Topic 3 - EM Algorithm

Assignment 2: Mixtures of t-distributions

Isin Altinkaya

09.11.22

Theoretical background

The t-distribution with

- Location parameter $\mu \in \mathbb{R}$
- Scale parameter $\sigma > 0$
- Shape parameter (d.f.) $\nu > 0$

has density

$$f(x \mid \mu, \sigma^2, \nu) = \frac{\Gamma((\nu+1)/2)}{\sqrt{\pi\nu\sigma^2}\Gamma(\nu/2)} \left(1 + \frac{(x-\mu)^2}{\nu\sigma^2}\right)^{-(v+1)/2}$$

Theoretical background

Given i.i.d. observations $x_1, ..., x_n$ from the two-component mixture of t-distributions has a density in the form of

$$pf(x \mid \mu_1, \sigma_1^2, \nu_1) + (1-p)f(x \mid \mu_2, \sigma_2^2, \nu_2)$$

for parameters

- $p \in (0, 1)$
- $\mu_1, \mu 2 \in \mathbb{R}$
- $\sigma_1, \sigma_2 > 0$
- $\nu_1, \nu_2 > 0$, fixed shape parameters

Theoretical background

We can see this as

$$X = Z \cdot Y_1 + (1 - Z) \cdot Y_2$$

where Z, Y_1, and Y_2 are independent,

$$P(Z = 1) = 1 - P(Z = 0) = p$$

and $Y_i \sim t(\mu_i, \sigma_i^2, \nu_i)$.

Expectation step

Given a current set of parameters Θ, we compute the membership weights of data point x_i in the component $k \in 1, 2$ as

$$\pi_{ik} = P(z_{ik} = 1 \mid x_i, \Theta)$$
$$= \frac{pf(x_i \mid \mu_1, \sigma_1^2, \nu_1)}{pf(x_i \mid \mu_1, \sigma_1^2, \nu_1) + (1-p)f(x_i \mid \mu_2, \sigma_2^2, \nu_2)}$$

Expectation step

On the expectation step at (t)-th iteration, we need to compute $Q(\Theta \mid \Theta^{(t)})$

$$E_{\Theta^{(t)}}(Z_{ij} \mid y_j) = \pi_{ik}^k$$

$$E_{\Theta^{(t)}}(U_j \mid y_j, z_j)$$

for $i = 1, ..., g; j = 1, ..., n$.

Conditional expectation of the complete data log likelihood,

$$Q(\Theta \mid \Theta^{(t)}) = Q_1(p \mid \Theta^{(t)}) + Q_2(\sigma^2 \mid \Theta^{(t)}) + Q_3(\nu \mid \Theta^{(t)})$$

Maximization step

On the M step at the $(t+1)$-th iteration of the EM algorithm, for $i = 1, ..., g$, we update the mixing proportions given by the average of the posterior probabilities using $Q_1(p \mid \Theta^{(t)})$

$$p_i^{(t+1)} = \sum_{j=1}^{n} \pi_{ij}^{(t)} / n$$

And to update the estimates of μ_i and σ_i^2 $(i = 1, ..., g)$, we use $Q_2(\sigma^2 \mid \Theta^{(t)})$.

Maximization step

$$\mu_i^{(t+1)} = \sum_{j=1}^{n} \pi_{ij}^{(t)} u_{ij}^{(t)} y_j / \sum_{j=1}^{n} \pi_{ij}^{(t)} u_{ij}^{(t)}$$

This corresponds to the log likelihood function formed from n independent observations $y_1, ..., y_n$ with common mean μ_i and covariance matrices $\sigma_i^2/u_1^k, ..., \sigma_i^2/u_n^k$.

Thus it is equivalent to computing the weighted sample mean and sample covariance matrix of $y_1, ..., y_n$ with weights $u_1^{(k)}, ..., u_n^{(k)}$.

$$(\sigma_i^2)^{(t+1)} = \frac{\sum_{j=1}^{n} \pi_{ij}^{(t)} u_{ij}^{(t)} (y_j - \mu_i^{(k+1)})(y_j - \mu_i^{(k+1)})^T}{\sum_{j=1}^{n} \pi_{ij}^{(t)}}$$

Data simulation

- S3 object oriented programming approach

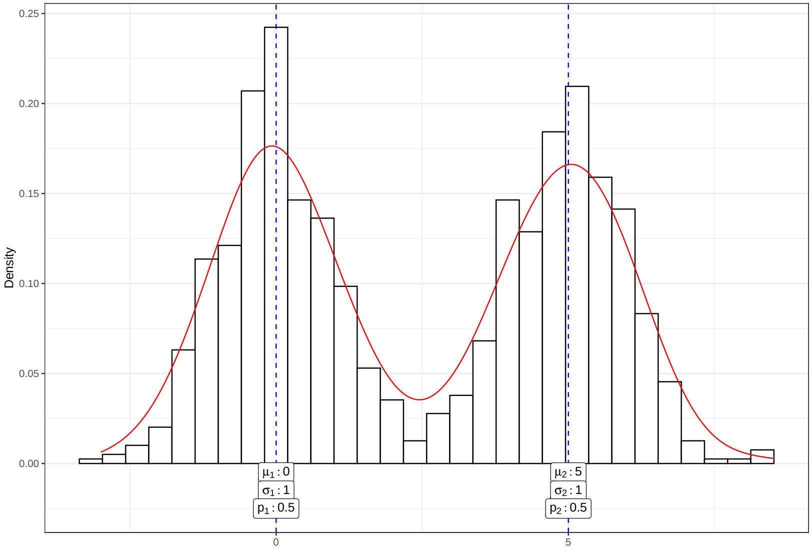
```
simulate_data_v1 <- function(n, p=0.5,
                             mus=c(0,1),
                             sigmas=c(1,1)){

  ps <- c(p,1-p) #membership weights

  data<-data.frame(x=seq(1,n),
                   y=c(rnorm(n=n*ps[1],mean=mus[1],
                             sd=sigmas[1]),
                       rnorm(n=n*ps[2],mean=mus[2],
                             sd=sigmas[2])))

  structure(class="sim_data",list(df=data,n=n,
      params=list(p=ps,mu=mus,sigma=sigmas)))
}
```

Data simulation



Computing observed Fisher information

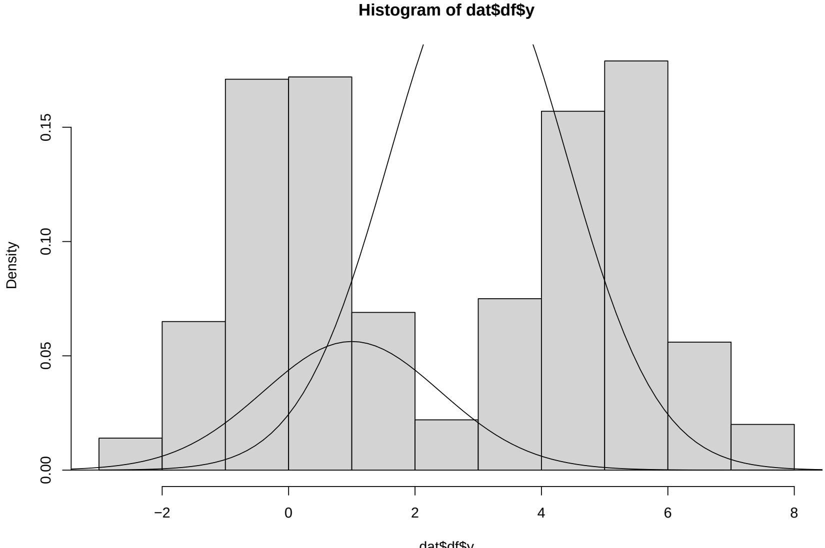
```r
dat <- simulate_data_v1(1000, p = 0.5, mus = c(0,
    5), sigmas = c(1, 1))
ihat <- optim(c(0.5, -1, 2, 1, 1), t_negll_v1,
    x = dat$df$y, hessian = TRUE, method = "BFGS")$hessian
# standard errors
sqrt(diag(solve(ihat)))
```

```
[1] 0.01603943 0.04494353 0.04758833
[4] 0.06539502 0.07313478
```

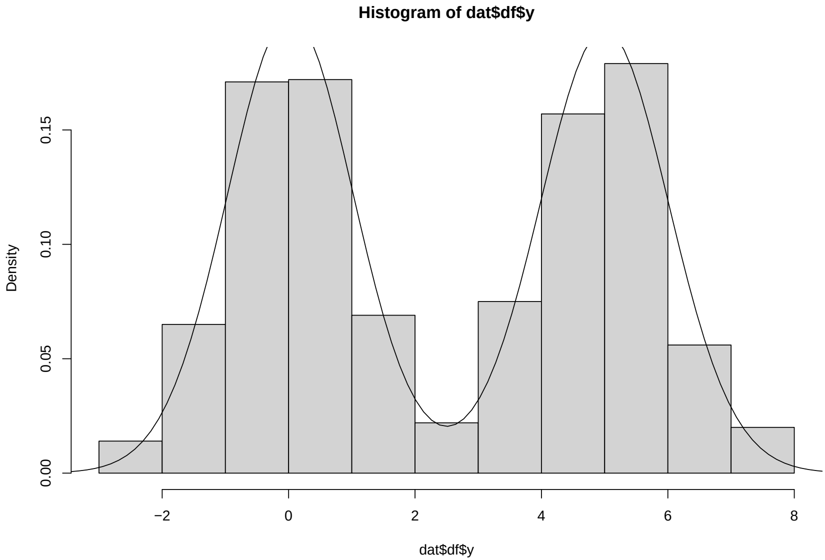
Before EM: Starting parameters

Simulation: p = 0.5, mus = c(0, 5), sigmas = c(1, 1)

Initial parameters: p = 0.2, mus = c(1, 3), sigmas = c(2, 2)



After EM: Estimated parameters



After EM: Estimated parameters

```
         p        mu1        mu2
0.50320620 0.02777232 4.99768279
    sigma1     sigma2
1.00672548 1.04121594
```