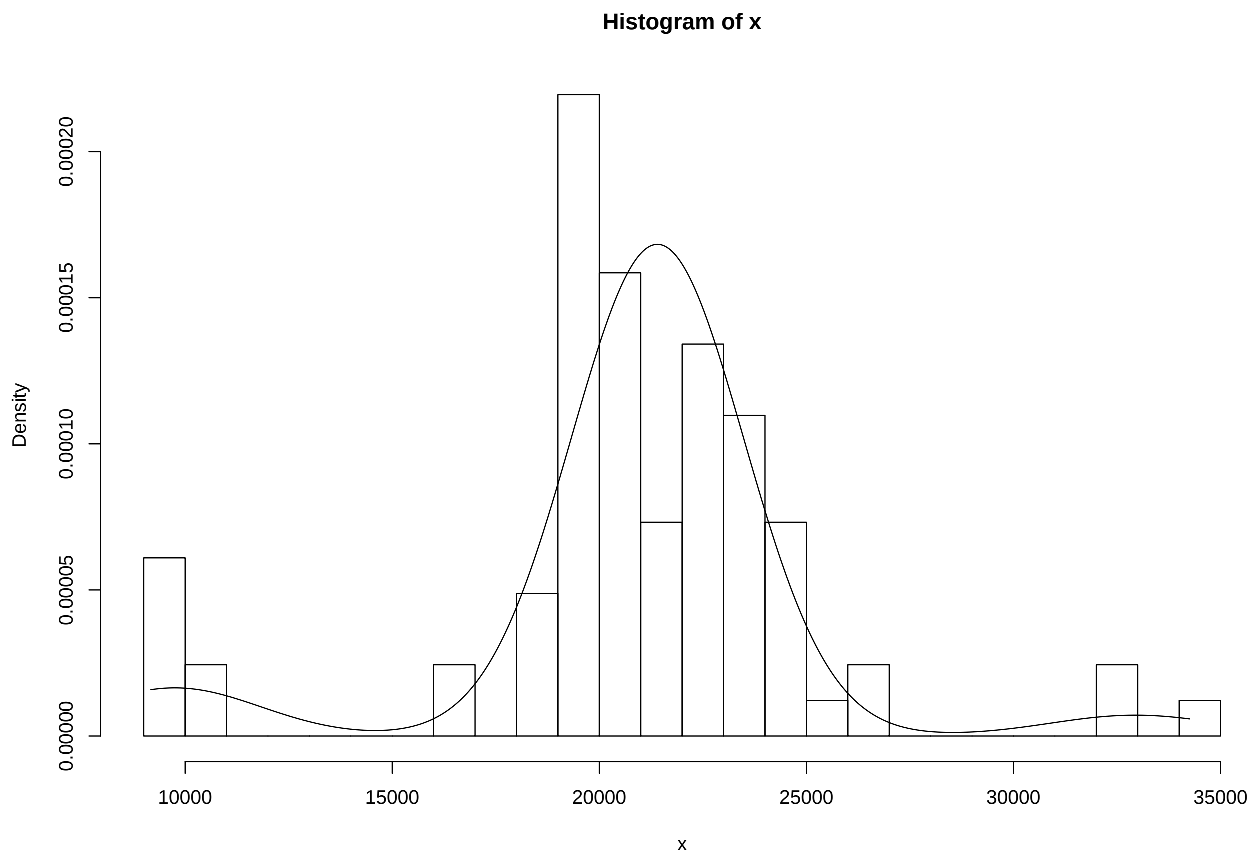
- Trace plots of EM iterations for parameters μ_1,μ_2,μ_3.



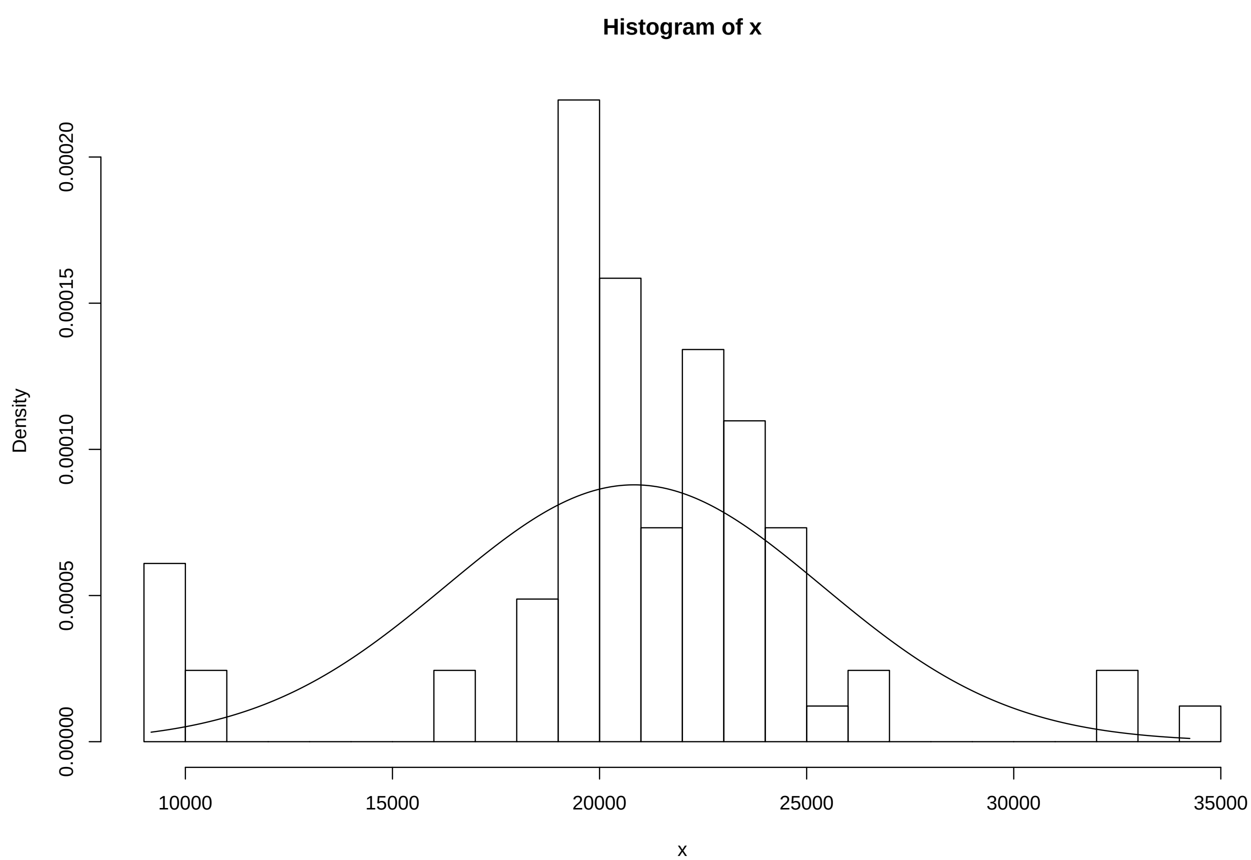
- Trace plots of EM iterations for parameters p_1, p_2, p_3.



- Histogram and estimated density curve for one EM solution.



- Histogram and estimated density curve for another EM solution.



These results suggest that the EM is run from different starting points. I considered 5 intial points but in other problems, more points may be required.

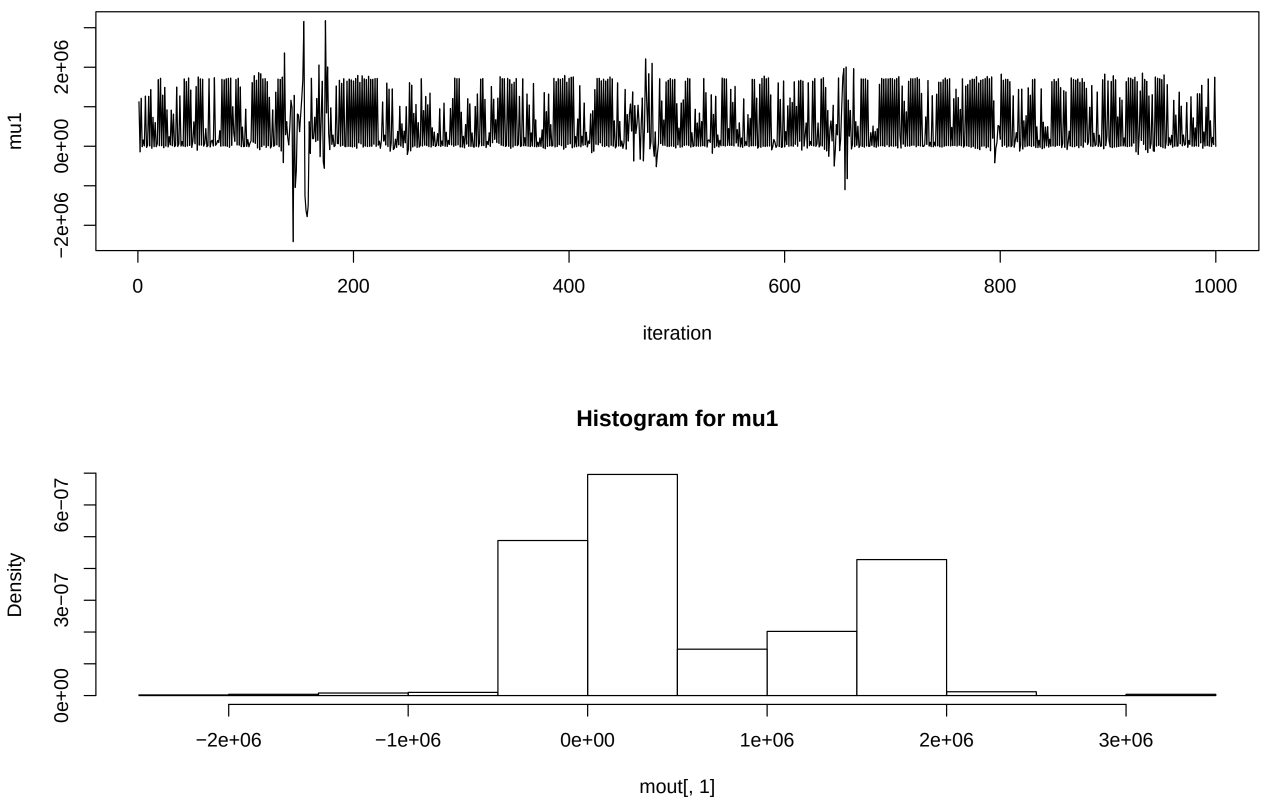
```
 out1
$mu
[1] 32944.347  9750.164 21403.202
$pr
[1] 0.03719463 0.08590094 0.87690443
$sig
[1] 4321231
 loglik(out1)
[1] -779.1162
 out4
$mu
[1] 32944.347 21403.202  9750.164
$pr
[1] 0.03719463 0.87690443 0.08590094
$sig
[1] 4321231
loglik(out4)
[1] -779.1162
 out5
$mu
[1] 21083.55 21083.56 20514.20
$pr
[1] 0.4436884 0.1135485 0.4427631
$sig
[1] 20533390
 loglik(out5)
[1] -806.8523
```

```
# Sta 590 Statistical Computing
# EM algorithm for mixture of normals example
# with k components
# mu is k-dimensional vector for the means
# sig is the variance
# pr is the vector of probabilities
# iter is the number of E-M iterations
#  x is the data vector
# n is the sample size
n <- 82
k <- 3
x <- c(9172,9350,9483,9558,9775,10227,
10406,16084,16170, 18419,18552,
18600,18927,19052,19070,19330,
19343,19349,19440,19473,19529,19541,
19547,19663,19846,19856,19863,19914,
19918,19973,19989,20166,20175,20179,
20196,20215,20221,20415,20629,20795,
20821,20846,20875,20986,21137,21492,
21701,21814,21921,21960,22185,22209,
22242,22249,22314,22374,22495,22746,
22747,22888,22914,23206,23241,23263,
23484,23538,23542,23666,23706,23711,
24129,24285,24289,24366,24717,24990,
25633,26960,26995,32065,32789,34279)
pst <- matrix(NA,n,k)
mat <- matrix(NA,n,k)
```

```
emmix <- function(mu,pr,sig)
 {
# calculation of pst
 for (i in 1:n)
  {
    aux <- dnorm(x[i],mu,sqrt(sig))*pr
    pst[i,] <- aux/sum(aux)
   }
   s <- sum(pst)
# calculation of pj's and mu'js
    for (j in 1:k){
     pr[j] <- sum(pst[,j])/s
     mu[j] <- sum(pst[,j]*x)/sum(pst[,j])
      }
# calculation of sigma^2
  for(i in 1:n){
     for(j in 1:k)
         {
        mat[i,j] <- (x[i]-mu[j])^2
         }
      }
  sig <- sum(mat*pst)/n
 return(mu,pr,sig)
  }
```

```
iter <- 500
# set initial values
mu <- rnorm(3,mean=mean(x),sd=sqrt(var(x)))
sig <- var(x)
pr <-rep(1/k,k)
mout <- matrix(NA,iter,7)
# EM-iterations
for (i in 1:iter)
 {
  out <- emmix(mu,pr,sig)
  mu <<- out$mu
  sig <<- out$sig
  pr <<- out$pr
  mout[i,]<-c(mu[1],mu[2],mu[3],pr[1],pr[2],pr[3],sig)
  print(i)
 }
# histogram and mixture estimate
hist(x,prob=T,nclass=30)
y <- seq(min(x),max(x),by=30.0)
z <- out$pr[1]*dnorm(y,out$mu[1],sqrt(out$sig))
+out$pr[2]*dnorm(y,out$mu[2],sqrt(out$sig))
+out$pr[3]*dnorm(y,out$mu[3],sqrt(out$sig))
lines(y,z)
loglik <- function(out)
 {
fxi <-  out$pr[1]*dnorm(x,out$mu[1],sqrt(out$sig))+
out$pr[2]*dnorm(x,out$mu[2],sqrt(out$sig))+
out$pr[3]*dnorm(x,out$mu[3],sqrt(out$sig))
ll <- sum(log(fxi))
}
```

- MCMC implementation. Trace plot and histogram for μ_1



- MCMC implementation. Trace plot and histogram for p_1